University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Doctoral Dissertations 1896 - February 2014

1-1-1972

An application of psychometric models and ATI methodology to the evaluation of instruction.

Daniel S. Sheehan
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_1

Recommended Citation
Sheehan, Daniel S., "An application of psychometric models and ATI methodology to the evaluation of instruction." (1972). *Doctoral Dissertations 1896 - February 2014*. 2730.
https://scholarworks.umass.edu/dissertations_1/2730

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations 1896 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

AN APPLICATION OF PSYCHOMETRIC MODELS AND ATI METHODOLOGY TO THE EVALUATION OF INSTRUCTION

A Dissertation Presented

by

DANIEL S. SHEEHAN

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

DOCTOR OF EDUCATION

December 1972

Educational Research--Science Education

AN APPLICATION OF PSYCHOMETRIC MODELS AND ATI METHODOLOGY TO THE EVALUATION OF INSTRUCTION

Daniel S. Sheehan

Abstract

The goal of this study was to investigate several strategies that pertain to the effectiveness of instructional programs in the schools. The research was divided into two components. One part was concerned with adapting instruction to individual differences and the other was focused on the evaluation of instruction.

The prevading theme for many current educational models is the individualized treatment of students. This particular research attempted to validate alternative instructional treatments by first designing differing treatments and then matching the treatments to specific learner characteristics.

The 285 students utilized in this study were randomly assigned to take their instruction in a three week segment of a science course under one of four instructional modes. The four instructional modes were labelled as follows: a) reading mode, b) media mode, c) programmed instruction mode, and d) teacher mode. Prior to beginning instruction on the module a wide selection of aptitude measures was administered to all of the students. As a criterion and delayed criterion measure a 40 item multiple-choice test was administered to all students immediately and one month after instruction was completed.

The test of the parallelism of slopes from regressing the criterion variable on each of the predictor variables was rejected with two predictor

variables. Rejection occurred when the Letter Sets Test was the predictor and the Module Posttest was the criterion and when the Delayed Module Posttest was used as the criterion and the Test Anxiety Scale for Children was used as the predictor.

The results obtained should provide some evidence to the teachers on how to individualize their program. Directions have been found for assigning students to the instructional treatments on the basis of Letter Sets and Test Anxiety Scale for Children test results.

A second area of concern in the study was the study of techniques for the evaluation of instruction. In spite of the obvious importance of evaluating the effectiveness of instruction, the development of appropriate tools has been slow. It was felt that the field of psychometrics offered several promising possibilities which could lead to the development of new evaluative tools. One of these is Wiley's latent partition analysis model. Another is the Tucker-Messick individual differences model for multidimensional scaling. The data appropriate for an LPA requires students to sort a group of concepts into categories considered to be homogeneous. Paired comparison judgments, which require the student to make similarity judgments between all possible paris of a set of concepts, provide the input for the Tucker-Messick model.

13 stimuli were chosen from the list of module objectives. To obtain the manifest categorizations needed for the latent partition analysis a sorting task questionnaire utilizing these 13 stimuli was developed and administered to students before and after instruction on the module. To obtain the input for the individual differences model for multidimensional scaling all possible pairs of the 13 concepts were arranged in a questionnaire

in paired comparison format. This questionnaire was also administered to the students both before and after instruction on the module.

The latent partition analyses revealed several interesting changes. Two concepts were placed in different latent categories after instruction. One ambiguity (a concept that had substantial loadings on more than one latent category) was removed by instruction and three others were created. The latent category composition of the high achievers differed from that of the low achievers.

The Tucker-Messick model revealed that instruction had the effect of producing an additional interpretable point of view. Instruction also appears to have provided the high achievers with more of a basis for organizing the concepts than it did for the low achievers.

ACKNOWLEDGMENTS

The completion of this study would not have been possible without the assistance of the following people. To all of them I extend my deepest thanks.

1. Dr. Ronald K. Hambleton---for his painstaking and knowledgeable guidance throughout this study. His complete dedication to his work will forever remain as a goal for me.

2. Dr. Leverne J. Thelen---for stimulating my interest for the need for conducting research in science education.

3. Dr. Hariharan Swaminathan---for his valuable statistical and methodological suggestions.

4. Deloris Fremouw---for her patience as well as for her outstanding secretarial skills.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1

Introduction

1.1 Background

The goal of this study was to investigate several strategies that pertain to the effectiveness of instructional programs in the schools. The research described in this study was divided into two components. One part was concerned with adapting instruction to individual differences and the other was focused on the evaluation of instruction. To highlight this distinction subheadings referring to these components will be used where-ever applicable.

Adapting Instruction to Individual Differences

The prevading theme for many current educational models such as Individually Prescribed Instruction (Glaser, 1968), Project PLAN (Flanagan, 1967), and a Model of School Learning (Carroll, 1963) is the individualized treatment of students. In individualizing instruction, one tries to provide learning opportunities that are in agreement with student needs, interests and aptitudes. Unfortunately, at this particular point in time we lack sufficient theoretical guidelines and empirical results to know just how this can best be done. Research in this area is critically needed if these models are to fulfill their potential.

What is the instructional method that will best adapt to a student's individual requirements and thus maximize his attainment of the instructional objectives? This, in essence, is the aptitude treatment interaction (ATI) problem, and is of utmost importance to the new instructional models (Glaser and Nitko, 1971). Cronbach and Snow (1969) state the problem formally:

> Assume that a certain set of outcomes from an educational program is desired. Consider any particular instructional treatment. In what manner do the characteristics of learners affect the extent to which they attain the outcomes from each of the treatments that might be considered? Or, considering a particular learner, which treatment is best for him?

Thus the basic premise of ATI theorists is that no single instructional process provides optimal learning for all students. Given predetermined educational goals, some students will be more successful with one instructional program and other students will be more successful with an alternative instructional program. Therefore, when instruction is differentiated for different types of students, a greater proportion of students will attain the instructional goals. The rationale for ATI use in school programs is convincing. This explains why nearly every new objective-based program builds in an ATI component (see for example, Hambleton, 1973).

ATI theorists attempt to build on individual differences as a way of establishing different paths toward the same educational goals. Hopefully, we will eventually develop an understanding of the factors that cause a pupil to respond to one instructional plan rather than to another. Yet this understanding can only be developed from a research basis.

Cronbach and Snow (1969) warn that the alternative instructional methods used by teachers must be validated. While many schools state that instruction is adapted to the individual, such adaptation has never been systematic because no one has known the principles that govern the matching of learner and instructional environment. Students themselves are poor judges of the educational environments that best suit them (Cronbach and Snow, 1969). Thus there are no short-term solutions to the problem of individual differences save the systematic design of investigations using alternative instructional schemes.

Evaluation of Instruction

In spite of the obvious importance of evaluating the effectiveness of instruction, it is perhaps surprising to note that the development of appropriate evaluative tools has been slow. It was our impression that the field of psychometrics offered several promising possibilities. [One psychometric model had been used by investigators in another study (Traub and Hambleton, 1971) with encouraging results.] As one of our purposes in this investigation we proposed the use of several psychometric models for the evaluation of instruction. The models were then used to evaluate instruction in a ninth grade science program and the results of the work are reported in Chapter 3. The models are discussed briefly in the next few pages and in more detail in chapter 2.

(a) Individual Differences Model for Multidimensional Scaling

In the words of Kruskal (1964) the problem of multidimensional scaling is to find n points whose interpoint distances match in some sense the experimental similarities of n stimuli. Thus multidimensional scaling methods utilize as input the observed similarity (or disimilarity) judgments among the stimuli under consideration. Relevant concepts from a subject matter area could be used as stimuli. It would then be possible to infer a concept space from similarity estimates between each pair of these concepts by multidimensional scaling methods.

Yet typical multidimensional scaling analyses do not concern themselves with individual differences in the observed similarity judgments.

They ignore any individual differences by averaging the similarity judgments over all individuals to obtain a single set of similarity values. The individual differences model for multidimensional scaling (Tucker and Messick, 1963), on the other hand, first isolates groups of individuals who tend to exhibit consistency in their similarity judgments. Single sets of similarity values are determined for each isolated group. These are then subjected to the usual multidimensional scaling analysis. Again where concepts are the stimuli, it is felt that the several concept spaces that result are more representative of actual people than the one "average" concept space that would result if the Tucker-Messick procedure was not used.

There is already some evidence to suggest that the Tucker-Messick individual differences model for multidimensional scaling is sensitive to the effects of instruction (Traub and Hambleton, 1971).

(b) Latent Partition Analysis

Latent partition analysis (LPA) is another psychometric model that may one day have widespread evaluative uses. The initial step with LPA was taken by Hess and Johnson (1971) who have shown that the model is sensitive to the effects of instruction in college physics. The various latent partition models operate by asking subjects to sort a group of stimuli into categories. It is assumed that although subjects construct different categories there is a hypothetical latent categorization of the stimuli underlying the individual sorts. From the sorting task data, the proportion of the sample, r_{ih}, to include both stimulus h and stimulus i in the same manifest category is computed. The matrix which contains these joint proportions is then factored to yield (1) the loadings of the stimuli on the underlying or latent categories, and (2) the probabilities of any pair of stimuli from two different latent categories occurring in the same manifest category. Again using concepts from a subject matter area as stimuli, it is possible to infer the latent categorization of the concepts from the sorting task data.

1.2 Review of the Literature

Adapting Instruction to Individual Differences

Bracht (1970) analyzed 90 research studies which could be examined for an ATI relationship. Bracht used two criteria for selecting the studies: (a) the study had to include a comparison of two or more alternative treatments for attaining a common set of objectives and (b) the study had to include one or more personological variables (aptitudes) so the comparison between alternative treatments could be made for subjects at different levels of the personological variable. In most of the studies the experimenter used a treatments-by-levels factorial design with analysis of variance. Bracht found only five studies with significant disordinal interactions. As was hypothesized by Bracht, the five disordinal interactions were obtained in experiments with controlled treatment tasks, factorially simple aptitude variables, and specific dependent variable measures. Yet numerous other studies, at least superficially, met these requirements and revealed no disordinal interactions. While admitting that the present picture did not look too good, Bracht seemed to feel that results would improve. In most of the studies, the alternative treatments were not developed with the ATI concept in mind. The analysis of an interaction effect was often an afterthought rather than a carefully planned part of the study.

The most complete ATI literature review was conducted by Cronbach and Snow (1969). They analyzed studies dealing with possible interactions of general ability with specific treatments such as programming techniques, meaningfulness of instruction, and complex instructional methods. Specialized abilities and various personality variables and their possible interactions with treatments were thoroughly reviewed. Summarizing the field, Cronbach and Snow labeled ATI research as "frustrating" and "disappointing". Few or no ATI effects have been solidly demonstrated. Most of the research has been inconclusive, either because questions were badly put or because investigations contradicted each other. According to Cronbach and Snow, the quality of analysis and reporting is such that the conclusions of the original author are about as likely to be incorrect as correct.

Cronbach and Snow provided some guidelines for future ATI research. The treatments used in past experiments have generally suffered from brevity and artificiality. They stated that we are not going to learn how students respond to instructional treatments by mimicing laboratory experiments, by presenting a single brief lesson repetitively until it is mastered, or by introducing utterly artificial motivational procedures. One possible approach is to use ongoing educational programs. While lacking randomness, these studies can employ large samples and collect data over long periods of time. Another alternative is to contrast two or more adaptations of regular instructional material. Contrasting versions of instructional materials or alternative programs of activities can be implemented without disarranging the school program.

Cronbach and Snow recommend process analysis for designing alternative treatments. To be differentially effective for various types of students, the alternative treatments should demand different abilities for successful performance. Within a certain treatment, the processes needed for mastery are identified. After process identification, judgments are made about the ability which is related most significantly to the processes. Then an attempt is made to develop an alternative treatment in which different processes are called for to attain the same instructional objectives. The ability to perform the second set of processes should be unrelated to the ability to perform the original set of processes. For example, one treatment could be designed to rely heavily on general ability and the other treatments could be aimed at more specific abilities. The treatments should result in the same level of objective attainment, if individualization is carried out properly.

A few tentative principles for the design of treatments can be extracted from the ATI literature. Alternative treatments that reduce the burden of semantic processing of verbal information seem likely to give flat-slope treatments (indicating a lack of dependence on general ability). This can perhaps be done by making the instructional presentation more obvious, through such communication devices as easier vocabulary, repetition and paraphrase, and audiotapes that the learner can hear while he is following the text with his eyes, etc.

Another possibility is the placing of a greater responsibility on the learner for organizing material in his own way. There are hints that the more able student responds positively to this kind of challenge, and does less well where his interpretation is constrained by a strong didactic structure. Cronbach and Snow continually point out that these results are highly equivocal.

In dealing with aptitudes, one could pit the fluid and crystallized segments of general ability against each other. Over the years, the high "nonverbal IQ" has been an embarrassment to educational psychology. Everyone has recognized its reality but no one has discovered a generally applicable instructional approach that serves the pupil who is strong on the nonverbal side and relatively weak on the verbal side (Cronbach and Snow, 1969).

In the area of personality variables, Cronbach and Snow labeled the results as almost entirely disappointing. Although hints of interaction can be found, the effects, weak at best, are inconsistent from one school subject to another and, within an experiment, from one measure of achievement to another. Because of the changeable nature of personality variables, interpretable results may only be obtained by collecting data at several points in time.

Thus if different learning styles exist, they are very elusive. As an example of the instability of ATIs, consider the research of Bunderson (1969). After revising an instructional treatment to simplify it and reduce the time consumed by its administration, he found that ATIs obtained with the original instruction were reversed for the revised version. A study by Burton and Goldbeck (1962) offers further evidence supporting the unpleasant possibility that reducing the difficulty of the learning task may reverse the ATI.

It is hoped that this review of the ATI literature puts the problem of matching learner characteristics to instructional treatments in its proper perspective. Pitfalls and drawbacks were considered to emphasize the existing state of the research. Guidelines for future ATI studies were stated in the hope that they may eventually alleviate the confusion that currently prevails.

Evaluation of Instruction

(a) Individual Differences Model for Multidimensional Scaling

Most multidimensional scaling studies have concerned themselves with uniform and relatively simple stimuli. A study by Root (1963), however, showed that multidimensional scaling could be used in determining the dimensionality of an extremely complex psychological attribute, the meaning of words. The scaling analysis of the interstimulus distances of twenty English nouns yielded seven dimensions. Five of the dimensions proved amenable to psychological interpretation.

In a 1964 study, Hall asked whether the dimensionality of a stimulus space changed after concept attainment. The experimental group learned a concept that related the parts of a set of figures. The control group was not introduced to the concept. The similarity of the set of figures was rated by both groups, both before and after the instructional period. The change in the structure of the stimulus space was estimated from multidimensional scaling analyses of the similarity ratings obtained before and after the concept was learned. Hall found that before concept attainment, the structure of the stimulus space was two- or three-dimensional. After the concept was learned, the structure of the stimulus space was estimated as four-dimensional. The additional dimension was related, in part, to the concept learned. These results showed that the concept that was taught was actively assimilated by the subjects. Once assimilated it served as an additional basis by which the subjects related the stimuli. The concept was not just stored as an inert collection of words, but affected the analysis and output of similarity judgments.

Traub and Hambleton (1971) examined the effect of instruction on the semantic space of concepts taken from a course on educational testing and measurement. The students' similarity judgments of all possible pairs of thirteen concepts were obtained before and after they had taken the course. Using an individual differences model for multidimensional scaling, the authors concluded that instruction had the effect of eliminating whatever differences existed in the way students viewed the concepts. The post-course semantic space was more highly organized and smaller as instruction removed possible extraneous dimensions.

It should be pointed out that only the Traub and Hambleton study utilized what essentially is a two-stage method of multidimensional scaling. The other studies used typical one-stage analyses whose input data was based upon an average of the responses over all individuals in the sample. Even if extensive differences existed in individual perceptual spaces, the one-stage scaling methods would blend them together in deriving the average structure for the sample.

The additional stage utilized in the Traub and Hambleton study was Tucker and Messick's (1963) individual differences model for multidimensional scaling or "points of view" analysis. This model gets at the number of factors that are needed to account for the covariation in the responses of the individuals in the sample. A set of distance values is obtained for each of these factors or "points of view." Each of these sets of distance values is analyzed by one of the standard methods of multidimensional scaling (stage two of the analysis). Being capable of isolating more than one "average" response pattern within a sample, it is felt that the prior utilization of the Tucker-Messick procedure results in the empirical isolation of consistent individual viewpoints that would remain hidden in the typical multidimensional scaling analysis. The luxury of dealing with group data is still present, yet some of the hazards of averaging across the total group are avoided.

(b) Latent Partition Analysis

The original latent partition analysis model was put forth by Wiley (1967) and generalized by Evans (1970a). Since its inception several investigators have demonstrated the feasibility of the implementation of sorting methodology for the operationalization of the concept of cognitive structure in subject matter areas. In a study by Johnson *et al*. (1970) 40 economics terms were given to two groups of ninth grade students who were instructed to sort the items into categories according to their knowledge of economics. In one group of students (Econ I) the teacher had used approximately six weeks time to cover two units of work relevant to economics. The two units were from the *Harvard Social Studies Curriculum* which stresses a legal-ethical approach to learning content in economics. The second group (Econ II) had used the entire semester to cover an economics course using the textbook *Comparative Economics Systems* which stressed the learning of analytical economics concepts. An analysis of the results by latent partition analysis procedures and subsequent multidimensional scaling analyses (Torgerson, 1958) revealed that the Econ II students formed fewer latent categories, evidenced greater discrimination between categories and evidenced a somewhat more integratively complex schemata operating in their grouping of the items.

In another study, Hess and Johnson (1971) extracted 100 physics concepts and terms from an elementary college physics course. These 100 items were randomly placed into two groups of 50 items each. One group was given to each half of a class of physics students with instructions to sort the items into categories according to their knowledge of physics. Two weeks later, at the end of the quarter, the same groups of items were again sorted by the same students. For the data analysis the students were divided into high, middle, and low achieving groups (on the basis of course examinations). Latent partition analysis and a subsequent multidimensional scaling were used to analyze the data. [As in the Johnson et al. study the matrix indicating the probabilities of any pair of items from two different latent categories occurring in the same manifest category was used as a measure of category relatedness. Conceiving these probabilities as distances, a multidimensional scaling analysis of these latent category probabilities yields a model of category relatedness (Hess and Johnson, 1971).] The data analysis revealed that changes in category relatedness within each achievement group resulted from instruction in physics. Changes in category relatedness between achievement groups were also apparent. The low group tended not to show a greater differentiation between the various categories upon course completion.

The implications of the preceding two studies for instructional evaluators should be obvious. Latent partition analysis methodology facilitated the detection of changes in cognitive structure that resulted from two different approaches to economics instruction. Also, it was possible to detect differences in cognitive structure among low, middle, and high achieving physics students.

The preceding review presents the existing state of knowledge in the adaptation of "points of view" and latent partition models to the instructional setting. These studies indicate the utility of the two models for assessing the structure among concepts and thus they provide a mechanism for assessing the effectiveness of instruction relative to this structure for a set of course-related concepts.

1.3 Purposes of the Investigation

Adapting Instruction to Individual Differences

Jamesville-DeWitt High School in DeWitt, New York, utilizes a new instructional model in its ninth grade science course. The course is organized into instructional modules which consist of instructional activities designed to teach a single major concept which has been expressed in terms of behavioral objectives. The instructional activities which make up a module are organized into smaller submodules called learning activity packages (LAPs). To assess a student's progress and diagnose his learning deficiencies, criterion-referenced tests are administered before he begins work on the module, before and after he completes the LAPs, and after he has completed the module (Hambleton and Gorth, 1971).

Wherever possible, alternative instructional activities have been developed. These have been primarily used as alternative LAP material and as remedial and enrichment activities. There has been no attempt to match the alternative instructional activities to specific learner characteristics. As Cronbach and Snow (1969) state, the alternative instructional activities used in a program must be validated. It may be that no relationship exists between feasible instructional activities and measurable learner characteristics. But if relationships exist, the individualized instruction component of the program can be implemented.

Thus the first purpose of our study was to investigate the interactions between a wide selection of aptitude variables and instructional treatments for a small segment of instruction (a module) in the ninth grade science program. Information bearing on the ATI question would be extremely important to science teachers in the Jamesville-DeWitt program.

To avoid creating confusion the specific instructional treatments used in this study were not labelled by any distinguishing features (such as the reading mode or the media mode). Such labelling diverts the focus from the treatment as a functional unit, and may even be erroneous because the relative amounts of the features are not known. That is, how much "reading" is in a reading mode or how much "media" is in a media mode? Thus the treatments in this study were simply designated as treatments 1, 2, 3, and 4.

The major features of the four instructional treatments were as follows:

Treatment 1 - Students assigned to this instructional treatment met in a teacher-led class and were presented a series of lectures by the ninth grade science teachers. The three ninth grade science teachers were instructed to prepare their classes in their usual manner of lecture preparation. The only restrictions imposed were that the lectures cover the module objectives and the audiotapes, videotapes, and programmed instruction per se not be utilized. The teachers could use demonstrations, assignments, or additional worksheets if they felt them appropriate.

The students were allowed to proceed as they would in an ordinary lecture situation. That is, they could take notes, read any books that they became interested in, do any assigned problems, etc. They were not specifically instructed to engage in any of these activities.

Treatment 2 - The instructional materials for this treatment consisted of one videotape, seven audiotapes, and four worksheets. Each student assigned to this treatment received a media handout consisting of an assignment section and a section containing the worksheets. The assignment section told the students which worksheets or parts of worksheets were to be used with the videotape and the various audiotapes. The students were generally required to respond in writing on the appropriate worksheets. Direction was provided by the videotape and by the audiotapes.

The students assigned to this treatment viewed the videotape in small groups in one of the science rooms. Enough cassette recorders were available at the Resource Center so that the audiotapes could be used by pairs of students or by single students.

Students were instructed to return their media handouts to one of the science rooms each day after they were finished using them.

Treatment 3 - This treatment consisted of a reading handout composed of an assignment section and a section of appropriate readings. The readings were extracted from several ninth grade science books and put together in such a manner so as to effectively cover the module objectives. Each student who was assigned to this instructional treatment worked alone on his booklet of readings. These students never responded by writing. They were only required to read the instructional material.

The students worked on their reading handouts during class time and then returned them to one of the science rooms each day after they were finished using them.

Treatment 4 - This treatment consisted of 185 frames of programmed instruction material divided into five booklets. One frame was presented per page. The student was instructed to read each frame and to write the necessary response in the space provided. He was then directed to turn the page to discover the correct answer, the answer appearing above the next frame. If the student's answer was correct he was to proceed to the next frame which appeared farther down the page. If he was not correct he was instructed to go back and re-read the frame until he was able to obtain the correct information. The student progressed through the booklets in this manner.

It should be added that all of the students assigned to this instructional treatment worked alone on one set of programmed instruction booklets. They returned their booklets to one of the science rooms each day after they were finished using them. As with treatments 2 and 3, this was done to prevent loss of the instructional materials.

It was extremely difficult to decide whether to develop specific ATI hypotheses. Given our instructional modes, the ATI literature could have been searched for studies with superficially similar instructional modes. Using the results of these studies as a basis, possible interactions with the instructional modes and aptitudes could have been hypothesized. Yet it was felt that this approach would be misleading because of the confusing state of the ATI literature.

General educational and psychological theory could also have been used as a basis for the hypothesis development. The drawback with this approach was the complex nature of our treatments. The individual treatments undoubtedly contained dimensions that are important in the learning process. The lack of controls in the experimental study, however, made it impossible for us to handle them in a suitable way. In fact there were probably several dimensions relating to the treatments that affected learning that we were unaware of. This position becomes more understandable if, for example, we imagine we are pharmacologists, not educators, and our task is to evaluate the ability of four antibiotics to kill the typhoid baccilus. What would we <u>have</u> to know about our antibiotic treatments? We would have to know the composition of the treatments. That is, we would have to know how much of a particular antibiotic we have in a treatment, or if we are using mixed treatments, which antibiotics and how much of each is present in the mixture. We would not consider labelling a particular treatment as "streptomycin" because we happened to shake a little streptomycin into our bacterial culture. Or if we wanted to test the combined effects of streptomycin and aureomycin we would not just haphazardly shake some of each into the culture tube. Rather, we would systematically test carefully controlled amounts of streptomycin with carefully controlled amounts of aureomycin. We would have to know exactly how much of the particular antibiotics we were using.

In spite of the problems in developing ATI hypotheses, there were some obvious differences among the instructional treatments that suggested several possible interactions.

It was hypothesized that two of the instructional treatments would give relatively flat-slope regression lines, though for possibly different reasons. Treatment 4 seemed to be more repetitive than the other treatments. Treatment 2 should enable the student to cover the relevant material with a minimum of semantic processing. Both of these treatments appeared to put less emphasis on traditional measures of academic aptitude. That is, they should favor the "less able" student.

Cronbach and Snow (1969) state that treatments which place more responsibility on the student for organizing material are responded to positively by the "more able" student. Such a student does less well in a restrictive instructional situation.

Treatment 3 allowed the student to function in a relatively independent manner. The subject matter is not as cohesively organized as it is in the other treatments. Various reading sources are coalesced to cover the module objectives. More able students may find this situation stimulating, while the less able students may not be able to function in such a relatively unrestrictive situation.

The interactions between certain treatment pairs and various aptitude variables seemed particularly promising. For example, when IQ is used as the aptitude variable, treatment 3 should give a much steeper regression line than treatment 1 because the increased amount of "organizational ability" needed for treatment 3 favors high ability students. An aptitude variable such as achievement motivation may also cause the regression line of treatment 3 to be steeper than that of treatment 1. The highly motivated students may respond positively to the increased demand for self-organization. The students lacking in achievement motivation may perform better when the organization is provided by an external source (as it appears to be in treatment 1). These hypotheses are very speculative though, since we do not know for certain if our treatments reduce the burden of semantic processing or require a higher level of student organizational ability. While certain treatments may seem to possess varying amounts of these attributes, they have not actually been measured on "semantic processing" or "organizational ability" scales. They have also not been measured on scales of other variables that may be relevant to the learning process.

Evaluation of Instruction

(a) Individual Differences Model for Multidimensional Scaling

There is some evidence to suggest that an individual differences model for multidimensional scaling is sensitive to detecting the effects of instruction thus making the model useful as an evaluative tool. Will this result also apply to a concept space defined by a series of chemical concepts? The second purpose of our investigation was to discover whether: 1) instruction affected the number of points of view and the dimensionality of the concept space for each point of view, 2) the type of instructional treatment affected the number of points of view and the dimensionality of the concept space for each point of view, and 3) the high and low achieving students (on the basis of Module Posttest scores) have the same (in terms of number of points of view and the number of dimensions) concept spaces. Answers to these questions would provide useful evaluative data for understanding and improving the quality of instruction in the instructional module. In addition to the results the feasibility of the model as an evaluative tool was under study.

(b) Latent Partition Analysis

The Johnson et al. (1970) and Hess and Johnson (1971) studies have indicated that latent partition analysis is also sensitive to the effects of instruction. Our final purpose was to consider the potential of latent partition analysis as an evaluative tool. Specifically we were interested in knowing whether: 1) instruction affected the mean category cohesiveness of the latent categorization, 2) the type of instructional treatment affected the mean category cohesiveness of the latent categorization, and 3) the high and low achieving students (on the basis of Module Pretest and Module Posttest scores) have the same (in terms of mean category cohesiveness) latent categorizations.

1.4 Significance of the Problem

Adapting Instruction to Individual Differences

Many new instructional models focus on making a more concerted effort toward meeting the needs of individuals. While philosophically sound, techniques of individualizing instruction must be developed before the proposed models can be fully implemented. ATI fits into this scheme because it represents perhaps the ultimate in individualization possible in a school setting. ATI information can be used to match specific instructional methods to selected learner characteristics. This matching of student to treatment should be optimal in the sense that selected outcome variables will be maximized.

Evaluation of Instruction

What do we usually mean when we say that a student has mastered a particular subject area? We usually mean that he has attained some criterion score on a measuring instrument that samples the content domain of the subject area. Yet, in the terminology of Bloom, most conventional measuring instruments sample only the knowledge and comprehension levels. These are the lowest levels of the cognitive domain. More complex levels of cognition are frequently stated as curriculum objectives, but all too often are not measured effectively.

There is a definite need for more appropriate measurement of curriculum objectives that are categorized at the higher cognitive levels of Bloom's taxonomy. By assessing these levels more appropriately we can make more definitive statements about the outcomes of a particular instructional treatment. Bloom feels that the emphasis on the lower level objectives has stiltified educational research (Bloom, 1956). He states that it is relatively easy to understand why so many of the research studies - on large and small classes, teaching by television vs. teaching by regular classroom procedures, lectures vs. discussions, demonstrations vs. laboratory experiences, the use of programmed learning materials or audiovisual techniques, and even the independent use of books and other printed materials - all give very similar results when measured by knowledge level evaluation instruments.

It is felt that the interrelationships of concepts are the critical aspect of any learning situation, both practically and theoretically. When we measure learning in terms of isolated bits of information, we are not effectively sampling the whole domain of learning outcomes. We slight the student by not adequately assessing his learning, and in the process, fail to provide an effective evaluation of the instructional treatment.

The more complex levels of the cognitive domain seem to require more sophisticated measuring instruments than those that are commonly in use. Objective tests, for the most part, only sample the knowledge and comprehension levels. While essay tests may, in some situations, utilize the more complex cognitive levels, they are frequently of low reliability. It would appear that new methods of evaluation are needed here.

Paired comparison judgments (analyzed using multidimensional scaling methods) require the student to make similarity judgments between all possible pairs of a set of concepts. Sorting tasks (analyzed using the latent partition analysis model) require the student to arrange a group of concepts into categories that he considers to be homogeneous. Using Bloom's classification system, it appears that these types of judgments and tasks would facilitate the assessment of higher levels of the cognitive domain, in particular the analysis and synthesis levels. As the student would be required to make similarity judgments or categorizations, he would have to analyze the relationships between the concepts and then put the relevant elements of the concepts together to form an interpretable whole.

Chapter II

Methodology

2.1 Design of the ATI Study

As previously stated, the first part of this study was concerned with investigating possible interactions between aptitude variables and instructional treatments. The specific instructional treatments were designed for one module of instruction in the ninth grade science program at Jamesville-DeWitt High School, DeWitt, New York. The module was entitled "The Structure of Matter" and the eleven objectives of this module were used as the basis for designing the instructional treatments. A list of the objectives appears in appendix A. The instructional treatments have been described in section 1.3.

Assignment to Instructional Treatments

There were seven periods throughout the school day during which ninth grade science was taught. In three of the seven periods only one science class was conducted per period. To circumvent the artificiality of teachers lecturing to extremely small groups, treatment 1 (the so-called teacher mode) was not used in these three periods. Students were assigned to the other three instructional treatments on a random basis.

In four of the seven instructional periods there were at least two science classes conducted per period. All four instructional treatments were used in these periods. Initially, however, to compensate for not using treatment 1 in the periods with only one science class per period, a compensatory number of students were randomly assigned to treatment 1. The remaining students in each of these four periods were randomly assigned to one of the four instructional treatments.

Pre-Instructional Treatment Measuring Instruments

Prior to the beginning of instruction on the "The Structure of Matter" module the following measurements were made on the students in the ninth grade science program:

1. Module Pretest
2. Lorge-Thorndike Intelligence Test
    a. Verbal IQ Score
    b. Non-Verbal IQ Score
3. SRA Achievement Series
    a. Science Test
        (1) Capitalization and Punctuation Score
        (2) Spelling Score
        (3) Grammatical Usage Score
        (4) Total Language Arts Score
    c. Arithmetic Test
        (1) Reasoning Score
        (2) Concepts Score
        (3) Computation Score
        (4) Total Arithmetic Score
    d. Reading Test
        (1) Comprehension Score
        (2) Vocabulary Score
        (3) Total Reading Score
    e. Total Battery Composite Score (includes all tests except the Work-Study Skills)
    f. Work-Study Skills Test

(1) References Score

(2) Charts Score

(3) Work-Study Skills Score

4. Survey of Study Habits and Attitudes

   a. Work Methods Score

   b. Delay Avoidance Score

   c. Study Habits Score

   d. Teacher Approval Score

   e. Education Acceptance Score

   f. Study Attitudes Score

   g. Study Orientation Score

5. Junior Index of Motivation (JIM Scale)

6. Test Anxiety Scale for Children

7. Children's Manifest Anxiety Scale

   a. Lie Scale Score

   b. Anxiety Scale Score

8. Intellectual Achievement Responsibility Scale

   a. Responsibility for Successes Score

   b. Responsibility for Failures Score

   c. Total Responsibility Score

9. Mathematics Test

10. Letter Sets Test

11. Student Attitude Questionnaire (eight concept scores and a total score)

Complete descriptions of the tests are presented in Appendix A.

<u>Post-Instructional Treatment Measuring Instruments</u>

As a criterion measure a 40 item multiple-choice test was administered to all students immediately after instruction had been completed. This test was also re-administered one month later. A description of the test appears in Appendix A.

To discover how the students felt about the instructional materials a Student Questionnaire was administered to all students after instruction had been completed. The questionnaire is described more fully in section 3.4.

2.2 Design of the Evaluative Study

As a basis for evaluating the four instructional treatments, 13 concepts were chosen from "The Structure of Matter" module. These 13 concepts were selected because they seemed to be representative of the concepts defined by the module objectives. The number of concepts was limited to 13 to prevent student boredom on some of the evaluative tasks (Traub and Hambleton, 1970). The particular concepts that were chosen were proton, ion, neutron, element, radical, compound, electron, nucleus, atomic number, atomic weight, mixture, Shell or energy level, and oxidation number. These concepts were arranged into a Student Sorting Task and a Similarity Judgment Questionnaire. Both the Student Sorting Task and the Similarity Judgment Questionnaire were given to all ninth grade science students just prior to beginning the instructional module and immediately after the module was completed. Both of these instruments will be described more fully in the following sections.

Individual Differences Model for Multidimensional Scaling

The input for the individual differences for multidimensional scaling analysis is the similarity judgments between all possible concept pairs. Using the previously mentioned concepts, all 78 possible concept pairs were arranged in questionnaire form (Traub and Hambleton, 1970). A method developed by Ross (1934) was used to order the 78 concept pairs in the questionnaire. This balanced ordering maintained the greatest possible spacing between pairs involving similar members, avoided regular repetitions, and eliminated time and space errors. A nine-point rating scale was placed beside each concept pair in the questionnaire. The lower numbers of the scale were to be used for concepts viewed as being similar, the higher numbers for concepts viewed as being different, and the middle of the scale was for indifference judgments. The questionnaire cautioned the students to be thoughtful and to focus on the attributes held in common by a pair of concepts. The students were also instructed to spread their judgments over the nine categories of the rating scale.

To provide evidence about the reliability of the similarity judgments, 13 additional concept pairs were appended to the original 78. These concept pairs were chosen from throughout the original group of 78 pairs so that each individual concept appeared twice. With these 13 pairs the order of the concepts was reversed from that of the first presentation of the pair in the questionnaire. Using the data from these additional 13 concept pairs, it was possible to get individual estimates of the reliability of the similarity judgments by correlating judgments on each pair summing across pairs. By summing over individuals for all 13 concept pairs on the two occasions and then correlating the two sets of numbers a group estimate of the reliability can be obtained. A group estimate of the reliability for each concept pair can be obtained by correlating the two responses to the pair summing across individuals. Of the three reliabilities, the group reliability for all 13 concept pairs appeared to be the most useful for our purposes. A copy of the Similarity Judgment Questionnaire is included in Appendix B.

Latent Partition Analysis

To obtain the manifest categorizations needed for the latent partition analysis a sorting task questionnaire with 13 concept cards were utilized. Each card had one of the previously listed concepts printed on it. The questionnaire instructed the students to sort the cards into categories. No restriction was placed on the number of categories that a student could form and the students were told to sort the concepts on whatever basis they thought appropriate. A sample categorization was presented on the questionnaire. Once the students completed the sorting task, they were told to record their categories with their component concepts on the questionnaire. A copy of the Student Sorting Task is included in Appendix B.

2.3 Methods of Analysis

ATI Methodology

Integral to the discussion of ATI are the terms ordinal and disordinal interaction. Berliner and Cahen (in press) cogently illustrate these concepts by presenting regression diagrams (similar to figures 2.3.1, 2.3.2 and 2.3.3 shown on the next three pages).

In Figure 2.3.1 there is no interaction between aptitude and treatment. The regression slope of criterion on aptitude is identical for the two treatments, although treatment one is superior to treatment two. The mean criterion score is always greater for treatment one than for treatment two, regardless of the aptitude level. If the outcome variable represented

Figure 2.3.1 Illustration of no interaction between aptitude and treatment.

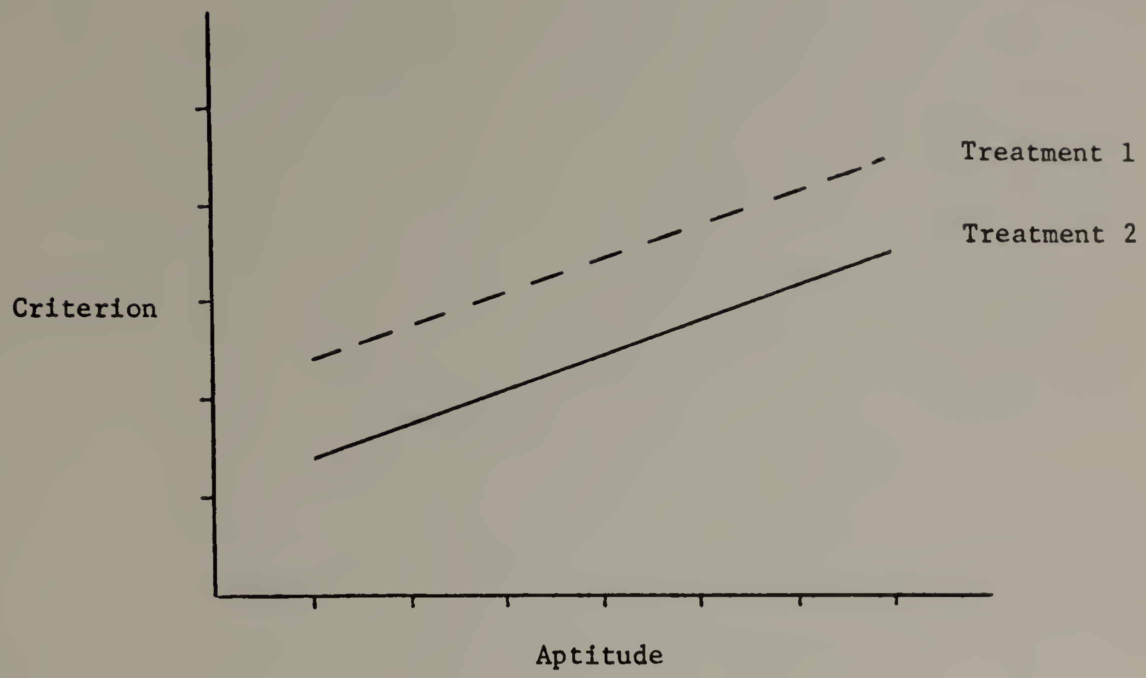


Figure 2.3.2 Illustration of an ordinal aptitude-treatment interaction.

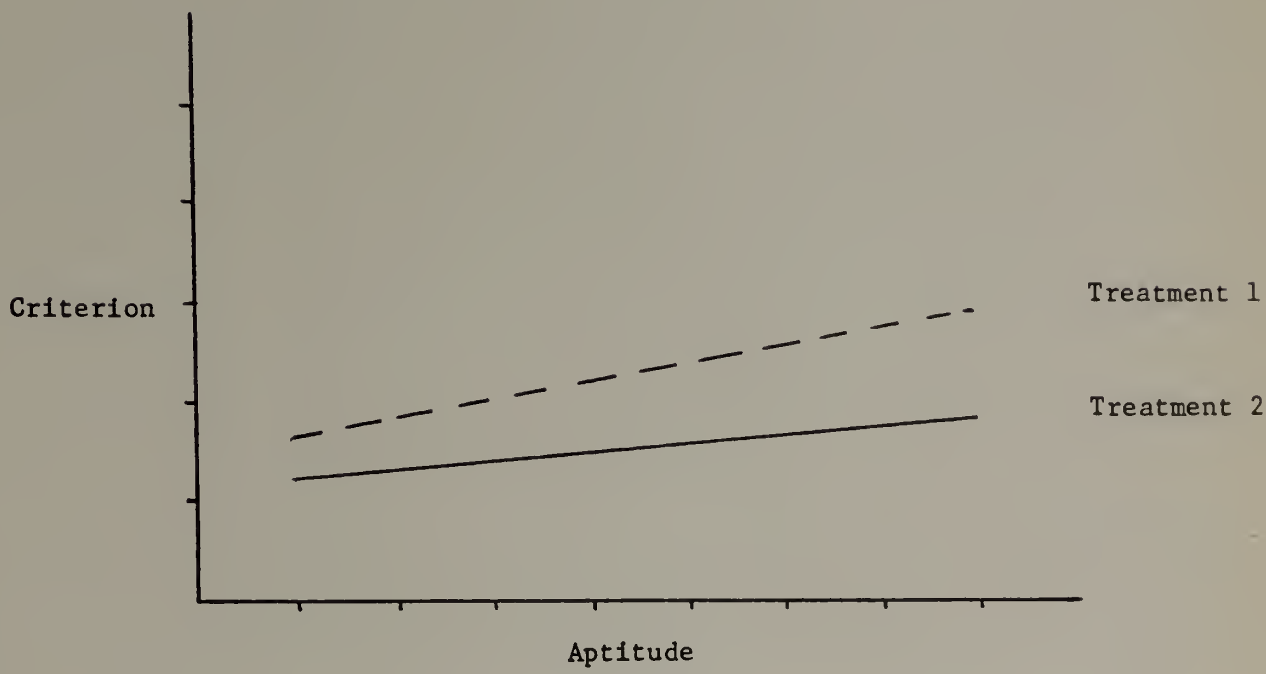
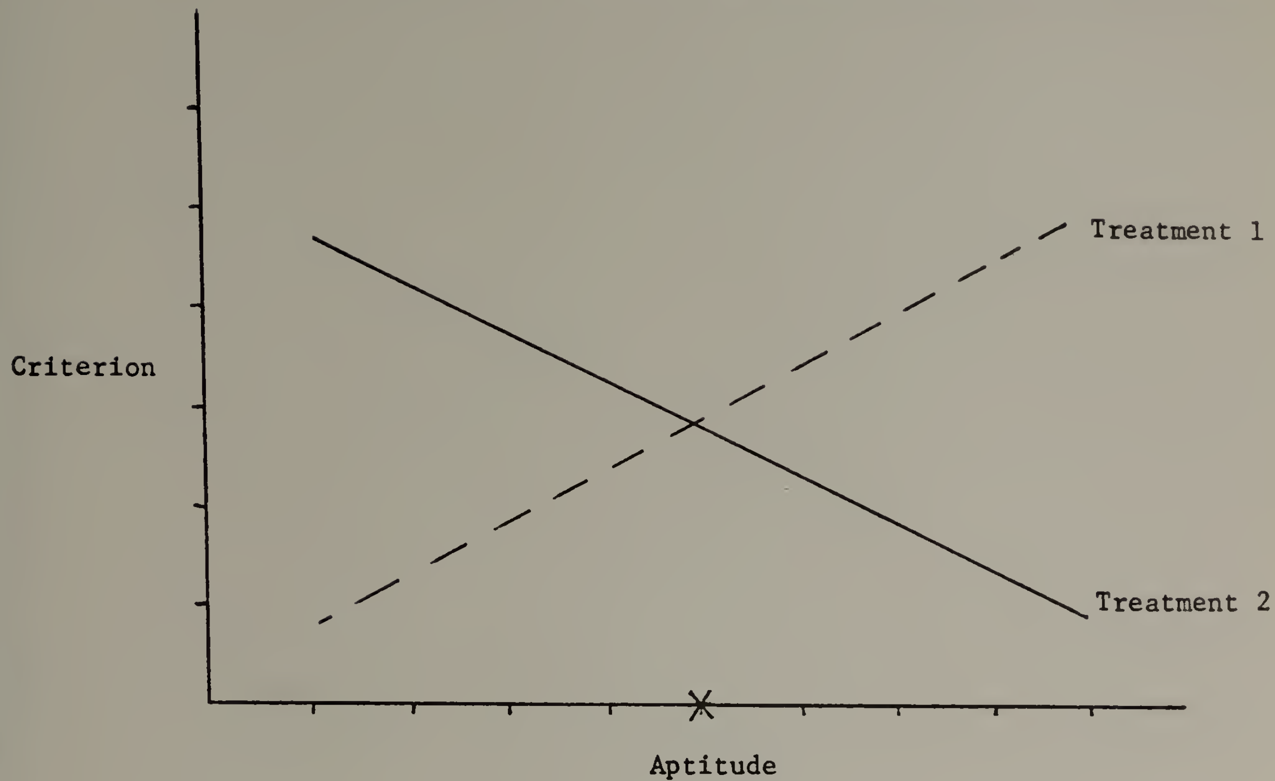


Figure 2.3.3 Illustration of a disordinal aptitude-treatment interaction.



a measure of scholastic success and if the treatment costs were comparable, one would assign all of the students to treatment one. This plot represents neither an ordinal nor a disordinal interaction (Berliner and Cahen, in press).

Berliner and Cahen (in press) state that the regression lines shown in Figure 2.3.2 represent an example of an ordinal interaction. Once again treatment one is superior to treatment two at all levels of the aptitude. If the cost of treatment two is the same as (or more than) that of treatment one, the ordinal interaction yields no more information to the decision maker than if no interaction was present (as in Figure 2.3.1). The rational decision would be to assign all students to the more effective treatment, i.e., treatment one.

A disordinal interaction is presented in figure 2.3.3. Berliner and Cahen (in press) define a disordinal interaction as the crossing of the regression lines within the observed range of the aptitude measure (note that in figure 2.3.2 the regression lines would have crossed outside of the range of the aptitude measure). The disordinal interaction shown in figure 2.3.3 indicates that students with aptitude scores above the score where the two lines cross would profit more if they were assigned to treatment one rather than to treatment two. On the other hand, students would profit more from treatment two than from treatment one if they had aptitude scores below the starred score. Students with aptitude scores equal to the starred score could be assigned to either treatment.

The standard procedure for ATI research is to use a test of the parallelism of regression lines. If the test of parallelism is rejected and if points within the range of observed scores on either side of the intersection score on the aptitude scale can be found where the predicted criterion scores are significantly different, the interaction is disordinal.

(The intersection score is the point on the aptitude score scale where both regression lines have the same predicted criterion score.) If the test of parallelism is rejected but only points on one side of the interaction score can be found where the regression lines differ significantly the interaction is merely ordinal.

The particular computer program used in our ATI research followed the analysis of covariance method outlined by Gulliksen and Wilks (1950). Tests for three statistical hypotheses are considered. The first is the hypothesis of the homogeneity of variance of the criterion scores about the regression line of criterion scores on aptitude scores for the collection of groups. If we reject this hypothesis, i.e., the homogeneity of variance of criterion scores about the regression lines cannot be assessed, we stop the hypothesis testing. If, however, we do not reject the initial hypothesis we then go on to test the hypothesis that the slopes of the regression lines are equal. Specifically the second hypothesis is that the slope of the regression of criterion scores on aptitude scores is the same for each treatment group. As was previously stated, from the point of view of ATI research this second hypothesis is the important one. It must be rejected for significant interactions to exist. To complete the cycle, the third hypothesis is a test to see if the regression lines are identical. Specifically the final hypothesis states that the intercept of the regression of criterion scores on aptitude scores is the same for each treatment group. This third hypothesis is only tested if we fail to reject the first two. From an analysis of covariance point of view, if all three hypotheses are accepted one may conclude that the groups are from the same basic population, or at least from populations having regression lines with the same slopes and the same intercepts, and have the same standard errors of estimate (Gulliksen and Wilks, 1950). From the ATI point of view, if the second hypothesis is not rejected no significant interaction exists.

It should be pointed out that throughout this discussion (and throughout the literature) the term "interaction" has been used rather loosely. In the one-factor analysis of covariance model there is no interaction term ($y_{ij} = \mu+\beta_j+\gamma(x_{ij}-\bar{x}_{..})+e_{ij}$). This model implies that for a given set of x's the y scores within each of r treatment groups lie on r parallel lines. Deviations among the treatment groups from this parallelism give us plots of the ordinal and disordinal interactions seen in figures 2.3.2 and 2.3.3.

If we now go ahead and block on the concomitant variable x and use these blocks as levels of a factor, the two-factor analysis of variance model emerges ($y_{ijk}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+e_{ijk}$). Only with this model do we have an actual interaction term [$(\alpha\beta)_{ij}$]. This term represents an interaction between the treatment levels and the levels of the previously blocked x variable. Yet plots of the cell means from an analysis of variance (one for each combination of treatment level and x variable level) yield data displays similar to figures 2.3.1 through 2.3.3. This point is essentially one of terminology, yet it may prove clarifying for some readers.

Once interactions have been found investigators want to determine regions of significant differences of outcomes for treatments as a function of aptitude level. The Johnson-Neyman technique allows us to determine these regions. For example, when dealing with one aptitude and two treatments, a Johnson-Neyman analysis might tell us that there is a region from (say) 6 to 12 on the aptitude scale, outside which there are significant differences in outcome. More specifically, persons above 12 perform significantly better on one treatment and persons below 6 perform significantly better on the other. Although the technique was originally developed for designs with just two groups and one criterion variable, it has been extended by Potthoff (1964) to designs with more than two groups and more than one criterion variable.

Evaluation of Instruction

(a) Individual Differences Model for Multidimensional Scaling

Tucker and Messick (1963) originated this method of multidimensional analysis. Their model assumes that estimates of interstimulus distances are available for each individual. Let

$x_{(jk)i}$ = an estimate of dissimilarity between stimuli j and k by individual i;

i,h = individuals 1, 2,..., N;

j,k = stimuli 1, 2, ..., n;

(jk) = stimulus-pairs 12, 13, 23, etc.; number of stimulus-pairs = n(n-1)/2.

Then X is a matrix of $x_{(jk)i}$, having n(n-1)/2 rows for the stimulus-pairs and N columns for the individuals.

The typical multidimensional scaling analysis involves an averaging of the $x_{(jk)i}$ values over the individuals to obtain a single number to represent the dissimilarity or distance between each pair of stimuli j and k.

Tucker and Messick, on the other hand, ask whether there is consistent covariation among individuals in these $x_{(jk)i}$ estimates by factoring X into its principal components. If only one factor is found to account for the consistent variance in X, then the appropriate average distance values may be analyzed to obtain a single representative multidimensional space. If, on the other hand, more than one factor is necessary to account for the variance in X, then more than one set of distance values will be obtained from the factor loadings to be subsequently analyzed by multidimensional scaling procedures.

In practice, however, matrix X cannot be factored directly because, in general, it is a non-symmetric matrix. One must compute an n(n-1)/2 by n(n-1)/2 matrix of cross products P (when the number of individuals, N, is greater than the number of stimulus-pairs, n(n-1)/2) where

$$P = XX'$$

Using a theorem developed by Eckart and Young (1936) for any arbitrary rank r we can produce a least squares approximation to X by

$$\hat{X}_r = U_r \Lambda_r W_r \qquad \text{where}$$

$\hat{X}_r$ = a least-squares, rank r approximation to X;

U_r = n(n-1)/2 by r section of an orthogonal matrix;

Λ_r = r by r diagonal matrix of latent roots;

W_r = r by N section of an orthogonal matrix.

The components U_r, Λ_r, and W_r are determined by factoring the matrix of cross-products, P, according to the method of principal components. Matrix P may be analyzed into principal components because, unlike X, it is a positive semi-definite matrix.

Thus $\hat{P}_r = \hat{X}_r \hat{X}_r' = U_r \Lambda_r^2 U_r'$ where Λ_r^2 is a diagonal matrix composed of the r largest latent roots of P, and U_r contains, as column vectors, the corresponding characteristic vectors of P.

The matrix W_r can be computed from the second equation ($\hat{X}_r = U_r \Lambda_r W_r$) by postmultiplying both sides by U_r' and Λ_r^{-1} respectively. Thus

$$W_r = \Lambda_r^{-1} U_r' X \qquad \text{where}$$

Λ_r has as its diagonal elements the square roots of the terms in Λ_r^2 which was obtained by the original factoring process.

At this point it might be well to mention how r or the number of factors is determined. Specifically the technique of "root staring" is used. As previously stated, the matrix P is decomposed into its roots and vectors. The square root of the diagonal matrix of roots is analyzed for a gap between the roots. The number of "large" roots before the gap is taken as the number of factors needed to account for the individual differences in X. When this technique is used one has to be aware that the first root will always be large because mean scores are not removed in forming matrix P (Tucker and Messick, 1963).

It might now appear that with our U_r matrix we have our r sets of distance values that can be subsequently analyzed by multidimensional scaling procedures. It might be said that we have recovered the sources of individual differences in the X matrix. Yet Roger Pennell (1971) has succinctly shown that the columns of U_r do not yield admissible sets of distances or sets which correlate highly with the original data sets. The columns of U_r do not recover the exact configurations of the stimuli.

To alleviate this problem we have to look at the rotated person-space and pick out representative clusters of persons from the unrotated person-space (W_r) (Cliff, 1968). We then get average loadings for each cluster and array each average as a column in matrix B. The B matrix is used to weight U_r. Thus

$$X_r^* = U_r \Lambda_r^{\frac{1}{2}} B^{-1} \qquad \text{where}$$

X_r^* represents judgments of distance made by idealized individuals. Pennell (1971) states that this averaging process is not subject to the same philosophical criticism as using a mean vector to represent the judgments of all the subjects. Here the components of individual differences have presumably been isolated, and, as well, groups of subjects that consistently respond alike. In some simulation studies conducted by Pennell (1971) he has shown that the columns of X_r^* correlate almost perfectly with the original distance vectors.

It is hoped that the preceeding discussion has brought out the essential features of the computer programs that this study has used to conduct a "points of view" analysis. The output of the "points of view" analysis or the column vectors of X_r^* are subsequently analyzed by multidimensional scaling procedures. These are described in the next section.

(a) Nonmetric Multidimensional Scaling

Shepard in 1962 and Kruskal in 1964 outlined the basic nonmetric multidimensional scaling model. This type of scaling is called non-metric (as opposed to metric) because only the ordinality of the data is preserved. In the metric models the proximity data are related to distances among points in a to-be-recovered coordinate space in a way that depends upon a function of some particular, specified form. As Shepard pointed out, in the nonmetric models the proximity data are related to distances among points in a to-be-recovered coordinate space by a function that is merely monotonic.

Given the vectors of the X_r^* matrix, each with n(n-1)/2 pairs of proximity measures, the computational problem of nonmetric multidimensional scaling is as follows. The model attempts to find a representation of the n stimuli as n points in a space of the smallest possible dimensionality such that the given proximity measures are, to a sufficient degree of approximation, monotonically related to the distances between the corresponding points in the spatial configuration or spatial arrangement. The solution of this problem is an iterative process designed to adjust the positions of the n points in an initial and usually arbitrary configuration until an explicitly defined measure of departure from the desired condition of monotonicity is minimized.

The following is a general outline of a computer program for nonmetric multidimensional scaling developed by J.B. Kruskal (1964).

The program starts by generating a coordinate, x_{it}, for each point i, on each of the arbitrary but orthogonal axes, t, by some random, arbitrary, or rational method.

The position and over-all size of the configuration (configuration meaning the pattern or structure produced by the arrangement of the points) is adjusted with respect to the coordinate system so that, e.g., $\Sigma x_{it}=0$ and $\Sigma x_{it}^2=0$. This process normalizes the configuration.

Next the distances among all n points are computed; e.g., using Euclidean metric, distances are given by

$$d_{ij} = \sqrt{\Sigma_t (x_{it}-x_{jt})^2}.$$

At this point the input data (a column from X_r^*) comes into the scheme of things. The explictly defined measure of departure from the desired monotonic relation between the given proximity data, δ_{ij}, and the computed interpoint distances, d_{ij}, is found. Kruskal calls this goodness of fit measure the stress.

Then the direction and relative distance in which each of n points should be moved to produce an improvement in the goodness of fit is conducted using, amongst other things, the method of steepest descent.

The achievement of a minimum is tested for. (A minimum signifies that further movement of the points will not lead to a further decrease in the stress.) If a minimum is not achieved, each of the n points is moved in the direction of and for the relative distance indicated by the current negative gradient. (The over-all magnitude of the adjustment is determined by a step-size.) Since a minimum has not yet been achieved the program goes back and re-normalizes the configuration and proceeds with the next iteration.

If a minimum is achieved and the goodness of fit measure (stress) is sufficiently small, the procedure is stopped. If the stress value is not sufficiently small the minimum may be merely local. (That is, in addition to the local minimum there is an overall minimum which results in the smallest possible stress value for that particular configuration.) If this is the case the program goes back and starts over by constructing a new configuration or arrangement of the points in space.

The particular nonmetric multidimensional scaling program that was used in this study was developed by Forrest Young (1968) and is called TORSCA-9. It uses the same essential logic as that put forth by the Kruskal program with but one major exception. The one major exception being that a different criterion is utilized to evaluate the goodness of the solution. (The Kruskal and Young programs also differ in their methods for establishing initial configurations.)

At the risk of being repetitive, let us suppose that there are n stimuli and that we have experimental values of dissimilarity between them (δ_{ij}).

For a configuration of points $(x_1, x_2, \ldots, x_n)$ in a t-dimensional space, with interpoint distances d_{ij}, Kruskal defines the "stress" of the configuration by

$$S = \sqrt{\frac{\Sigma(d_{ij}-\hat{d}_{ij})^2}{\Sigma d^2_{ij}}} .$$

The $\hat{d}_{ij}$ values are those numbers which minimize the "stress" subject to the constraint that the $\hat{d}_{ij}$ values have the same rank order as the dissimilarity values (δ_{ij}). The best solution is that configuration of points which minimizes the "stress" function in as low a dimensional space as possible subject to the constraint that the configuration's disparities ($\hat{d}_{ij}$ values) have the same rank order as the dissimilarity values (δ_{ij}) (Kruskal, 1964).

In place of the "stress" function, Young defines a function α where

$$\alpha = \frac{1}{2} + \frac{1}{2} \frac{\Sigma \hat{d}_{ij} d_{ij}}{\sqrt{(\Sigma \hat{d}_{ij})^2 (\Sigma d_{ij})^2}} .$$

(This function attempts to evaluate the degree to which a plot of d_{ij} versus $\hat{d}_{ij}$ can be fitted by an equation of the form y=ax. The y=ax equation is used because the relation between d_{ij} and $\hat{d}_{ij}$ is linear and both variables are referred to a common origin.) The best solution in Young's program is that configuration of points which maximize the α function in as low a dimensional space as possible subject to the constraint that the configuration's disparities ($\hat{d}_{ij}$ values) have the same rank order as the dissimilarity values (δ_{ij}). The same reasoning is present; the criterion for a solution differs slightly.

(b) Latent Partition Analysis

The original idea for latent partition analysis can be credited to Wiley (1967). The model attempts to relate the manifest categorizations of sorters to a hypothetical underlying categorization of the items.

The fundamental equation of the model is as follows:

$$S = \Phi'\Omega\Phi + \Delta^2 \quad \text{where,}$$

given that j,k = items 1, 2, ..., K;

m,n = manifest categories 1, 2, ..., M;

u,v = latent categories 1, 2, ..., L;

S is a K by K matrix with the (j,k)th entry of S being the probability that the average judge puts item j in the same manifest category as item k. Φ is a L by K matrix with the (u,j)th entry of Φ being 1 or 0 accordingly as item j is in latent category u or not. Ω is a L by L matrix with the (u,v)th entry of Ω being the probability that the average judge puts an item from latent category u together with one from latent category v into the same manifest category. Δ^2 is a K by K diagonal matrix with the (j,j)th entry of Δ^2 being the diversity of item j. The diversity of item j is the probability of item j being included in two different manifest categories under independent partitioning.

The iterative solution proposed by Wiley focuses on the estimation of Δ^2. The solution yields estimates of Φ and Ω, or $\hat{\Phi}$ and $\hat{\Omega}$. The $\hat{\Phi}$ matrix gives, in effect, the loadings of the items on each of the latent factors. The $\hat{\Omega}$ matrix gives the probabilities that items from various pairs of latent categories will appear in the same manifest category. The diagonal entries of this matrix can be considered to be the probabilities of latent category confusions. These entries provide an index of category cohesiveness reflecting the amount of agreement on the placement of the items into a given category. Thus both the $\hat{\Phi}$ and $\hat{\Omega}$ matrices give us an interpretable solution to the latent partition problem.

The particular computer program that was used in this study was developed from Wiley's model. Specifically the program is able to accept raw data and use the raw data to build up the joint proportions matrix. Hotelling's iterative procedure is used to decompose the joint proportions (S) matrix. The decomposition is performed on successive reproductions of the S-matrix until the diagonal elements of the reproduced S-matrix stabilize. The eigenvectors are then submitted to a raw quartimax rotation and scaled by the sums of the columns. The program outputs S, $\hat{\Phi}$, $1-\hat{\Delta}^2$, and $\hat{\Omega}$ (Harasym and Precht, 1971).

Chapter III

Results

3.1 Results of the ATI Study

The summary statistics for each of the four instructional treatment groups are presented in Table 3.1.1. The Module Posttest and the Delayed Module Posttest were used as criteria in the search for interactions since they seemed to be the most appropriate. Most of the remaining variables reported in the Table 3.1.1 were chosen as aptitudes. It is clear from the table that the four groups did not differ significantly on the various aptitude variables.

The intercorrelations of the predictor and criterion variables for each of the four instructional treatment groups are shown in Tables 3.1.2-3.1.5. Correlations significantly different from zero (at the .05 level) were underlined in the tables.

The previously described analysis of covariance method of Gulliksen and Wilks (1950) was used to search for interactions. Initially the Module Posttest was used as the criterion and each of the 24 predictor variables was used in turn as the independent variable. The results of these analyses are shown in Table 3.1.6.

Using the same predictor variables, the analyses were repeated using the Delayed Module Posttest as the criterion. These results are shown in Table 3.1.7.

Table 3.1.1

Summary Statistics of the Four Instructional Treatment Groups on the Variables Considered in the Study

Variable	Treatment Group							
	G1 (N=49)		G2 (N=66)		G3 (N=75)		G4 (N=72)	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
Lorge-Thorndike IQ Test								
Verbal IQ	107.8	12.5	113.9	13.9	115.6	15.4	112.3	12.6
Non-Verbal IQ	115.9	15.0	118.9	17.4	120.0	14.6	120.0	11.9
SRA Achievement Series								
Total Science Score	5.9	1.6	6.5	1.5	6.7	1.4	6.2	1.5
Capitalization & Punctuation Score	5.6	1.8	6.2	1.7	6.3	1.7	6.6	1.6
Grammatical Usage Score	6.1	1.4	6.1	1.5	6.5	1.6	5.8	1.5
Spelling Score	5.9	1.9	5.8	1.8	6.1	1.8	6.2	1.5
Total Language Arts Score	5.9	1.7	6.0	1.6	6.4	1.5	6.4	1.5
Reasoning Score	5.9	1.5	6.3	1.6	6.2	1.6	6.0	1.7
Concepts Score	6.3	1.5	6.9	1.9	6.9	1.6	6.9	1.4
Computation Score	6.1	1.5	6.4	1.6	6.3	1.4	6.7	1.2
Total Arithmetic Score	6.4	1.3	6.8	1.6	6.7	1.4	6.8	1.3
Comprehension Score	6.2	1.3	6.7	1.3	6.8	1.4	6.5	1.5
Vocabulary Score	6.3	1.5	7.0	1.4	7.1	1.4	6.9	1.3
Total Reading Score	6.2	1.3	7.0	1.3	7.1	1.5	6.8	1.5
Total Battery Composite Score	6.4	1.4	6.9	1.5	7.0	1.4	6.8	1.3
References Score	6.4	1.4	6.4	1.3	6.9	1.4	6.7	1.2
Charts Score	6.0	1.6	6.1	1.7	6.4	1.6	6.4	1.4
Total Work-Study Skills Score	6.3	1.3	6.4	1.3	6.6	1.5	6.6	1.2
Survey of Study Habits & Attitudes								
Delay Avoidance Score	20.9	6.9	20.5	9.2	19.1	8.4	20.3	9.7
Work Methods Score	23.3	8.1	22.5	8.3	22.6	8.4	24.5	8.6
Study Habits Score	44.2	13.7	43.0	16.1	41.6	15.4	44.8	17.4
Teacher Approval Score	27.5	10.7	27.3	8.8	26.0	8.6	27.0	11.6
Education Acceptance Score	25.1	8.4	25.3	8.3	24.2	7.5	25.6	9.8
Study Attitudes Score	52.7	18.4	52.6	15.9	50.2	15.2	52.5	20.7
Study Orientation Score	96.9	30.4	95.0	30.0	91.8	28.5	95.7	37.0
JIM Scale Score	132.5	19.7	129.4	18.2	126.6	21.3	131.2	21.3
Intellectual Achievement Responsibility Scale								
Responsibility for Success Score	12.6	3.1	12.7	3.0	12.8	2.1	13.0	2.8
Responsibility for Failures Score	11.8	3.1	11.8	2.8	11.8	2.7	12.1	2.8
Total Responsibility Score	24.4	5.5	24.5	5.2	24.6	4.0	25.0	4.5
Children's Manifest Anxiety Scale								
Anxiety Scale Score	16.6	7.7	17.5	7.8	16.8	6.4	16.9	7.0
Lie Scale Score	1.2	1.4	1.3	1.5	1.2	1.6	1.4	1.5
Test Anxiety Scale for Children Score	10.2	6.5	9.4	5.8	8.7	5.4	9.9	6.0
School Anxiety Scale Score	18.9	4.7	19.2	4.7	18.7	4.1	20.0	4.1

Table 3.1.1 (continued)

Summary Statistics of the Four Instructional Treatment Groups on the Variables Considered in the Study

Variable	Treatment Group							
	G1 (N=49)		G2 (N=66)		G3 (N=75)		G4 (N=72)	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Mathematics Test Score	12.0	4.2	13.0	5.7	12.7	5.2	13.4	5.5
Letter Sets Score	21.7	5.4	20.5	6.0	21.7	4.7	22.8	5.4
Module Pretest Score	10.3	5.4	9.8	6.9	9.4	5.8	10.5	6.5
Module Posttest Score	30.5	7.2	26.8	9.6	27.3	8.5	27.8	8.1
Delayed Module Posttest Score	24.5	8.3	21.7	9.3	21.5	8.5	22.6	8.2
Student Attitude Questionnaire Concept Scores								
Teachers	1.2	.9	1.1	0.9	1.0	1.0	1.3	1.4
Science	1.0	1.1	.8	1.0	0.8	0.9	1.0	1.2
Student Freedom in Class	1.1	1.2	1.0	1.0	1.2	1.1	1.2	1.2
School	1.2	1.2	0.9	1.0	0.9	1.1	0.9	1.2
Testing	0.2	0.9	0.0	0.8	0.3	1.0	0.2	1.1
Individualized Instruction	1.0	1.1	0.6	1.1	0.8	1.0	1.1	1.3
Resource Center	1.1	1.3	1.0	1.2	1.1	1.2	1.2	1.4
Enrichment Program	1.1	1.2	0.8	1.4	1.0	1.1	1.0	1.3
Attitude Towards School (average of the above eight concept scores)	1.0	0.8	0.8	0.7	0.9	0.7	1.0	1.0
Final Grade in Science	86.2	6.2	86.1	8.1	88.6	6.1	86.2	6.8

Table 3.1.2

Intercorrelations Among the ATI Variables for Instructional Treatment Group 1[1,2]

Group 1 (N=49)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1. Verbal IQ		72	72	71	53	53	76	66	24	50	41	08	13	10	25	36	-12	01	-53	-29	61	13	19	34	30	-18
2. Non-Verbal IQ			63	61	75	58	73	81	15	30	25	-15	-08	-12	04	16	-19	29	-26	-10	66	31	12	55	63	-33
3. SRA Science Score				54	64	47	70	54	07	44	28	12	20	16	23	26	-06	-13	-51	-23	36	07	02	63	41	-27
4. SRA Total Language Arts Score					59	58	79	62	06	55	33	12	18	15	25	47	15	-16	-48	-36	45	55	01	34	20	18
5. SRA Total Arithmetic Score						73	88	79	05	51	31	03	13	08	19	33	22	-05	-52	-36	62	25	-11	47	51	00
6. SRA Total Reading Score							82	78	10	48	31	04	16	10	21	35	01	12	-47	-50	47	14	07	22	28	-22
7. SRA Total Battery Composite Score								81	10	63	40	10	25	18	30	46	15	-17	-67	-54	56	33	-02	41	38	-02
8. SRA Total Work-Study Skills Score									13	51	35	14	21	18	27	31	05	03	-51	-40	59	36	-02	35	39	-11
9. SSHA Delay Avoidance Score										65	89	68	64	69	82	19	04	-06	-11	-08	01	04	-20	27	39	08
10. SSHA Work Methods Score											92	71	71	74	86	56	31	-24	-32	-27	14	19	-03	38	36	07
11. SSHA Study Habits Score												76	75	78	93	43	21	-17	-25	-20	09	13	-11	36	40	08
12. SSHA Teacher Approval Score													86	97	93	33	28	-27	-12	-13	-04	09	-15	24	41	36
13. SSHA Education Acceptance Score														95	92	41	39	-16	-13	-11	00	13	-22	26	32	30
14. SSHA Study Attitudes Score															96	38	34	-23	-13	-13	-02	11	-19	26	38	35
15. SSHA Study Orientation Score																42	30	-22	-19	-17	03	13	-16	32	41	25
16. Junior Index of Motivation Score																	41	-04	-27	-28	16	23	-05	29	20	03
17. Intellectual Achievement Responsibility Scale Total Score																		-14	-23	-08	20	01	-40	03	-06	28
18. Children's Manifest Anxiety Scale Score																			44	18	07	-24	01	11	03	-28
19. Test Anxiety Scale for Children Score																				69	-39	-31	08	11	-13	04
20. School Anxiety Scale Score																					-34	-30	-19	-21	-22	06
21. Mathematics Test Score																						-34	-05	33	49	-19
22. Letter Sets Score																							-20	29	24	15
23. Module Pretest Score																								10	05	-03
24. Module Posttest Score																									67	-04
25. Delayed Module Posttest Score																										01
26. Attitude Towards School Score																										

[1]Correlations significantly different from zero at the .05 level are underlined.

[2]Decimals have been omitted.

Table 3.1.3

Intercorrelations Among the ATI Variables for Instructional Treatment Group 2[1,2]

Group 2 (N=66)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1. Verbal IQ		79	47	55	63	65	70	64	15	35	27	35	40	40	36	45	14	09	-16	-09	44	54	-07	57	51	14
2. Non-Verbal IQ			62	38	67	58	70	72	10	20	16	31	33	34	26	52	13	04	-07	07	55	71	05	61	57	34
3. SRA Science Score				17	68	57	67	62	12	08	11	16	06	12	08	44	-02	18	-21	-08	61	53	01	66	74	25
4. SRA Total Language Arts Score					46	59	72	50	23	44	35	48	33	43	44	09	24	20	-11	-14	11	40	-13	44	25	-17
5. SRA Total Arithmetic Score						78	89	83	11	14	13	34	11	25	18	51	11	10	02	06	62	68	-06	71	67	16
6. SRA Total Reading Score							90	73	-02	20	09	32	14	25	18	34	11	11	-24	-07	46	50	-24	62	48	20
7. SRA Total Battery Composite Score								84	11	21	16	39	15	29	24	41	10	07	-16	-02	49	62	-09	72	55	14
8. SRA Total Work-Study Skills Score									-03	10	03	18	04	11	06	40	05	15	04	10	50	73	08	61	61	14
9. SSHA Delay Avoidance Score										70	93	41	69	60	81	16	18	-15	-08	-09	14	02	-01	21	13	28
10. SSHA Work Methods Score											91	55	66	66	83	23	23	-25	-32	-33	15	18	-17	21	24	28
11. SSHA Study Habits Score												52	73	68	89	21	22	-21	-21	-22	16	10	-09	22	19	30
12. SSHA Teacher Approval Score													70	93	79	48	19	-17	-30	-25	03	23	-13	35	38	26
13. SSHA Education Acceptance Score														92	91	30	25	-11	-19	-14	-02	10	-13	30	26	22
14. SSHA Study Attitudes Score															92	43	24	-15	-27	-21	01	18	-14	35	35	26
15. SSHA Study Orientation Score																33	24	-20	-30	-26	05	14	-11	28	25	25
16. Junior Index of Motivation Score																	33	-18	-14	00	32	38	-02	32	32	28
17. Intellectual Achievement Responsibility Scale Total Score																		00	10	13	25	19	-14	11	01	03
18. Children's Manifest Anxiety Scale Score																			60	36	00	05	-06	08	09	-21
19. Test Anxiety Scale For Children Score																				63	-02	-03	05	-12	-06	-07
20. School Anxiety Scale Score																					00	01	12	-10	-04	14
21. Mathematics Test Score																						62	-03	56	52	19
22. Letter Sets Score																							-03	70	67	26
23. Module Pretest Score																								10	22	-09
24. Module Posttest Score																									87	25
25. Delayed Module Posttest Score																										27
26. Attitude Towards School Score																										

[1] Correlations significantly different from zero at the .05 level are underlined.

[2] Decimals have been omitted.

Table 3.1.4

Intercorrelations Among the ATI Variables for Instructional Treatment Group 3[1,2]

Group 3 (N=75)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1. Verbal IQ		69	70	67	62	70	75	67	03	37	22	31	28	31	29	46	04	-03	-36	-36	55	41	31	47	53	08
2. Non-Verbal IQ			56	57	66	54	69	62	03	36	22	30	20	27	27	38	-01	07	-24	-38	62	49	22	62	54	13
3. SRA Science Score				50	56	67	68	63	17	33	27	25	27	27	29	46	-09	-29	-28	-18	47	13	50	44	42	-15
4. SRA Total Language Arts Score					65	65	78	75	-05	18	06	16	01	09	08	43	-15	25	-06	-31	44	32	15	30	24	-01
5. SRA Total Arithmetic Score						69	88	79	20	45	34	28	09	21	30	48	-12	00	-33	-35	77	30	43	59	60	21
6. SRA Total Reading Score							87	80	15	23	20	23	04	15	19	41	-11	-12	-11	-24	47	10	36	43	46	-17
7. SRA Total Battery Composite Score								89	12	31	23	20	01	11	19	44	-17	09	-12	-28	59	20	33	55	59	-02
8. SRA Total Work-Study Skills Score									13	24	20	34	13	25	25	49	00	06	00	-17	59	34	37	62	62	00
9. SSHA Delay Avoidance Score										67	92	60	70	69	86	40	17	-26	-06	-07	09	17	-02	24	22	22
10. SSHA Work Methods Score											91	59	67	66	85	41	18	-47	-37	-27	35	37	10	37	38	17
11. SSHA Study Habits Score												65	75	74	93	44	19	-40	-23	-18	24	30	05	33	34	21
12. SSHA Teacher Approval Score													79	95	86	60	36	-26	-12	-16	33	28	09	34	47	22
13. SSHA Education Acceptance Score														94	91	51	46	-39	-15	-16	29	29	12	24	45	15
14. SSHA Study Attitudes Score															93	59	43	-34	-14	-17	33	31	11	31	48	20
15. SSHA Study Orientation Score																55	33	-39	-20	-19	30	32	08	34	44	22
16. Junior Index of Motivation Score																	18	-12	-20	-19	41	31	21	42	39	08
17. Intellectual Achievement Responsibility Scale Total Score																		-06	02	-11	00	-04	-08	-01	28	07
18. Children's Manifest Anxiety Scale Score																			54	29	-08	01	-13	-03	-03	13
19. Test Anxiety Scale for Children Score																				61	-28	-03	-19	-17	02	-10
20. School Anxiety Scale Score																					-21	-07	02	-27	-06	-01
21. Mathematics Test Score																						46	27	63	63	10
22. Letter Sets Score																							03	58	45	27
23. Module Pretest Score																								24	41	06
24. Module Posttest Score																									79	13
25. Delayed Module Posttest Score																										07
26. Attitude Towards School Score																										

[1]Correlations significantly different from zero at the .05 level are underlined.

[2]Decimals have been omitted.

Table 3.1.5

Intercorrelations Among the ATI Variables for Instructional Treatment Group 4[1,2]

Group 4 (N=72)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1. Verbal IQ		69	40	55	60	64	70	48	06	26	16	14	10	13	19	35	42	-23	-26	-24	62	32	-04	48	32	-23
2. Non-Verbal IQ			31	38	57	38	58	40	03	05	04	-05	03	-01	03	32	31	-14	-28	-14	70	61	05	55	53	-02
3. SRA Science Score				32	59	58	72	47	00	00	00	-02	09	03	02	12	08	-34	-47	-36	41	05	23	55	38	-04
4. SRA Total Language Arts Score					42	55	72	33	10	27	19	17	13	15	18	19	52	02	-12	-18	18	18	03	31	27	-15
5. SRA Total Arithmetic Score						58	78	59	-17	-07	-13	-07	-06	-07	-10	04	16	-02	-32	-27	65	24	13	56	51	-28
6. SRA Total Reading Score							81	62	02	24	13	23	21	23	19	22	39	-21	-37	-32	34	23	05	44	32	-26
7. SRA Total Battery Composite Score								62	13	26	19	19	24	22	21	25	37	-16	-38	-35	59	30	14	62	48	-17
8. SRA Total Work-Study Skills Score									-09	02	-04	08	04	07	02	11	21	00	-28	-30	54	18	09	43	49	07
9. SSHA Delay Avoidance Score										80	95	74	78	79	87	46	27	-27	-24	-12	01	12	22	21	18	52
10. SSHA Work Methods Score											94	82	84	86	89	46	29	-25	-39	-30	-01	07	26	17	20	37
11. SSHA Study Habits Score												82	85	87	92	48	30	-28	-33	-22	00	10	25	20	20	48
12. SSHA Teacher Approval Score													86	97	85	55	33	-18	-35	-30	-06	11	21	11	23	41
13. SSHA Education Acceptance Score														96	88	41	29	-15	-34	-23	-05	06	26	16	22	35
14. SSHA Study Attitudes Score															90	50	32	-17	-36	-28	-06	09	24	14	23	40
15. SSHA Study Orientation Score																47	31	-23	-29	-19	-01	07	24	20	24	34
16. Junior Index of Motivation Score																	47	-26	-34	-17	30	42	-10	38	28	26
17. Intellectual Achievement Responsibility Scale Total Score																		-14	-15	00	15	24	-18	25	21	08
18. Children's Manifest Anxiety Scale Score																			49	35	-10	-01	-10	-19	-15	-26
19. Test Anxiety Scale for Children Score																				66	-29	-28	-22	-48	-49	-10
20. School Anxiety Scale Score																					-19	-09	-13	-16	-28	-06
21. Mathematics Test Score																						53	06	51	57	05
22. Letter Sets Score																							01	56	47	17
23. Module Pretest Score																								28	45	26
24. Module Posttest Score																									71	11
25. Delayed Module Posttest Score																										21
26. Attitude Towards School Score																										

[1] Correlations significantly different from zero at the .05 level are underlined.

[2] Decimals have been omitted.

Table 3.1.6

Analysis of Covariance Results for the Four Instructional Treatment Groups with the Module Posttest as the Criterion

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
Verbal IQ	0.2	3	0.6	3,169	5.8**	3,174
Non-Verbal IQ	1.1	3	0.1	3,169	4.9**	3,174
SRA Total Science	1.0	3	0.9	3,150	6.7**	3,155
SRA Total Language Arts	0.7	3	0.4	3,150	3.4*	3,155
SRA Total Arithmetic	0.2	3	0.4	3,150	5.3**	3,155
SRA Total Reading	0.3	3	1.7	3,149	5.4**	3,154
SRA Total Battery Composite	0.8	3	1.3	3,149	6.6**	3,154
SRA Total Work-Study Skills	1.1	3	0.8	3,150	4.2**	3,155
SSHA Delay Avoidance	6.4	3	0.1	3,203	1.4	3,208

* - significant at the .05 level.
** - significant at the .01 level.

Table 3.1.6 (continued)

Analysis of Covariance Results for the Four Instructional Treatment Groups with the Module Posttest as the Criterion

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
SSHA Work Methods	6.9	3	0.6	3,203	1.6	3,208
SSHA Study Habits	6.8	3	0.5	3,203	1.5	3,208
SSHA Teacher Approval	4.8	3	1.4	3,203	1.6	3,208
SSHA Education Acceptance	5.4	3	0.5	3,208	1.6	3,208
SSHA Study Attitudes	4.9	3	1.1	3,203	1.6	3,208
SSHA Study Orientation	6.0	3	0.7	3,203	1.5	3,208
Junior Index of Motivation	4.9	3	0.3	3,228	2.0	3,233
Intellectual Achievement Responsibility Scale	4.6	3	0.9	3,228	2.4	3,233

Table 3.1.6 (continued)

Analysis of Covariance Results for the Four Instructional Treatment Groups with the Module Posttest as the Criterion

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
Children's Manifest Anxiety Scale	4.0	3	0.9	3,237	1.8	3,242
Test Anxiety Scale for Children	6.3	3	1.8	3,237	2.4	3,242
School Anxiety Scale	4.2	3	0.4	3,237	1.9	3,242
Mathematics Test	1.6	3	1.3	3,227	3.3*	3,232
Letter Sets Test	0.2	3	3.5*	3,228	3.4*	3,233
Module Pretest	5.3	3	0.5	3,231	2.1	3,236
Attitude Towards School	3.2	3	0.8	3,210	1.6	3,215

* - significant at the .05 level.

Table 3.1.7

Analysis of Covariance Results for the Four Instructional Treatment Groups with the Delayed Module Posttest as the Criterion

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
Verbal IQ	1.1	3	0.2	3,124	3.4*	3,129
Non-Verbal IQ	0.2	3	0.7	3,124	2.6	3,129
SRA Total Science	3.6	3	1.5	3,112	3.5*	3,117
SRA Total Language Arts	0.7	3	0.1	3,112	2.0	3,117
SRA Total Arithmetic	1.9	3	0.1	3,112	3.7*	3,117
SRA Total Reading	2.5	3	0.3	3,111	3.2*	3,116
SRA Total Battery Composite	3.5	3	0.2	3,111	4.1**	3,116
SRA Total Work-Study Skills	2.7	3	0.1	3,112	3.3*	3,117
SSHA Delay Avoidance	0.7	3	0.7	3,149	0.7	3,154

* - significant at the .05 level.
** - significant at the .01 level.

Table 3.1.7 (continued)

Analysis of Covariance Results for the Four Instructional Treatment Groups with the Delayed Module Posttest as the Criterion

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
SSHA Work Methods	0.5	3	0.5	3,149	0.9	3,154
SSHA Study Habits	0.7	3	0.8	3,149	0.8	3,154
SSHA Teacher Approval	0.6	3	1.0	3,149	0.9	3,154
SSHA Education Acceptance	0.6	3	0.9	3,149	1.1	3,154
SSHA Study Attitudes	0.6	3	1.1	3,149	1.0	3,154
SSHA Study Orientation	0.8	3	0.9	3,149	0.8	3,154
Junior Index of Motivation	0.5	3	0.4	3,173	1.3	3,178
Intellectual Achievement Responsibility Scale	0.7	3	1.3	3,173	1.5	3,178

Table 3.1.7 (continued)

Analysis of Covariance Results for the Four Instructional Treatment Groups with the Delayed Module Posttest as the Criterion

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedeom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
Children's Manifest Anxiety Scale	0.4	3	0.4	3,177	1.0	3,182
Test Anxiety Scale for Children	2.0	3	2.1	3,177	1.4	3,182
School Anxiety Scale	0.7	3	0.5	3,177	1.0	3,182
Mathematics Test	1.3	3	0.2	3,171	2.6	3,176
Letter Sets Test	1.5	3	1.5	3,170	1.9	3,175
Module Pretest	1.2	3	1.3	3,173	1.3	3,178
Attitude Towards School	1.0	3	0.6	3,158	0.7	3,163

Questionnaire evidence (Table 3.4.1 and 3.4.2) seemed to indicate that instructional treatment 2 met with some technical difficulties. Accepting the possibility that this treatment may have been ineffective, the analyses were repeated excluding this treatment group. The results of these analyses are shown in Tables 3.1.6a and 3.1.7a.

In each of the four tables the second hypothesis is the important one. The F statistic tests the parallelism of slopes from regressing the criterion variable on each of the predictor variables for the various instructional treatment groups. As can be seen in the tables, the F statistic is significant at the .05 level with three of the predictor variables. The test of the parallelism of slopes was rejected when the Letter Sets Test was the predictor and the Module Posttest was the criterion and all four instructional treatment groups were considered. This result also held up when instructional treatment group 2 was removed (Table 3.1.6a). The third rejection of the parallelism of slopes hypothesis occurred when the Delayed Module Posttest was used as the criterion, the Test Anxiety Scale for Children was used as the predictor, and instructional treatment group 2 was removed (Table 3.1.7a).

At this point, it should be noted that 96 ATIs were run (although only 24 were completely independent) and since only three were detected there is a distinct possibility that they are simply chance results and hence worthy of little attention. Nevertheless the analysis was continued on these "significant" ATIs but the results should be interpreted with extreme caution.

To determine which pairs of treatment groups contributed to the rejection of the equal slopes of regression hypothesis each pair of treatment groups was subjected to a test of parallelism of slopes. In addition,

Table 3.1.6a

Analysis of Covariance Results for the Three Instructional Treatment Groups with the Module Posttest as the Criterion (excluding treatment group 2)

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
Verbal IQ	0.2	2	0.3	2,126	8.5**	2,130
Non-Verbal IQ	0.7	2	0.1	2,126	7.6**	2,130
SRA Total Science	1.0	2	0.2	2,112	9.4**	2,116
SRA Total Language Arts	0.1	2	0.0	2.112	5.1*	2,116
SRA Total Arithmetic	0.1	2	0.2	2,112	7.6**	2,116
SRA Total Reading	0.3	2	0.3	2,111	7.6**	2,115
SRA Total Battery Composite	0.8	2	0.9	2,111	9.4**	2,115
SRA Total Work-Study Skills	0.8	2	0.3	2,112	6.3*	2,116
SSHA Delay Avoidance	2.2	2	0.2	2,154	1.6	2,158

* - significant at the .05 level.
** - significant at the .01 level.

Table 3.1.6a (continued)

Analysis of Covariance Results for the Three Instructional Treatment Groups with the Module Posttest as the Criterion (excluding treatment group 2)

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
SSHA Work Methods	1.8	2	1.0	2,154	2.1	2,158
SSHA Study Habits	2.0	2	0.8	2,154	1.8	2,158
SSHA Teacher Approval	1.3	2	1.7	2,154	1.8	2,158
SSHA Education Acceptance	2.0	2	0.5	2,154	1.9	2,158
SSHA Study Attitudes	1.6	2	1.2	2,154	1.9	2,158
SSHA Study Orientation	1.7	2	1.1	2,154	1.8	2,158
Junior Index of Motivation	1.0	2	0.5	2,173	2.3	2,177
Intellectual Achievement Responsibility Scale	1.9	2	1.3	2,173	2.8	2,177

Table 3.1.6a (continued)

Analysis of Covariance Results for the Three Instructional Treatment Groups with the Module Posttest as the Criterion (excluding treatment group 2)

	Hypothesis					
Aptitude	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
Children's Manifest Anxiety Scale	1.6	2	1.0	2,180	2.5	2,163
Test Anxiety Scale for Children	2.5	2	2.7	2,180	3.5*	2,163
School Anxiety Scale	1.3	2	0.4	2,180	2.6	2,163
Mathematics Test	0.2	2	1.4	2,174	3.1*	2,178
Letter Sets Test	0.1	2	3.2*	2,174	4.8*	2,178
Module Pretest	1.9	2	0.4	2,175	3.0	2,179
Attitude Towards School	1.2	2	0.6	2,159	2.4	2,163

* - significant at the .05 level.

Table 3.1.7a

Analysis of Covariance Results for the Three Instructional Treatment Groups with the Delayed Module Posttest as the Criterion (excluding treatment group 2)

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
Verbal IQ	1.1	2	0.1	2,93	4.4	2,97
Non-Verbal IQ	0.1	2	0.8	2,93	3.3*	2,97
SRA Total Science	0.6	2	0.0	2,84	4.1*	2,88
SRA Total Language Arts	0.7	2	0.1	2,84	2.5	2,88
SRA Total Arithmetic	1.4	2	0.1	2,84	4.6*	2,88
SRA Total Reading	2.5	2	0.1	2,83	4.2*	2,87
SRA Total Battery Composite	3.5	2	0.2	2,83	5.6**	2,87
SRA Total Work-Study Skills	2.6	2	0.3	2,84	3.7*	2,88
SSHA Delay Avoidance	0.4	2	1.0	2,113	1.0	2,117

* - significant at the .05 level.
** - significant at the .01 level.

Table 3.1.7a (continued)

Analysis of Covariance Results for the Three Instructional Treatment Groups with the Delayed Module Posttest as the Criterion (excluding treatment group 2)

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
SSHA Work Methods	0.2	2	0.7	2,113	1.5	2,117
SSHA Study Habits	0.3	2	1.1	2,113	1.2	2,117
SSHA Teacher Approval	0.3	2	1.5	2,113	1.5	2,117
SSHA Education Acceptance	0.2	2	1.3	2,113	1.7	2,117
SSHA Study Attitudes	0.4	2	1.6	2,113	1.6	2,117
SSHA Study Orientation	0.3	2	1.4	2,113	1.3	2,117
Junior Index of Motivation	0.1	2	0.7	2,132	1.9	2,136
Intellectual Achievement Responsibility Scale	0.0	2	2.0	2,132	2.3	2,136

Table 3.1.7a (continued)

Analysis of Covariance Results for the Three Instructional Treatment Groups with the Delayed Module Posttest as the Criterion (excluding treatment group 2)

Aptitude	Hypothesis					
	Homogeneity of Variance (χ^2 Statistic)	Degrees of Freedom	Equal Slopes of Regression (F Statistic)	Degrees of Freedom	Equal Intercepts of Regression (F Statistic)	Degrees of Freedom
Children's Manifest Anxiety Scale	0.1	2	0.3	2,135	1.6	2,139
Test Anxiety Scale for Children	1.2	2	3.2*	2,135	2.2	2,139
School Anxiety Scale	0.2	2	0.5	2,135	1.5	2,139
Mathematics Test	0.9	2	0.3	2,131	3.6*	2,135
Letter Sets Test	0.3	2	1.0	2,130	2.6	2,134
Module Pretest	0.3	3	1.8	2,131	2.1	2,135
Attitude Towards School	0.8	2	0.6	2,119	1.1	2,123

* - significant at the .05 level.

a Johnson-Neyman analysis was performed on each of these pairs of treatments. [The Johnson-Neyman technique defines a region of homogeneity about the crossover point of two non-parallel regression lines (Potthoff, 1964).] Figures 3.1.1-3.1.6 represent the six pairs of treatment groups when the Letter Sets Test was used as the predictor and the Module Posttest was used as the criterion. As can be seen from these figures only two pairs of treatment groups had significantly different slopes at the .05 level (Figures 3.1.1 and 3.1.2).

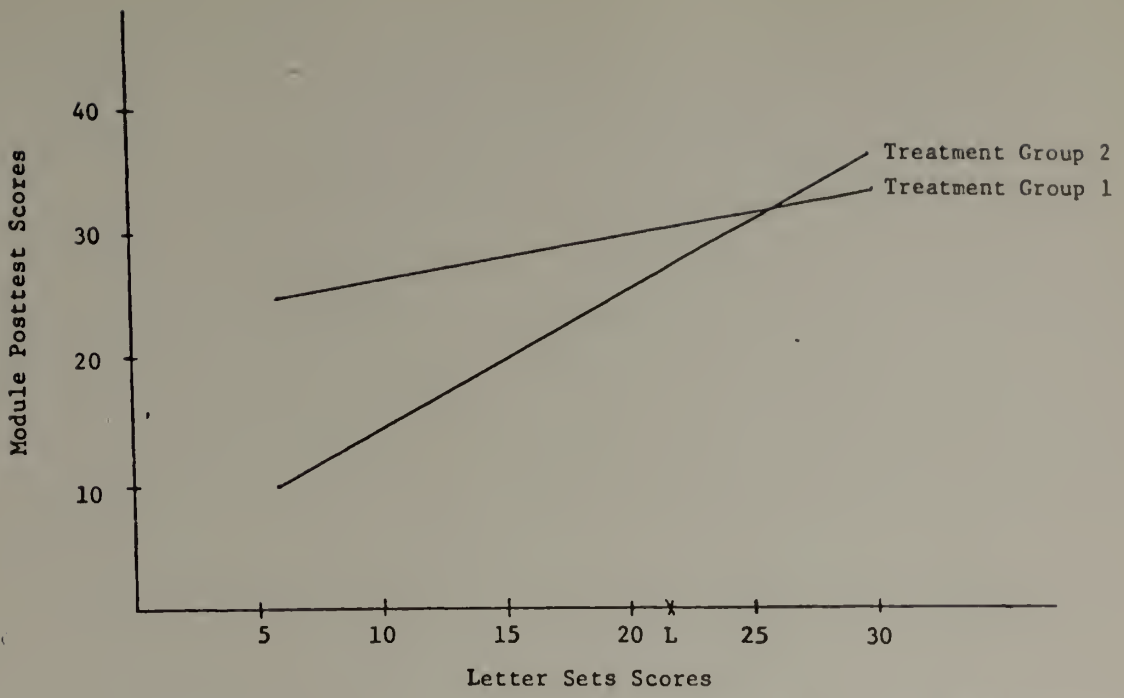


Figure 3.1.1 The regression of posttest scores on Letter Sets Scores for instructional treatment groups 1 and 2.

Figure 3.1.2 The regression of posttest scores on Letter Sets Scores for instructional treatment groups 1 and 3.

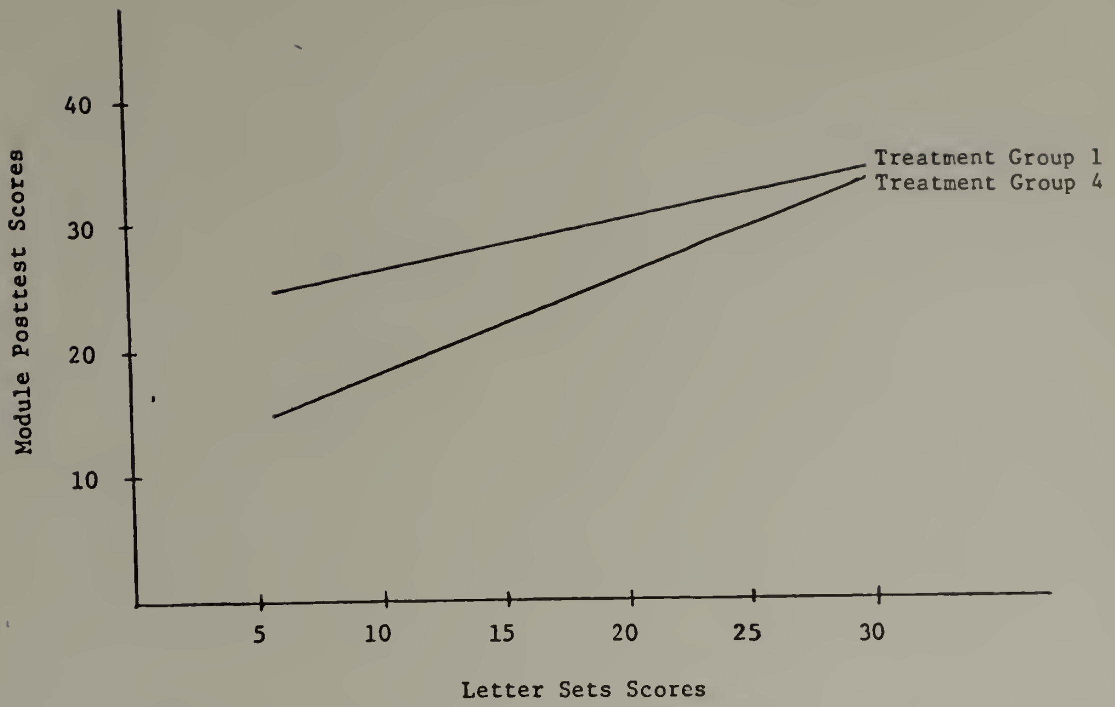


Figure 3.1.3 The regression of posttest scores on Letter Sets Scores for instructional treatment groups 1 and 4.

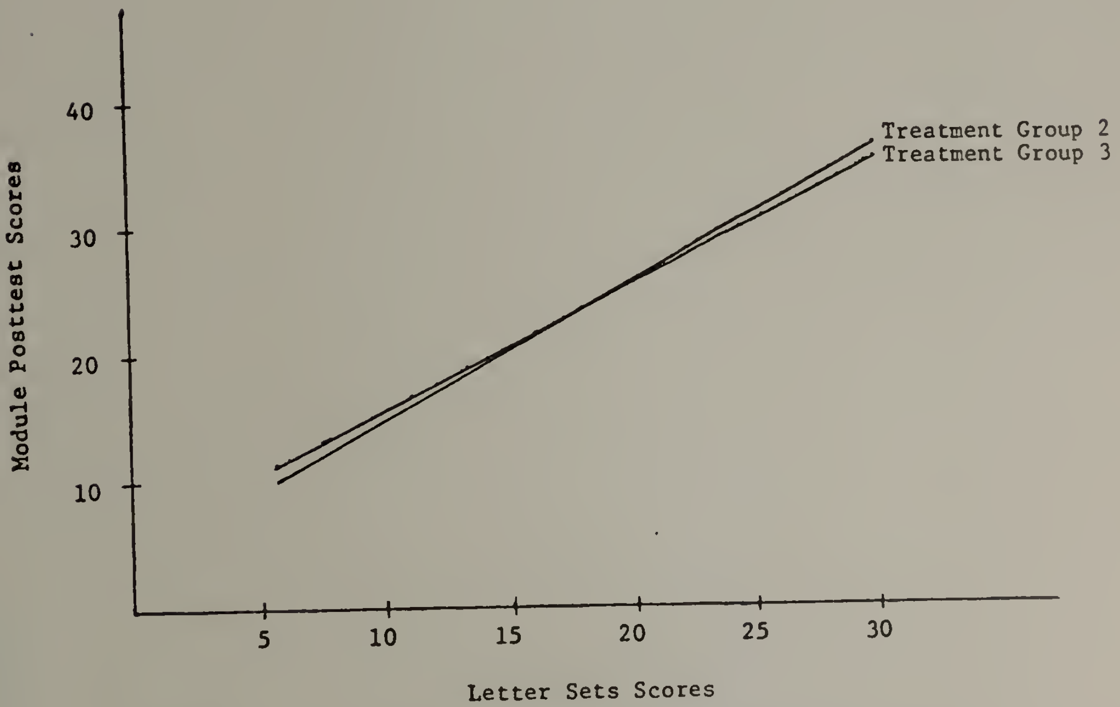


Figure 3.1.4 The regression of posttest scores on Letter Sets Scores for instructional treatment groups 2 and 3.

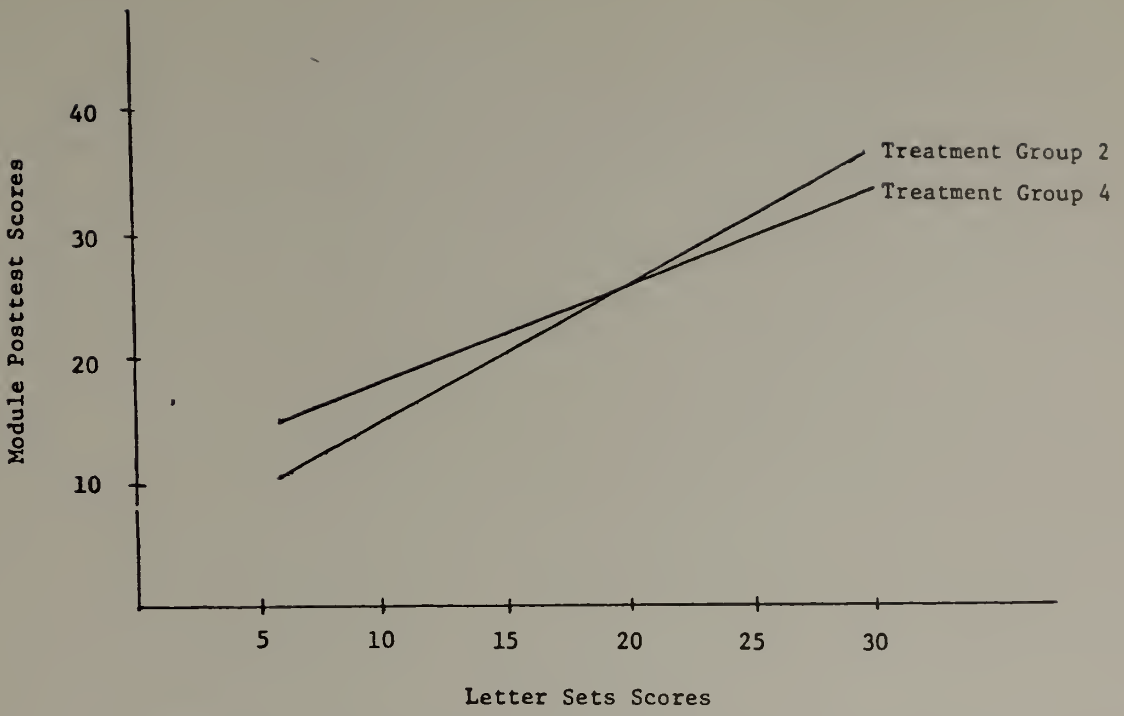


Figure 3.1.5 The regression of posttest scores on Letter Sets Scores for instructional treatment groups 2 and 4.

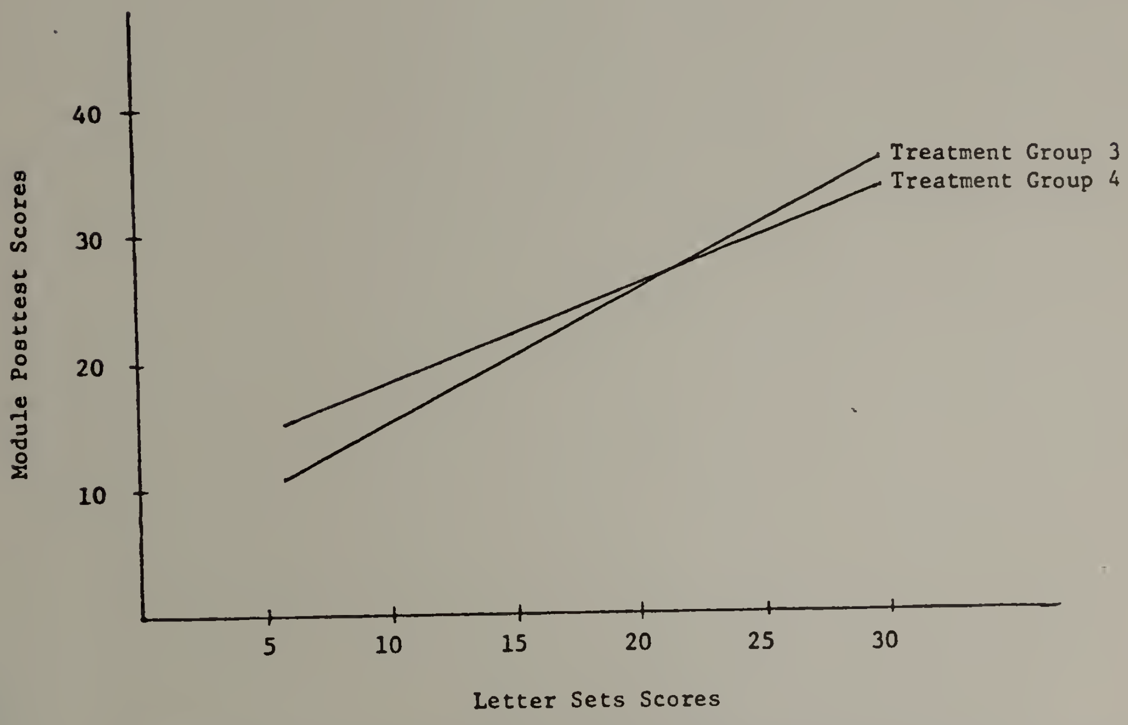


Figure 3.1.6 The regression of posttest scores on Letter Sets Scores for instructional treatment groups 3 and 4.

For Figure 3.1.1 the Johnson-Neyman analysis revealed a region of homogeneity with a lower limit of 22.0 and an upper limit of 35.2. For Figure 3.1.2 the region of homogeneity has a lower limit of 23.0 and an upper limit of 48.6. Because the upper limits are out of the range of the predictor variable only the lower limits are illustrated. Thus, for example, the mean criterion scores for treatment groups 1 and 2 (Figure 3.1.1) differ significantly (at the .05 level) when Letter Sets Test scores are below 22.0. To assign students to treatment group 1 with 95% confidence that the criterion scores will be significantly higher than those that would have resulted if assignment was made to treatment 2, the students must score lower than 22.0 on the Letter Sets Test.

For all practical purposes, the interactions between Letter Sets Test scores and treatment groups 1 and 2 and 1 and 3 are ordinal. The regression lines just barely cross at the high end of the range of the predictor variable. This was mirrored in the fact that the upper limit of the Johnson-Neyman solution was outside of the range of the Letter Sets Test.

Assuming that the interaction between treatments and the Letter Sets Test is significant one might make assignments to treatments in the future in the following way. The high scorers (25 to 30) would appear to benefit equally from treatments 1, 2, or 3. All of the students who score below 25 would perhaps be better off in treatment group 1.

As to what are the particular qualities of these treatments that cause such varying dependences on the ability measured by the Letter Sets Test one can only speculate. Other studies have shown that the Letter Sets Test loads heavily on an induction factor where induction is defined as the forming and trying out of hypotheses (French, Ekstrom, and Price, 1963). Perhaps future research will be able to "dissect" the instructional treatments and find the link between the Letter Sets Test and the basic learning variables in the treatments.

It was previously mentioned that the test of the parallelism of slopes was also rejected when the Letter Sets Test was used as the predictor, the Module Posttest was the criterion, and only three instructional treatment groups were considered (group 2 was removed). No illustrations are given for this case because the regression lines would be exactly the same as those previously considered.

The third rejection of the parallelism hypothesis occurred when the Test Anxiety Scale for Children was the aptitude, the Delayed Module Posttest was the criterion, and instructional treatment group 2 was excluded (Table 3.1.7a). It should be added that the .05 significance level was almost reached when the Module Posttest was the criterion (an F of 2.7 compared to a .05 cut-off value of about 3.0 - see Table 3.1.6a). Again to determine which pairs of treatment groups contributed to the significant F values, each pair of treatment groups was subjected to a test of parallelism of slopes. A Johnson-Neyman solution was also performed on these pairs. The plots of these pairs of treatment groups are presented in Figures 3.1.7-3.1.9. Only treatment groups 3 and 4 had significantly different slopes at the .05 level (Figure 3.1.9). The Johnson-Neyman region of non-significance for these two groups lies between the Test Anxiety Scale for Children scores of 5.8 and 23.3 (at the .05 level). This region of non-significance lies within the predictor variable range with 34% of the sample falling below the region and 1% falling above the region. This interaction would be termed disordinal. On the basis of this region of non-

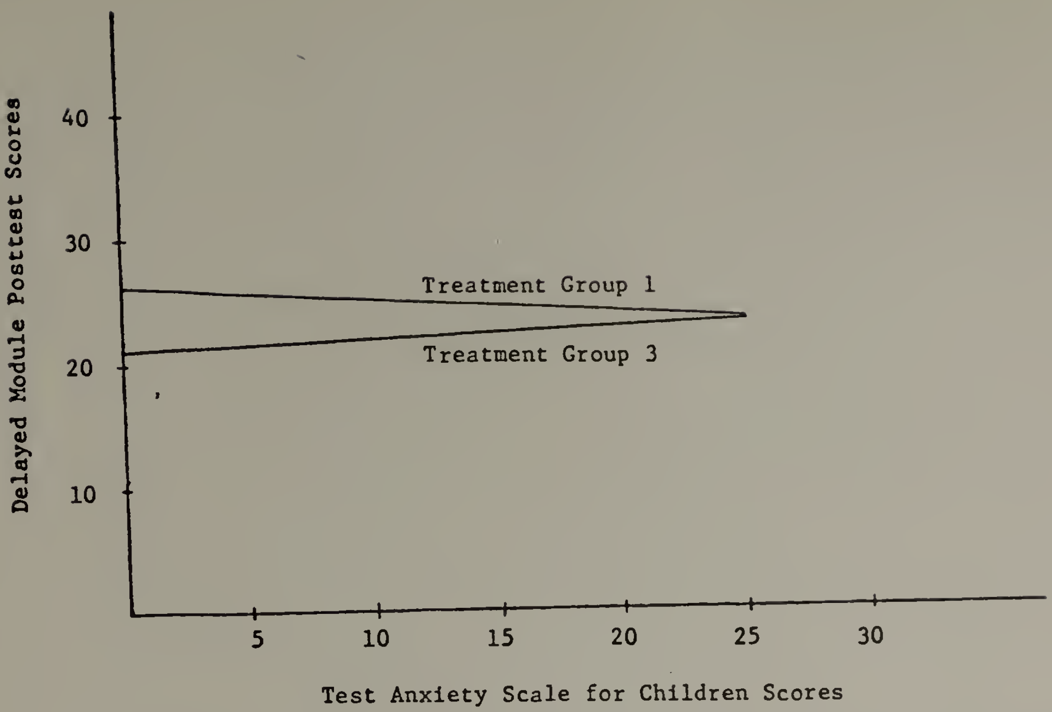


Figure 3.1.7 The regression of delayed posttest scores on Test Anxiety Scale for Children Scores for instructional treatment groups 1 and 3.

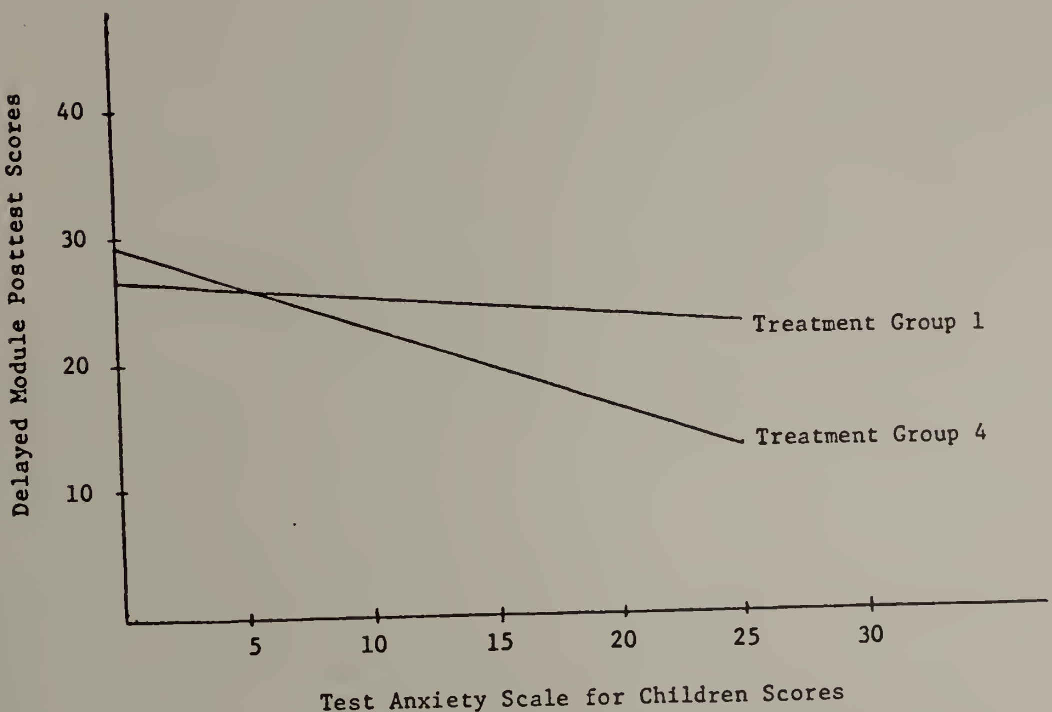


Figure 3.1.8 The regression of delayed posttest scores on Test Anxiety Scale for Children Scores for instructional treatment groups 1 and 4.

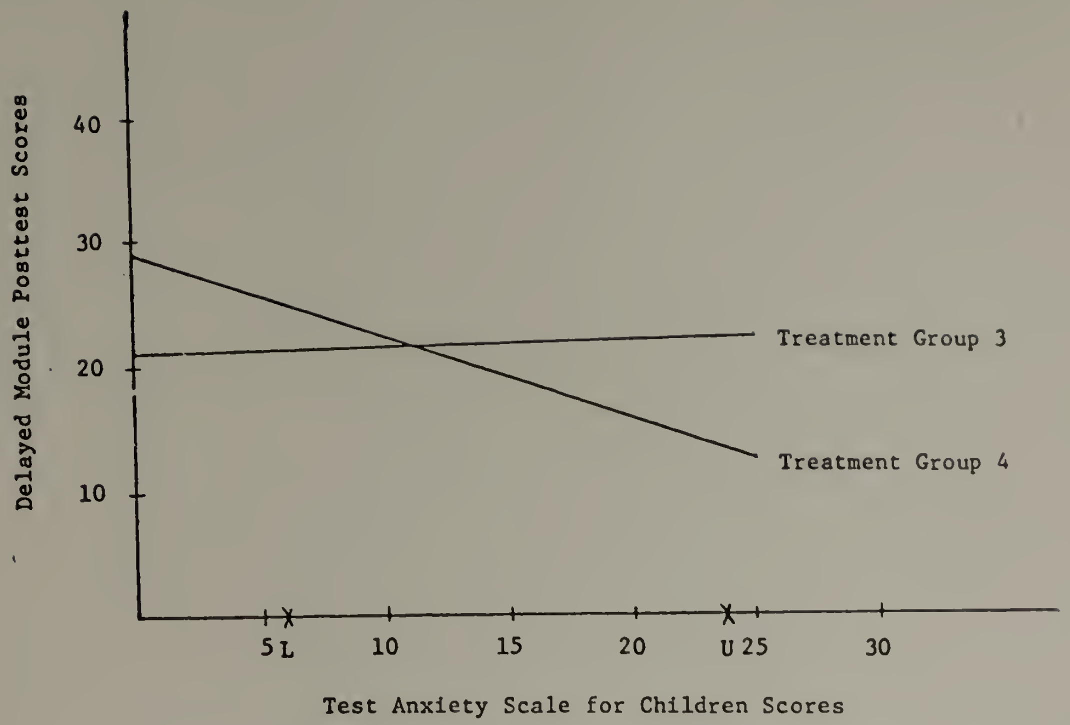


Figure 3.1.9 The regression of delayed posttest scores on Test Anxiety Scale for Children Scores for instructional treatment groups 3 and 4.

significance one would assign students to treatment 4 if they had anxiety scores below 6. Students with anxiety scores above 23 would be assigned to treatment 3. This would result in the maximum criterion scores for both groups.

Also, it is clear from Figure 3.1.7 that treatment 1 tends to be superior to treatment 3 throughout the range of anxiety scores. The difference between the two treatments becomes smaller, however, as the anxiety scores increase. The regression line for treatment 3 in Figures 3.1.7 and 3.1.9 indicates that the students in treatment 3 tend to achieve the same criterion scores irrespective of their anxiety scores. The criterion performance of the students in treatment 4, however, is markedly affected by their anxiety scores. Low anxious students score much higher on the criterion variable than do high anxious students.

These results seem to suggest the following about our treatments. Achievement by students in treatment 3 is about the same irrespective of anxiety level. There is a tendency for treatment 1 to penalize high anxious students. This tendency is even more marked in treatment group 4. As for underlying reasons for these differences, again one can only speculate.

The following course of action appears feasible for assigning students to one of the three treatment groups on the basis of anxiety scores. Low anxious students (scores less than 6) would seem to profit more from treatment 4. High anxious students (scores above 6) would profit more from treatment 1.

While there was no particular interest in testing for differences among treatment groups on the criterion measures, the results as indicated by the test of the third hypothesis of equal intercepts of regression reported in Tables 3.1.6, 3.1.6a, 3.1.7, and 3.1.7a indicate that generally the teacher mode was superior to the other modes.

Given the substantial amount of time that was spent in developing instructional materials; selecting, administering and analyzing aptitude test scores; and conducting the ATI analyses; the results were quite disappointing. Better results had certainly been expected. There exist a multitude of possible explanations which will be discussed at some length in the conclusions section.

3.2 Results of the Evaluation Study

Individual Differences Model for Multidimensional Scaling

(a) Reliability of the Similarity Judgments

As was previously stated, 13 additional concept pairs were appended to the original 78 to provide evidence about the reliability of the similarity judgments. The reliability estimates of the judgments calculated on the 13 concept pairs for each student exhibited extreme variability. To arrive at a reliability estimate across individuals, each individual reliability coefficient was transformed to a Fisher's Z. The Z values were then summed over individuals and averaged. The average Fisher's Z value was then converted back into a reliability coefficient. For the pretest data the reliability was .26 and for the posttest data the reliability was .47. It is obvious that the average individual reliability estimate in both the pretest and posttest situation is substantially lower than is desirable.

Fortunately in this analysis, group data and not individual data is of utmost concern. A group estimate of the reliability on each test occasion (pretest and posttest) was obtained by summing over individual judgments for each of the 13 concept pairs. The group reliability coefficient found by correlating the mean judgments on each pair for each occasion across the 13 concept pairs was .90 for the pretest data and .95 for the posttest data. These coefficients indicate that the group estimates of the similarity judgments are extremely consistent.

(b) Mean Differences

For 41 of the 78 pairs of concepts, the mean difference between the pre- and post- instruction similarity judgments was statistically significant at the .05 level. For 18 of the pairs the direction of change was toward judgments of greater similarity. For 23 of the 41 concept pairs the direction of change was toward judgments of greater dissimilarity. Of the 37 pairs for which mean differences between pre- and post- instruction judgments were not statistically significant, the change for 13 pairs was in the direction of increased similarity and for 24 pairs was in the direction of decreased similarity. The 41 concept pairs for which mean differences were statistically significant are listed in Table 3.2.1. Beside each pair are recorded the pre- and post-instruction means of similarity judgments and the associated t-statistic for the difference between the means.

Of the 41 concept pairs in Table 3.2.1, eight involve the concept mixture. The direction of change for these eight concept pairs was in the direction of increased dissimilarity. Instruction appears to have had the effect of heightening the students' perceptions of the dissimilarities involving the concept mixture and the remaining concepts.

There were 18 concept pairs judged as being significantly more similar after instruction. From these 18 pairs two groups appear to emerge. The first group includes the concepts proton, neutron, nucleus, electron, atomic number, and atomic weight. The eight concept pairs involving these concepts (proton-neutron, proton-atomic weight, proton-atomic number, atomic number-nucleus, atomic number-electron, neutron-atomic weight, neutron-nucleus, atomic weight-nucleus) seem to involve the component parts of the atom and their relationship to the concepts atomic number and atomic weight. The interrelationships of these concepts were more apparent to the students after instruction.

The second group whose interrelationships became more apparent after instruction include the concepts, radical, ion, oxidation number, electron, and shell or energy level. The seven pairs involving these concepts

Table 3.2.1

Statistical Data on Pairs of Concepts for Which Pre- and Post-Instruction Mean Differences Were Statistically Significant

Concept Pair	Pre-Instruction Mean	Post-Instruction Mean	t [1]
1. proton-ion	4.70	5.60	4.79
2. *proton-nucleus	3.92	3.47	-2.59
3. *proton-atomic number	4.91	2.99	-9.29
4. *proton-atomic weight	4.88	3.79	-6.47
5. *proton-mixture	5.51	6.28	4.44
6. atomic number-compound	5.27	6.52	6.63
7. *atomic number-nucleus	5.21	4.43	-3.89
8. *atomic number-mixture	5.29	6.23	5.27
9. atomic number-shell or energy level	5.21	4.75	-2.21
10. atomic number-ion	5.61	5.23	-2.25
11. *atomic number-radical	5.11	5.76	3.46
12. *atomic number-electron	5.04	3.76	-6.19
13. *electron-shell or energy level	4.96	3.52	-6.78
14. electron-oxidation number	5.67	4.09	-8.36
15. electron-mixture	6.36	6.59	1.19
16. *electron-ion	4.58	4.14	-2.14
17. neutron-shell or energy level	5.50	5.96	2.38
18. *neutron-atomic weight	5.08	4.06	-5.10
19. neutron-ion	4.46	5.69	6.43
20. neutron-radical	5.49	6.03	3.46
21. neutron-compound	5.42	6.01	3.28
22. *neutron-nucleus	4.27	3.78	-2.56
23. neutron-mixture	5.81	6.29	2.42
24. *atomic weight-compound	5.49	6.34	4.47
25. atomic weight-nucleus	4.69	4.01	-3.60
26. atomic weight-shell or energy level	5.18	5.61	2.32
27. atomic weight-oxidation number	4.77	5.34	3.10
28. mixture-compound	3.38	4.15	3.97
29. *mixture-nucleus	5.78	6.55	4.06
30. mixture-shell or energy level	5.39	6.18	4.38
31. *mixture-oxidation number	5.62	6.28	3.37
32. mixture-ion	5.55	6.14	3.42
33. shell or energy level-radical	5.27	5.54	1.40
34. shell or energy level-oxidation number	5.08	4.21	-4.23
35. *shell or energy level-ion	4.95	4.53	-2.00
36. *ion-element	4.78	5.21	2.25
37. *ion-radical	5.45	4.91	-2.81
38. ion-nucleus	4.82	5.35	2.72
39. ion-oxidation number	5.31	4.73	-2.81
40. radical-compound	5.41	4.58	-4.56
41. *radical-nucleus	5.14	5.63	3.22

NOTE: Similarity was rated on a scale from 1 to 9 with 1 the most extreme similarity rating, and 9 the most extreme dissimilarity rating. For this analysis, N=190.

*The concepts in the starred pairs appeared in reverse order in the Similarity Judgment Questionnaire.

[1]For a two-tailed test, $t_{.05,189} \simeq 1.98$.

(electron-shell or energy level, electron-oxidation number, electron-ion, shell or energy level-oxidation number, shell or energy level-ion, ion-oxidation number, ion-radical) reflect the "electrical" commonality among this group.

Two of the three remaining concept pairs with significant post-instruction changes in the direction of greater similarity (atomic number-shell or energy level and atomic number-ion) reflect the relationship between the concept atomic number and the electrical properties of the atom. The increased similarity in the third pair (radical-compound) seems to reflect the rules used in compound formation and the part that radicals play in these rules.

There were 23 concept pairs whose direction of change was significant and in the direction of increased dissimilarity. The 8 pairs involving the concept mixture were previously considered. The remaining 15 pairs did not appear to represent any particular pattern other than crossovers between the previously mentioned two groups. That is, concept pairs made up of a concept from the "component parts of the atom and atomic number, atomic weight" group and a concept from the "electrical" group were generally judged as being more dissimilar after instruction.

Another interesting aspect is revealed by Table 3.2.1. As can be seen from the table, students' perceptions of some concepts were more affected by instruction than their perceptions of others. For example, the concept element appeared in only one pair (out of 12 possible pairs) for which the mean similarity judgment was significantly more dissimilar after the course than before it. At the other extreme, the concept ion was contained in nine pairs for which the mean judgments of similarity were significantly more similar or dissimilar after than before instruction. It appeared that instruction had the effect of altering student's perceptions of some concepts more than it did of others.

(c) Points of View Analysis

As was previously stated, the Tucker-Messick procedure for determining the number of points of view is essentially subjective. Initially the matrix of cross-products is decomposed into its eigenroots and eigenvectors. The square-roots of the eigenroots are then plotted as a function of their rank in terms of size. One looks for a "gap" in the curve (Green, 1966). This gap is taken as the separator between the significant and insignificant dimensions of the cross-products matrix. The number of eigenroots before the gap is taken as the number of points of view.

In this study, a pre-instruction cross-product matrix was formed for the following group:

1) Total Class

Post-instruction cross-product matrices were formed for the following groups:

1) Total Class

2) High Achievers on the Module Posttest

3) Low Achievers on the Module Posttest

Reported in Table 3.2.2 and Table 3.2.2a are the square-roots of the largest 15 eigenroots and associated cumulative proportions of the traces for the cross-product matrices of the previously mentioned pre- and post-instruction groups. The gap in the eigenroots is indicated by black lines under the eigenroot value and the associated cumulative proportion of the trace. The number of points of view is taken to be the number of eigenroots above the

Table 3.2.2

Square-Roots of the Largest 15 Eigenroots
and the Associated Cumulative Proportions of the
Trace of the Pre-and Post-Instruction Cross-Product
Matrices [for the Total Class (N=210)]

Eigenroot	Pre-Instruction Matrix		Post-Instruction Matrix	
	Square-Root of Eigenroot	Cumulative Proportion of the Trace	Square-Root of Eigenroot	Cumulative Proportion of the Trace
1	659.96	.896	677.40	.891
2	57.65	.902	80.14	.903
3	49.40	.907	59.82	.910
4	44.33	.911	46.66	.914
5	41.33	.915	44.34	.918
6	40.60	.918	43.93	.922
7	39.38	.922	43.24	.926
8	37.88	.924	39.49	.929
9	37.84	.927	38.59	.932
10	36.82	.930	38.12	.934
11	36.74	.933	37.31	.937
12	35.54	.936	36.10	.940
13	35.04	.938	35.71	.942
14	34.23	.940	34.49	.944
15	33.84	.943	34.30	.947

Table 2.2.2a

Square-Roots of Largest 15 Eigenroots and Associated Cumulative Proportions of the Trace of Two Cross-Product Matrices (the Post-Instruction Similarity Judgments Data Divided into Two Groups on the Basis of Module Posttest Scores)

Eigenroot Number	Post Instruction Data: High Achievers[3] Square-Root of Eigenroot	High Achievers[3] Cumulative Proportion of the Trace	Low Achievers[4] Square-Root of Eigenroot	Low Achievers[4] Cumulative Proportion of the Trace
1	482.71	.903	434.75	.898
2	<u>54.37</u>	<u>.914</u>	46.07	.908
3	40.11	.920	<u>35.59</u>	<u>.914</u>
4	35.44	.925	31.64	.919
5	33.48	.930	30.71	.923
6	32.81	.934	30.43	.927
7	32.02	.938	30.07	.932
8	30.58	.941	28.73	.936
9	29.85	.945	26.72	.939
10	28.50	.948	25.82	.942
11	27.52	.951	25.44	.945
12	26.79	.954	25.29	.948
13	26.17	.956	24.58	.951
14	25.40	.959	24.17	.954
15	23.98	.961	23.43	.957

3-N = 97 4-N = 90

black line.

The number of points of view for each set of data reported in Table 3.2.2 were as follows:

Pre-Instruction

1) Total Class = 2

Post-Instruction

1) Total Class = 3

2) High Achievers on Module Posttest = 2

3) Low Achievers on Module Posttest = 3

These results seemed to suggest that the number of points of view required to account for the data was slightly higher for the posttest data than for the pretest data.

(d) Configuration of Concepts

The inter-concept distances associated with each point of view serve as input to a nonmetric multidimensional scaling analysis (Young, 1968). To obtain these inter-concept distances, it is necessary to isolate clusters of people representing the various points of view. Average loadings for each cluster are then obtained, and after arraying the averages as columns in a matrix B, the matrix of stimulus pairs by points of view is post-multiplied by the inverse of the B matrix. The resulting matrix gives us the distances of the stimulus pairs for the various points of view. This method has been advocated by Pennell (1971).

While clusters representing points of view could easily be identified with artificial data, this was not the case with real data. Thus instead of identifying clusters of students and deriving average loadings for each point of view, the student-space was spanned by selecting students who loaded highly on a particular point of view. The inter-concept distance matrix derived by this method represents the perceptions of a person whose judgments are extreme representations of the particular point of view. The distances do not represent the perceptions of a person whose judgments are an average representation of the point of view.

<u>Criterion for Determining Dimensionality</u>

The number of dimensions accepted was determined by considering Young's measure of goodness of fit (the α function). The magnitude of this measure is directly proportional to the number of dimensions in the configuration. As a rule of thumb, Young suggests that values in excess of 0.999 are needed for a satisfactory solution (Young, 1968). Thus in all but two cases the lowest dimensional solution that had an index of fit in excess of 0.999 was chosen. The best fitting configuration for the first post-instruction point of view for the total class gave an α value of 0.99900 (and a Kruskal stress value of 0.063). The best fitting configuration for the first post-instruction point of view for the low achievers had an α value of 0.99821 (and a Kruskal stress value of 0.085). These two solutions exceed the maximum stress level which can be accepted at a 0.05 significance level for 13 points in 5 dimensions (Wagenaar and Padmos, 1971). For these two solutions the probability that the fit is attributable to chance is greater than 0.05. Since the validity of the student's judgments in these cases is open to question they were not reported.

Cluster Analysis

Johnson's hierarchical clustering schemes which have been found to be useful in clarifying a multidimensional scaling solution partition stimuli into groups on the basis of similarity or dissimilarity measures among those stimuli. The input for this technique consists of the n(n-1)/2 similarity or dissimilarity measures among the n stimuli. At each stage of the clustering process two or more stimuli or clusters separated by a minimum distance are replaced by a composite cluster composed of those stimuli or clusters. The distances between the remaining stimuli and the composite cluster are recomputed at each stage to satisfy the ultrametric inequality. If a and b are two stimuli clustered at a particular level but not at the next higher level, and if z is another stimulus at the next higher level, then the ultrametric inequality states that the distance between a and z equals the distance between b and z. Yet this relationship rarely holds with real data. Johnson has proposed two variations to handle real data. In his "connectedness" method if stimuli a and b are clustered at a particular level but not at the next higher level, the distance from the (a,b) cluster to any third stimulus z is defined as the minimum of the distance between b and z. In Johnson's "diameter" method the distance from the (a,b) cluster to any third stimulus z is defined as the maximum of the distance between a and z or between b and z. Johnson cites examples to show that both methods yield very similar results (Johnson, 1967).

(d.1) Multidimensional Scaling Configurations for the Total Class

Table 3.2.3 presents the multidimensional scaling configurations for the two pre- and two of the three post-instruction points of view for the total class. The first post-instruction point of view was not reported because a suitable fit was not attained. The two pre-instruction points of view for the total class were represented in five and four dimensions. The two "good" post-instruction points of view were represented in three and four dimensions.

One useful criterion for evaluating multidimensional configurations is interpretability. A meaningful solution is one in which the dimensions can be interpreted in terms of the logical groupings of the concepts. Since instruction, hopefully, develops a logical interrelationship among the concepts, the interpretabilities of the solutions should increase after instruction.

With this particular group of concepts, several subgroups seem apparent. One subgroup represents the component parts of the atom. It includes the concepts nucleus, proton, neutron, and electron. An "electrical" subgroup includes the concepts ion, shell or energy level, oxidation number, radical, and electron. A third subgroup includes the concepts mixture, compound, and possibly radical, and element. A fourth grouping could include atomic number, atomic weight, proton, neutron, nucleus, and electron. While other logical groupings are undoubtably possible, such logical groupings should be more apparent after instruction.

The stimulus dimensions for the first and second pre-instruction points of view for the total class are shown in Figures 3.2.1 and 3.2.2. The stimulus dimensions for the second and third post-instruction points of view for the total class are shown in Figures 3.2.3 and 3.2.4. (The

Table 3.2.3
Multidimensional Scaling Configurations for Pre- and Post-Instruction Points of View of 13 Concepts for the Total Class

Concept	Pre-Instruction									Post-Instruction						
	Dimensions of First Point of View					Dimensions of Second Point of View				Dimensions of Second Point of View			Dimensions of Third Point of View			
	1	2	3	4	5	1	2	3	4	1	2	3	1	2	3	4
protron	-20	42	02	06	-06	31	-20	01	-18	38	11	-41	14	36	10	-03
ion	32	43	15	09	-15	04	-16	-42	-08	-33	-52	03	-05	05	-43	02
neutron	50	09	-30	09	00	42	-08	-18	04	19	75	-19	10	51	-00	02
element	45	-39	10	01	01	-04	09	-03	-29	00	13	03	-20	-06	-03	-51
radical	07	-04	05	-59	-09	-49	08	02	20	-14	-17	56	-18	-51	02	-21
electron	-06	-68	-01	03	-02	17	-26	-22	10	10	-36	-07	28	22	-19	25
compound	-05	07	-41	-02	21	-08	56	-04	-12	-01	20	69	-54	-22	-24	-33
nucleus	-18	-02	36	-40	-15	34	-13	11	05	11	41	-39	-04	30	20	34
atomic number	03	-09	-01	07	65	-06	-23	32	-16	17	-06	-50	40	06	12	06
atomic weight	-21	-01	-48	08	-32	05	-09	46	01	47	40	-21	11	12	44	06
mixture	04	01	14	44	-07	-05	62	-01	-04	-05	34	30	-57	-25	-03	-34
shell or energy level	-07	16	49	12	07	-11	-02	-04	55	12	-63	-10	18	-12	-04	50
oxidation number	-64	06	-10	03	-08	-52	-19	02	-07	-02	-62	28	38	-45	08	17

Note: Decimals have been omitted. Coefficients with an absolute value greater than .30 have been underlined. Both the pre-instruction and post-instruction configurations were derived from the judgments of 210 students.

Figure 3.2.1 Stimulus dimensions for first pre-instruction point of view for the total class (only stimuli whose loadings exceed ± .10 are represented).

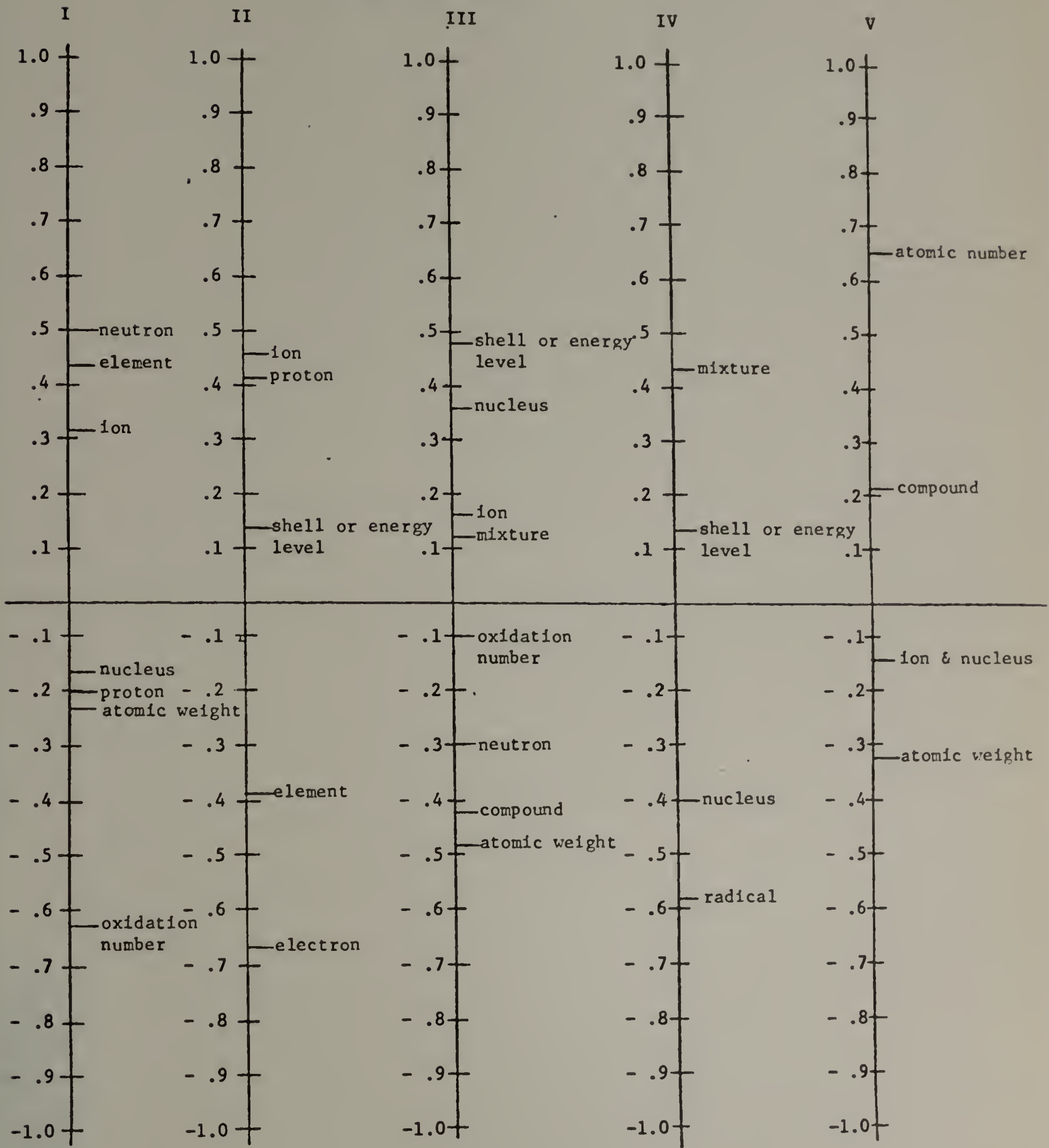


Figure 3.2.2 Stimulus dimensions for second pre-instruction point of view for the total class (only stimuli whose loadings exceed ± .10 are represented).

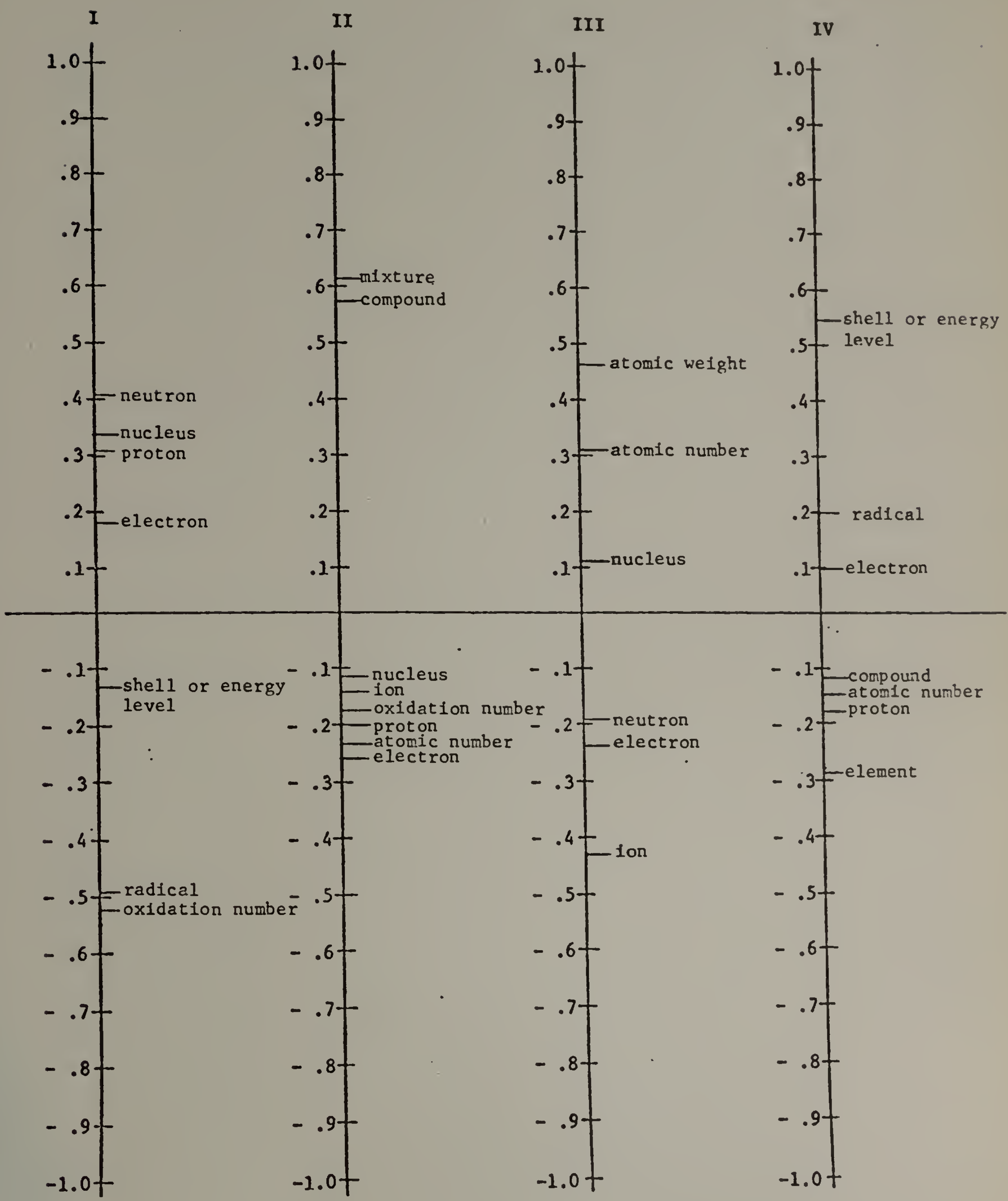


Figure 3.2.3 Stimulus dimensions for second post-instruction point of view for the total class (only stimuli whose loadings exceed ± .10 are represented).

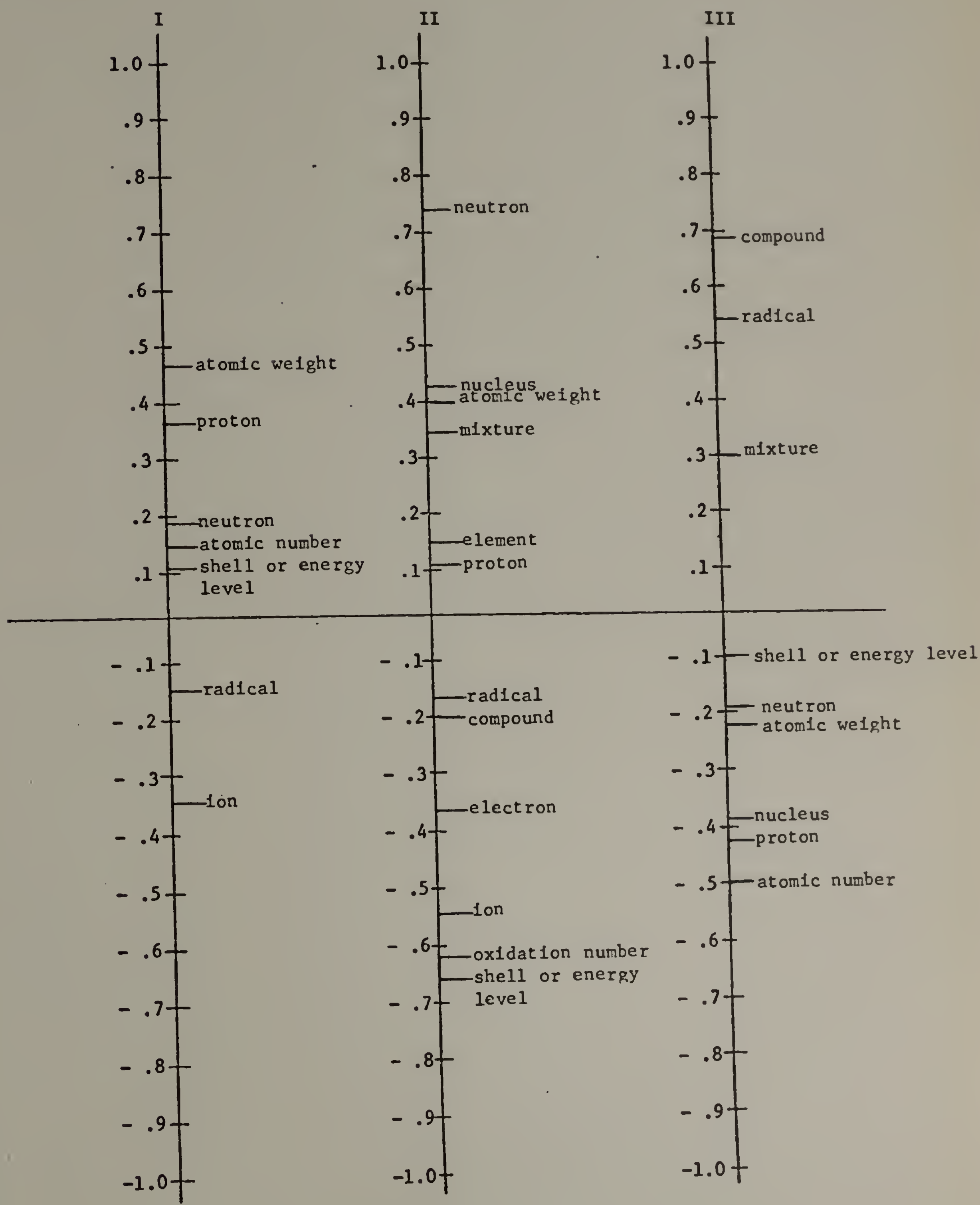
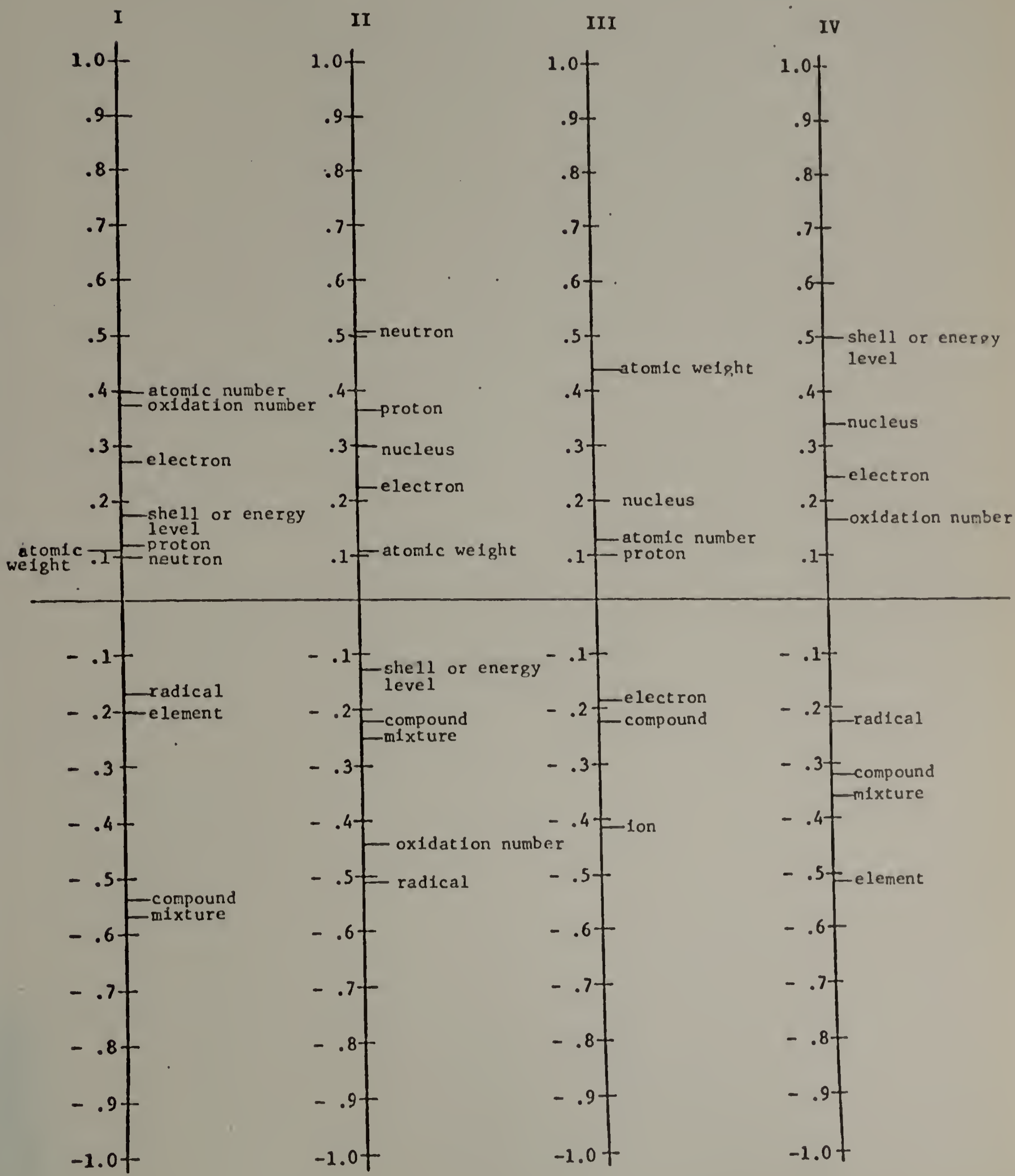


Figure 3.2.4 Stimulus dimensions for third post-instruction point of view for the total class (only stimuli whose loadings exceed ± .10 are represented).



first post-instruction point of view is not represented because of its poor index of fit value.)

It is apparent that the dimensions of the second pre-instruction point of view (Figure 3.2.2) conformed well with the dimensions of the third post-instruction point of view (Figure 3.2.4). Pre-dimension 1 and post-dimension 2 contrasted the component parts of the atom with the "electrical" concepts. Pre-dimension 2 and post-dimension 1 contrasted the concepts compound and mixture with concepts atomic number, electron, and proton. Pre-dimension 3 and post-dimension 3 contrasted the concepts atomic number, atomic weight, and nucleus with the concept ion. Pre-dimension 4 and post-dimension 4 contrasted element with the "electrical" concepts. Thus students adopting the second pre-instruction point of view apparently related the concepts much the same way as the students adopting the third post-instruction point of view.

An examination of the first pre-instruction point of view (Figure 3.2.1) showed a complete breakdown of the logical groupings. Concept groupings such as ion, proton; electron, element; or shell or energy level, nucleus showed little knowledge of the concepts involved. This point of view appeared to involve the perceptions of those students uninitiated to the meanings of the concepts.

The second post-instruction point of view for the total class (Figure 3.2.3) was presented in three dimensions. Dimension 1 contrasted ion with the concepts atomic weight, proton, neutron, and atomic number. The second dimension contrasted neutron, nucleus, atomic weight, and mixture with the "electrical" concepts. Dimension 3 contrasted compound, radical, and mixture with atomic number, proton, nucleus, atomic weight, and neutron. It is interesting that in this point of view the concepts nucleus, proton, and neutron were interrelated with the concepts atomic number and atomic weight. The concept electron was found in the "electrical" subgroup. It appeared that the component parts of the atom as an entity in themselves were ignored in this point of view.

In terms of interpretability, the effect of instruction appeared to have been the production of an additional interpretable point of view.

To help clarify these points of view consider their clustering solutions.

Figure 3.2.1a shows the hierarchical clustering solution for the first pre-instruction point of view for the total class. An examination of the clusters shows why this point of view represented the perceptions of those students who were unaware of the concept meanings. Such clusterings as the following do not have any logical basis: electron and element; ion and neutron, radical and nucleus; proton and shell or energy level; and atomic weight and compound.

Examination of Figure 3.2.2a and 3.2.4a shows the similarity between the second pre-instruction point of view for the total class and the third post-instruction point of view for the total class. While the clusterings of concepts like radical, ion, and element differed between the two points of view, other similarities emerged. These included the following clusters: compound and mixture; atomic number and atomic weight; proton and neutron; shell or energy level and oxidation number; and the larger cluster of nucleus, electron, neutron, and proton. These similarities seemed to substantiate the idea that a subgroup of students was exposed to the concepts prior to instruction.

Figures 3.2.3a and 3.2.4a contrast the second and third post-instruction points of view for the total class. The clusters that emerged for the second

Figure 3.2.1a Tree structure representing the hierarchical clustering solution for the first pre-instruction point of view for the total class.

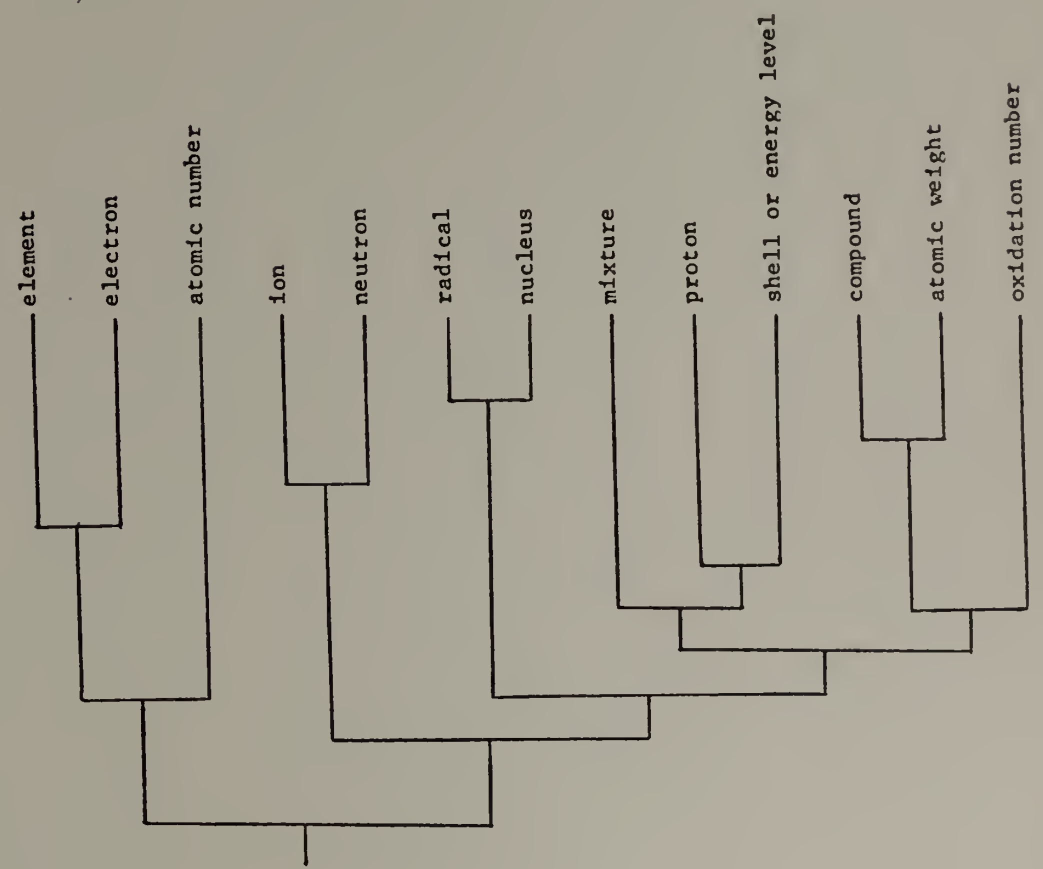


Figure 3.2.2a Tree structure representing the hierarchical clustering solution for the second pre-instruction point of view for the total class.

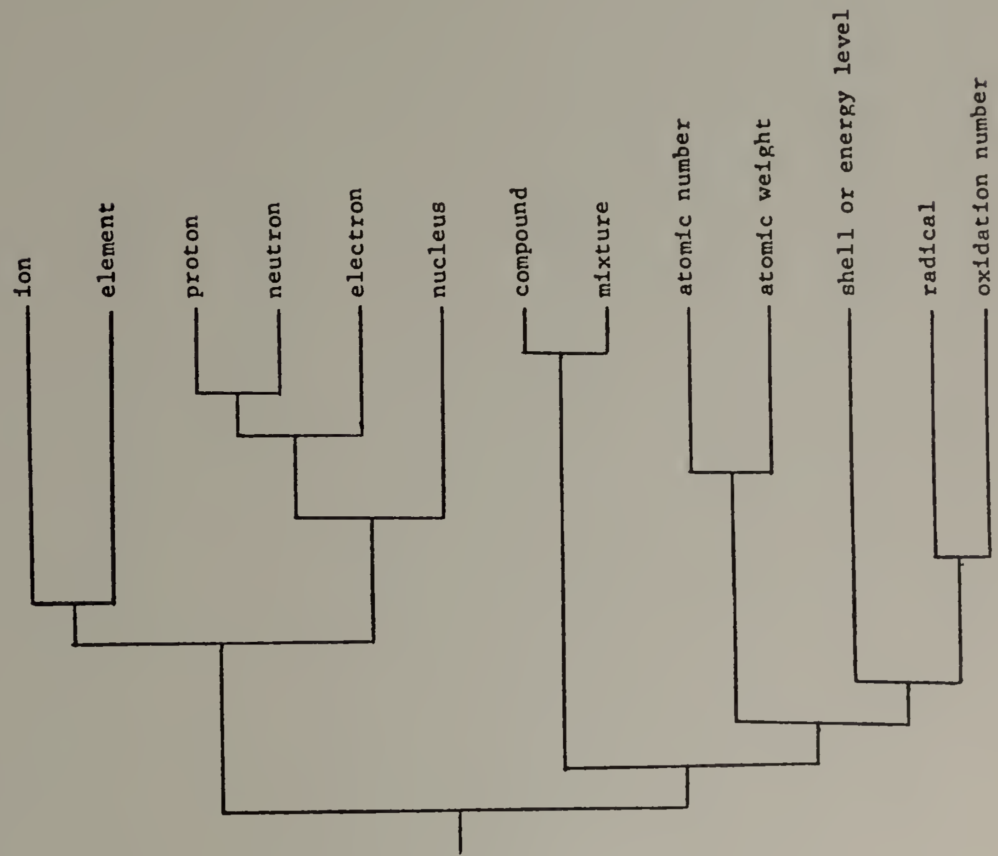


Figure 3.2.3a Tree structure representing the hierarchical clustering solution for the second post-instruction point of view for the total class.

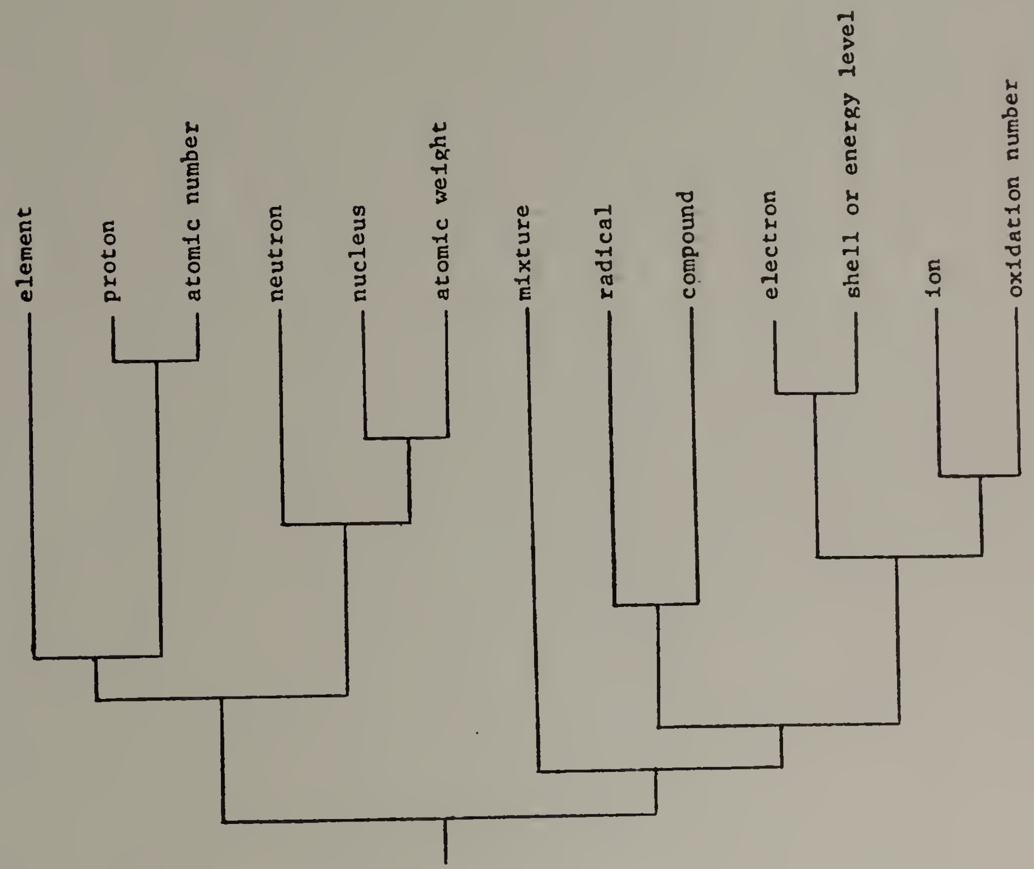
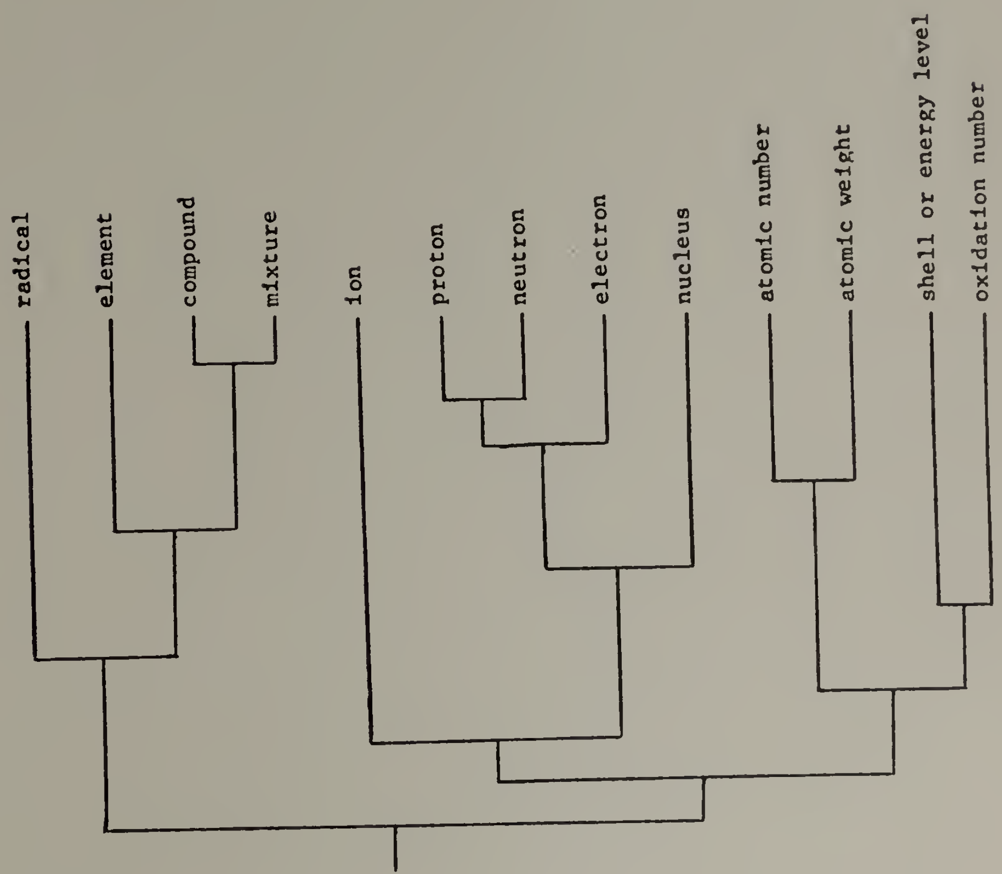


Figure 3.2.4a Tree structure representing the hierarchical clustering solution for the third post-instruction point of view for the total class.



point of view were:

1) element, proton, atomic number

2) neutron, nucleus, atomic weight

3) mixture

4) radical, compound

5) electron, shell or energy level, ion, oxidation number.

The clusters that emerge from the third point of view were:

1) radical, element, compound, mixture

2) ion

3) proton, neutron, electron, nucleus

4) atomic number, atomic weight

5) shell or energy level, oxidation number.

Important distinctions were the splitting up of the component parts of the atom subgroup (found in the third point of view) in the second point of view. The second point of view grouped the components of the atom around the concepts atomic weight and atomic number. Thus proton and atomic number cluster as do neutron, nucleus and atomic weight. The concepts ion and electron were clustered with shell or energy level and oxidation number in the second point of view but not in the third. An isolated concept in the second point of view was mixture while ion was isolated in the third.

(d.2) Multidimensional Scaling Configurations for the High and Low Achievers

Table 3.2.4 shows the multidimensional scaling configurations for the two post-instruction points of view for the high achievers and for two of the three post-instruction points of view for the low achievers. The first post-instruction point of view for the low achievers was not reported because a suitable fit was not attained. The two post-instruction points of view for the high achievers are represented in four and three dimensions. The

Table 3.2.4
Multidimensional Scaling Configurations for Post-Instruction Points of View of 13 Concepts for Those Students Who Scored in the Top and Bottom Halves on the Module Posttest Distribution

Concept	Top Half							Bottom Half							
	Dimensions of First Point of View				Dimensions of Second Point of View			Dimensions of Second Point of View				Dimensions of Third Point of View			
	1	2	3	4	1	2	3	1	2	3	4	1	2	3	4
proton	-12	05	-47	03	-26	-08	-36	01	40	02	-01	-25	33	08	-34
ion	-23	-46	33	06	48	-13	05	-10	05	-54	04	-44	-39	-03	37
neutron	-14	-07	-53	03	-62	02	-06	-08	47	-03	17	01	19	-46	-30
element	20	-09	01	67	-03	14	02	61	-01	-09	-15	01	00	-37	11
radical	19	-24	44	31	12	-08	54	03	-73	-01	07	14	00	06	61
electron	-28	-33	-14	-19	16	-23	-04	-22	52	-01	10	-42	06	13	-06
compound	67	-03	15	16	03	30	49	67	-17	-14	08	40	-59	11	14
nucleus	01	-07	-29	-38	-46	-02	-21	-22	01	-06	58	-00	10	00	-48
atomic number	-42	11	-02	-08	-03	-13	-45	-23	09	47	-22	-34	-11	-28	-29
atomic weight	-01	47	-11	-15	-40	14	-35	-08	-00	59	06	02	46	-07	-04
mixture	61	05	26	07	31	79	18	54	-33	-12	-15	102	-17	-04	12
shell or energy level	-10	02	12	-61	29	-42	-02	-56	01	-12	-04	-06	29	43	14
oxidation number	-28	58	25	10	39	-30	22	-36	-30	05	-53	-07	-17	44	02

Note: Decimals have been omitted. Coefficients with an absolute value greater than .30 have been underlined. The post-instruction configuration for the top half of the posttest distribution was derived from the judgments of 97 students. The post-instruction configuration for the bottom half of the posttest distribution was derived from the judgments of 90 students.

two "good" post-instruction points of view for the low achievers are represented in four dimensions.

Figures 3.2.5 and 3.2.6 show the stimulus dimensions for the first and second points of view for those students who scored in the top half of the module posttest distribution. Figures 3.2.7 and 3.2.8 show the stimulus dimensions for the second and third points of view for those students who scored in the bottom half of the module posttest distribution. It will be remembered that the scaling solution for the first point of view of the low achievers could have occurred by chance with a probability exceeding 0.05. This point of view may have represented the judgments of those students who failed to assimilate the interrelationships of the concepts from instruction.

Dimension 1 of the first point of view for the high achievers (Figure 3.2.5) contrasted the concepts compound, mixture, and element with the concepts atomic number, electron, and shell or energy level. Dimension 2 contrasted oxidation number and atomic weight with several "electrical" concepts. Dimension 3 contrasted a combination of the mixture, compound and "electrical" subgroup with the component parts of the atom. Dimension 4 contrasted element, radical, and compound with shell or energy level and nucleus. The groupings of concepts on several of these dimensions must be questioned. In particular, the grouping of oxidation number with atomic weight on dimension 2 and the closeness of shell or energy level to nucleus on dimension 4 appeared to be contrary to the meanings of these concepts.

Dimension 1 of the second point of view for the high achievers (Figure 3.2.6) contrasted the "electrical" concepts with atomic weight, neutron, nucleus and proton. Dimension 2 contrasted the mixture, compound, element subgroup with the "electrical" concepts. Dimension 3 contrasted radical

Figure 3.2.5 Stimulus dimensions for first post-instruction point of view for those students who scored in the top half on the module posttest (only stimuli whose loadings exceed ±.10 are represented).

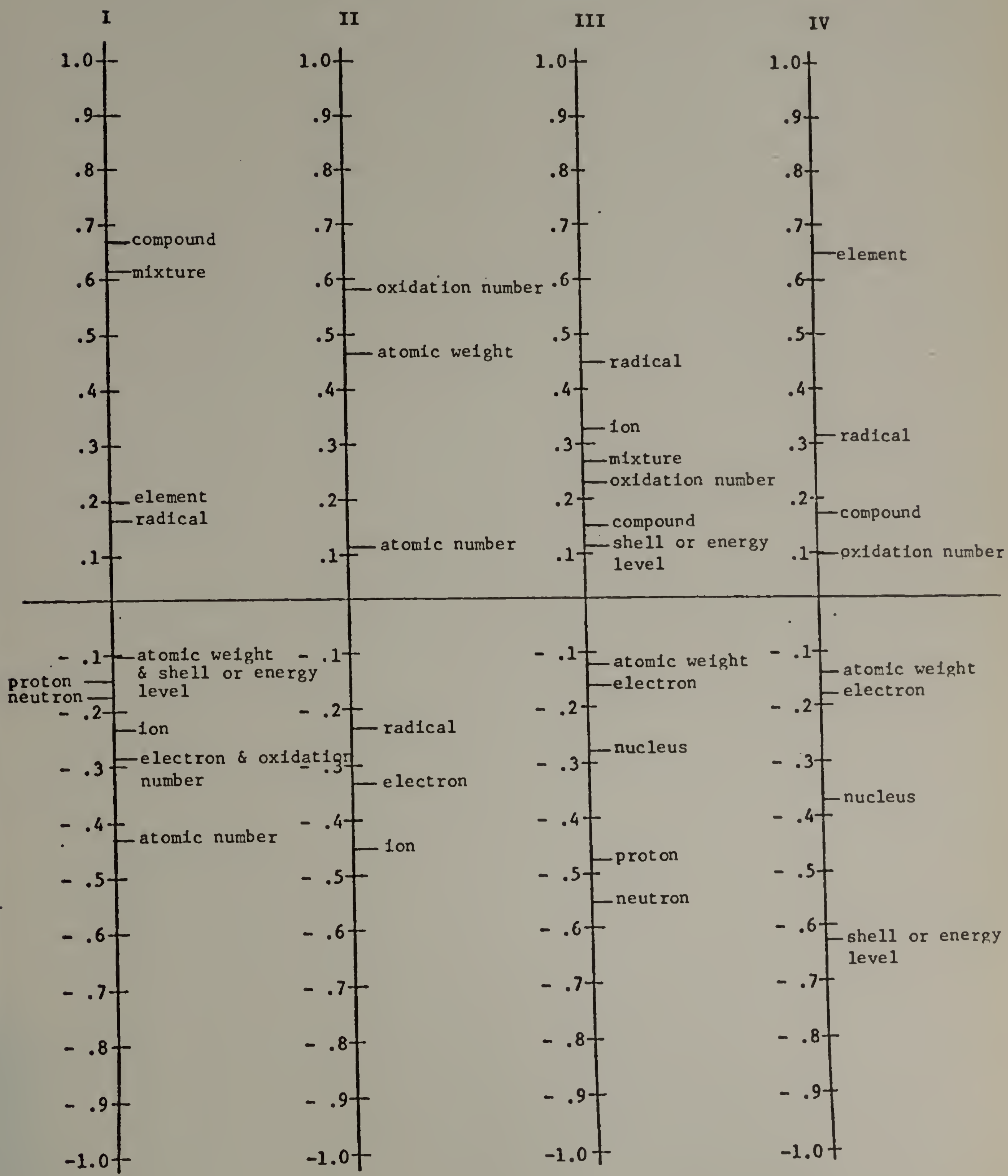


Figure 3.2.6 Stimulus dimensions for second post-instruction point of view for those students who scored in the top half on the module posttest (only stimuli whose loadings exceed ±.10 are represented).

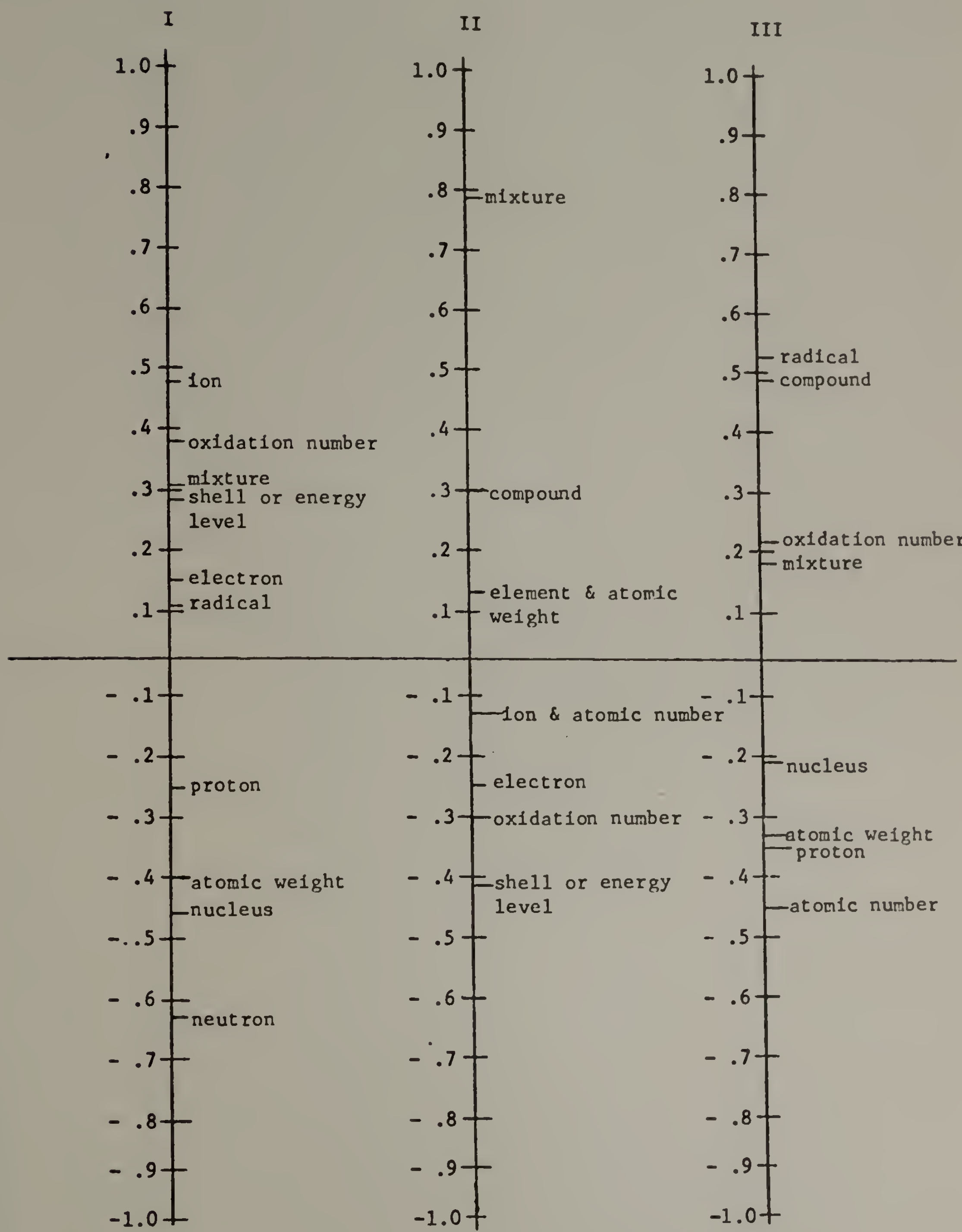


Figure 3.2.7 Stimulus dimensions for second post-instruction point of view for those students who scored in the bottom half on the module posttest (only stimuli whose loadings exceed ±.10 are represented).

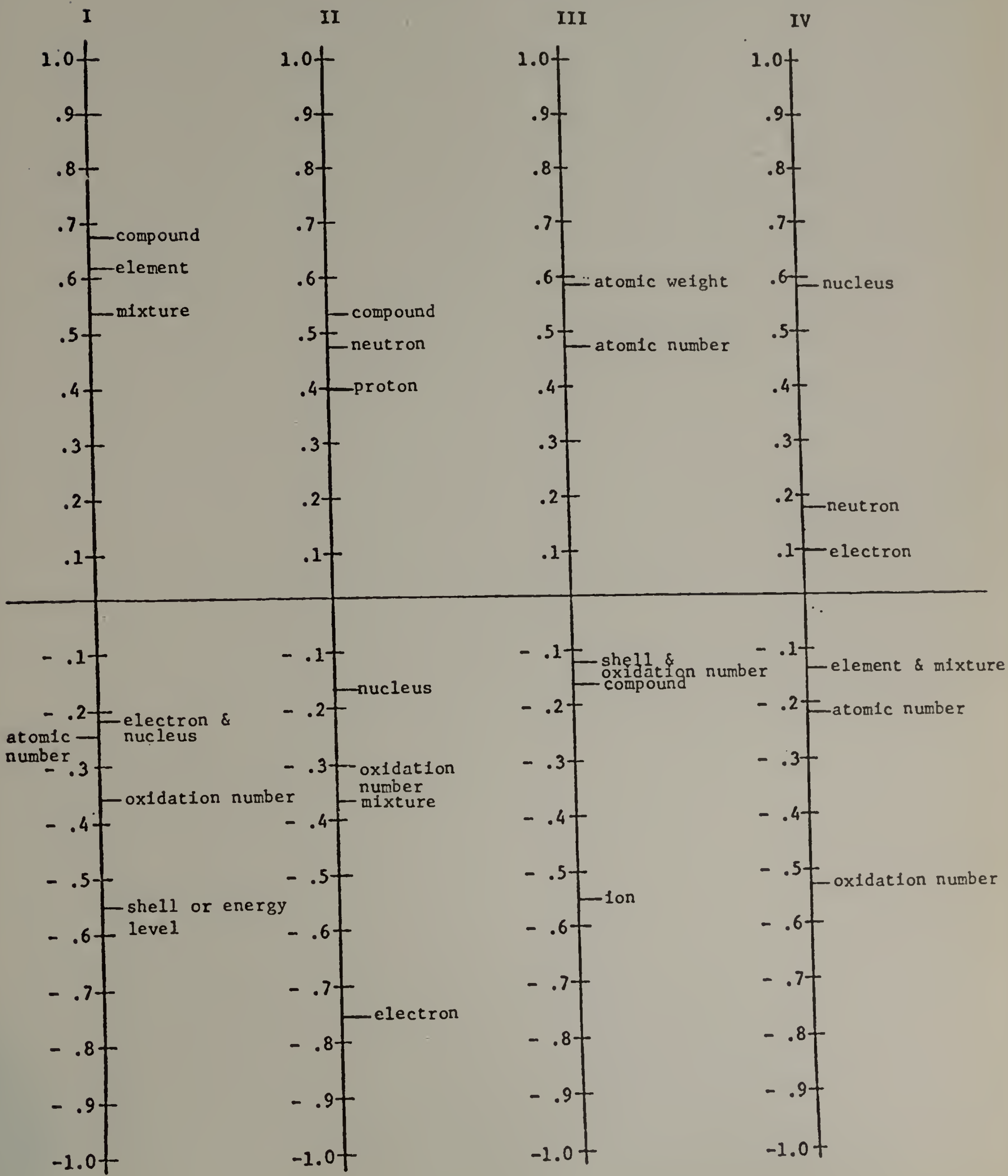
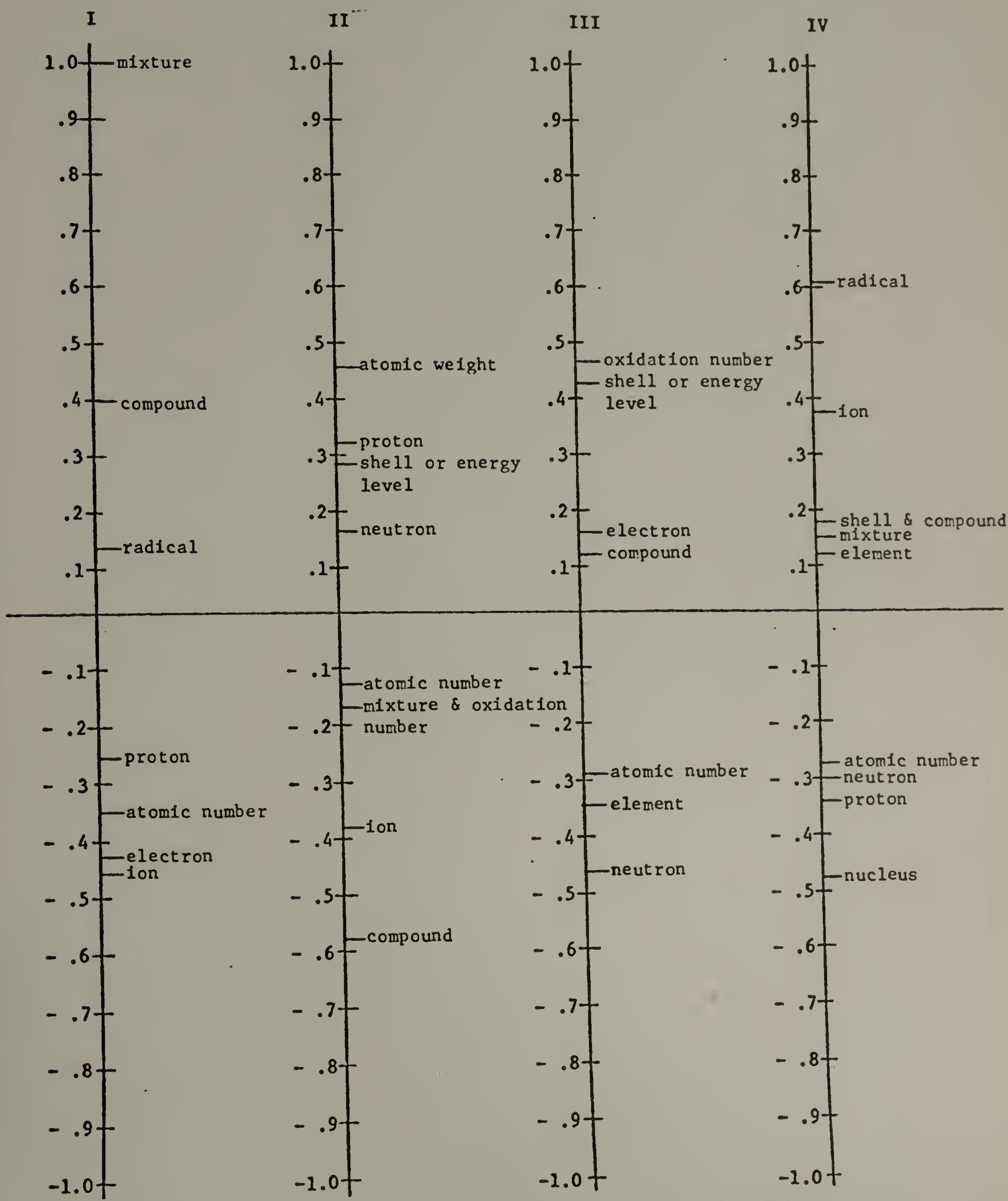


Figure 3.2.8 Stimulus dimensions for third post-instruction point of view for those students who scored in the bottom half on the module posttest (only stimuli whose loadings exceed ±.10 are represented).



and compound with the subgroup atomic number, proton, atomic weight, and neutron. The dimensions of this point of view appeared to have been defined in terms of logical subgroupings of the concepts.

Figure 3.2.7 presents the second post-instruction point of view for the low achievers. Dimension 1 contrasted the compound, element, mixture subgroup with two "electrical" concepts. Dimension 2 contrasted compound, neutron, and proton with the concepts electron, mixture, and oxidation number. Dimension 3 contrasted atomic weight and atomic number with ion. Dimension 4 contrasted nucleus with oxidation number and atomic number.

Figure 3.2.8 shows the stimulus dimensions for the third post-instruction point of view for the low achievers. Dimension 1 contrasted mixture and compound with ion, electron, atomic number, and proton. Dimension 2 contrasted atomic weight, proton, and neutron with compound and ion. Dimension 3 contrasted two "electrical" concepts with neutron, element, and atomic number. Dimension 4 contrasted radical and ion with nucleus, proton, neutron, and atomic number.

It should be pointed out that whereas the overall effect of instruction was to increase (slightly) the number of points of view with which the students perceived the chemical concepts, the most concise and interpretable solution was obtained from the high achievers. Both points of view associated with the high achieving students were interpretable, and the dimensionality (four dimensions and three dimensions) was lower than that of the pre-instruction points of view for the total class (five dimensions and four dimensions).

Figures 3.2.5a and 3.2.6a show the clustering solutions for the first and second post-instruction points of view for those students who scored in the top half of the posttest score distribution. The first point of view

Figure 3.2.5a Tree structure representing the hierarchical clustering solution for the first post-instruction point of view for those students who scored in the top half on the module posttest.

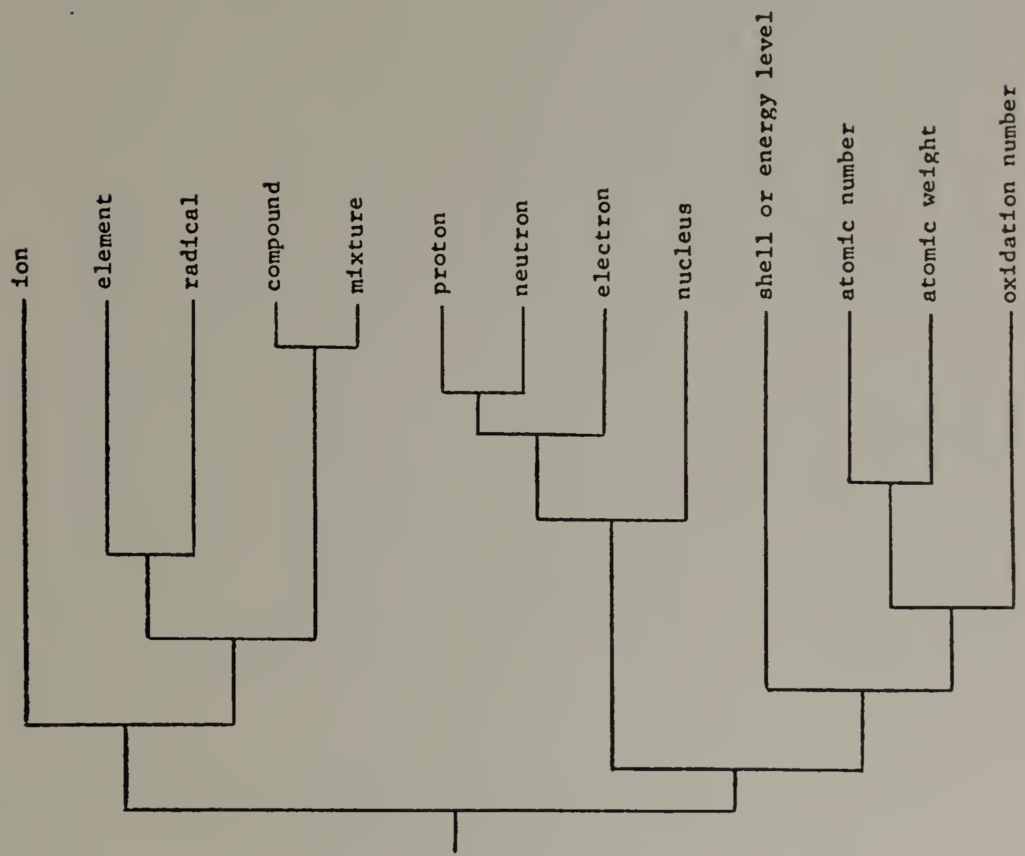
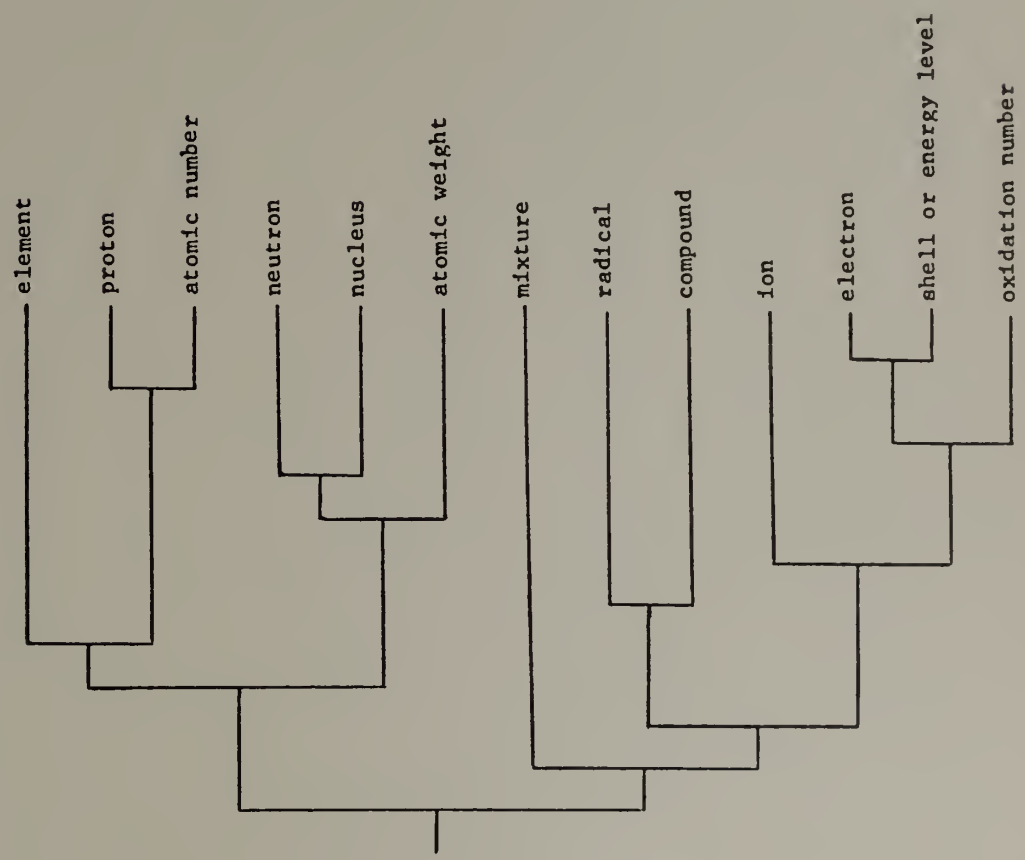


Figure 3.2.6a Tree structure representing the hierarchical clustering solution for the second post-instruction point of view for those students who scored in the top half on the module posttest.



(Figure 3.2.5a) showed the following clusters:

1) ion
2) element, radical, compound, mixture
3) proton, neutron, electron, nucleus
4) shell or energy level
5) atomic number, atomic weight, oxidation number.

The second point of view (Figure 3.2.6a) showed the following clusters:

1) element, proton, atomic number
2) neutron, nucleus, atomic weight
3) mixture
4) radical, compound
5) ion, electron, shell or energy level, oxidation number.

The second point of view split up the component parts of the atom. Proton was grouped with atomic number, and neutron and nucleus were grouped with atomic weight. Electron was found in a subgroup with the concepts shell or energy level, oxidation number, and ion. In the first point of view (Figure 3.2.5a) no such "electrical" cluster was found. The concepts atomic number and atomic weight were clustered with oxidation number and possibly shell or energy level. (The grouping of oxidation number with atomic weight was apparent in dimension 2 of Figure 3.2.5.) The component parts of the atom clustered together without being split up by the concepts atomic number and atomic weight. Electron was found in the component parts of the atom cluster and not in the "electrical" cluster.

At this point it would be well to mention that these clustering solutions were very helpful in interpreting the scaling configurations for the points of view. Many of the sub-clusters reappeared as the defining entities of the various dimensions of the configurations.

Figures 3.2.7a and 3.2.8a show the clustering solutions for the second and third post-instruction points of view for those students who scored in the bottom half of the posttest score distribution. Four clusters emerged from the second point of view (Figure 3.2.7a):

1) radical, element, compound, mixture
2) ion, proton, neutron, electron, nucleus
3) atomic number, atomic weight
4) shell or energy level, oxidation number.

The third point of view (Figure 3.2.8a) could be thought of as having five clusters:

1) compound, mixture
2) electron, proton, atomic number
3) element, nucleus, neutron, atomic weight
4) ion, radical
5) shell or energy level, oxidation number.

The third point of view had a clearly defined "electrical" subgroup (clusters 4 and 5). The component parts of the atom (electron, proton, neutron, and nucleus) are regrouped around the concepts atomic number and atomic weight. Electron was clustered with proton and atomic number, not with the "electrical" concepts. In the second point of view the component parts of the atom clearly emerged as a subgroup. Atomic number and atomic weight formed a cluster by themselves. The "electrical" subgroup was not as clearly defined as it was in the third point of view.

(e) General Comments

There appeared to be a slight tendency for instruction to reduce the dimensionality of the configurations. Four of the post-instruction points

Figure 3.2.7a Tree structure representing the hierarchical clustering solution for the second post-instruction point of view for those students who scored in the bottom half on the module posttest.

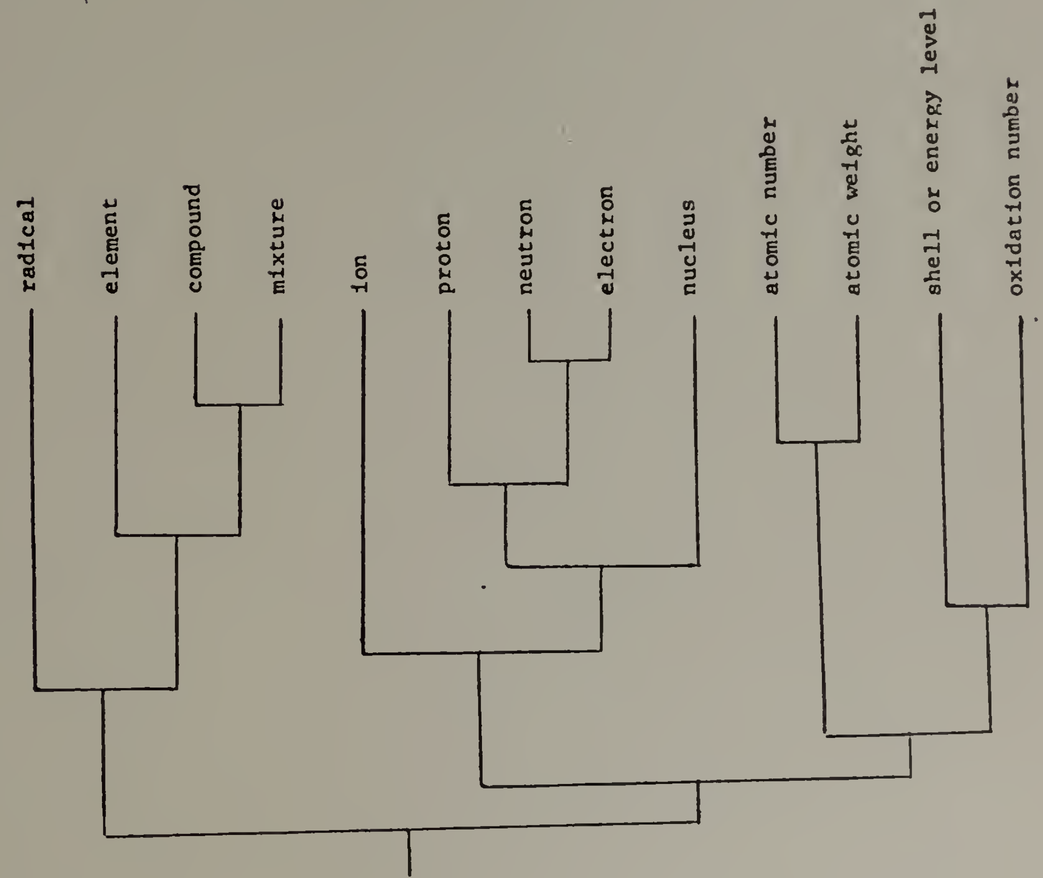
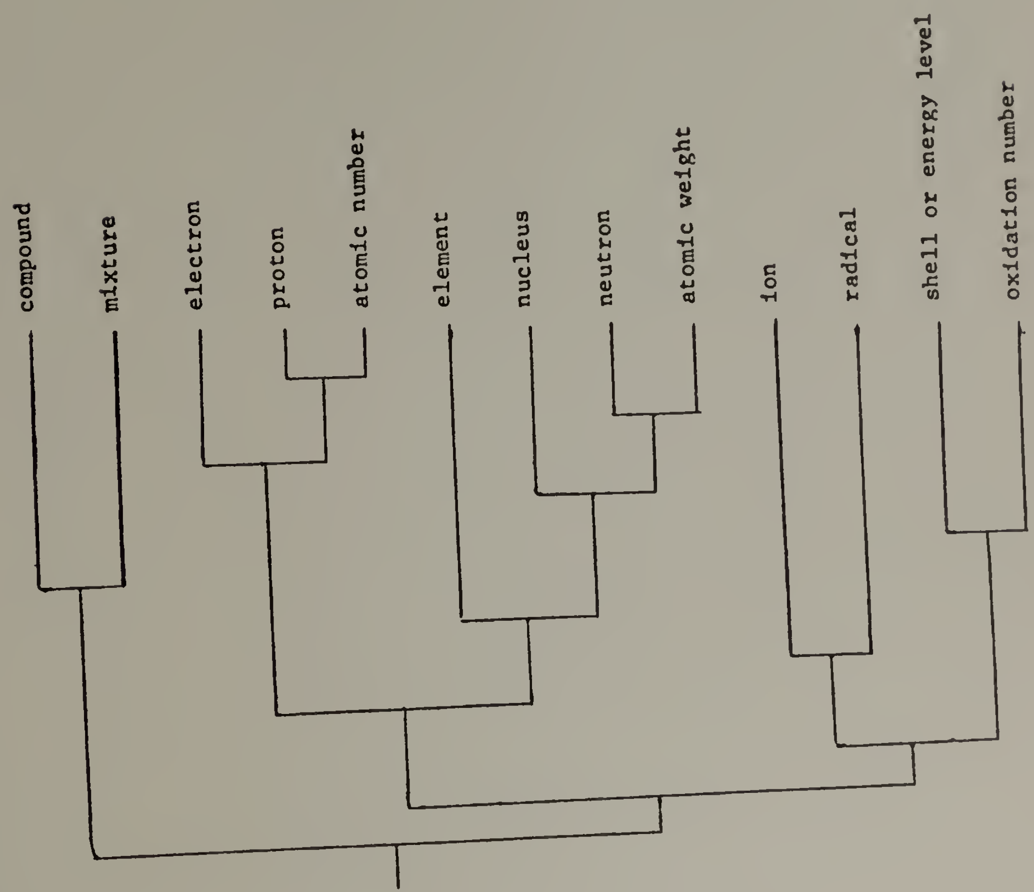


Figure 3.2.8a Tree structure representing the hierarchical clustering solution for the third. post-instruction point of view for those students who scored in the bottom half on the module posttest.



of view were represented in four dimensions, and two were represented in three dimensions. One pre-instruction point of view was represented in five dimensions, and one was represented in four dimensions.

There were differences between the high and low achieving students on the number of points of view and on the dimensionality of the concept spaces. The high achieving students tended to have the smallest number of points of view and lowest dimensionality. Perhaps more importantly, both points of view associated with the high achievers were "good" and interpretable.

It should be added that points of view and multidimensional scaling analyses were conducted on the four instructional treatment groups. Few differences were found between the groups, and because of this reason these analyses were not reported here.

The full potential of this technique as an evaluative tool will only be demonstrated after a criterion solution is obtained. The responses of chemical experts would lead to a criterion concept space. Student solutions could then be compared to this "correct" solution. If a method could be developed for determining the "distance" between two solutions comparisons would be put on a quantitative basis. Methods of instruction could be compared by examining the post-instruction concept spaces for the various instructional treatments. Those treatments producing concept spaces with the smallest discrepancies from the criterion space would be considered most effective.

This method of analysis may even have a place in ATI research. The differences between individual concept spaces and the criterion space could be computed. These differences would serve as measures of the criterion variable. Alternative instructional treatments could then be matched to specific learner aptitudes.

Latent Partition Analysis

(a) Latent Partition Analyses for the Total Class

An examination of the results of the pre- and post-instruction sorting data showed that the mean category cohesiveness values (average of the diagonal entries of the omega matrix) are not very useful in evaluating the effects of instruction. While the previously stated questions concerning mean category cohesiveness values seemed tenable from past studies, they have not proved particularly useful in this study.

LPA analysis of the pre-instruction sorting data for the whole class gave a mean category cohesiveness value of .67. The mean category cohesiveness value for the post-instruction sorting data for the whole class was .68. Thus the tendency of the sorters to place the chemical concepts into a given category was about the same before instruction as it was after instruction.

The most interesting aspect of the latent partition analyses centered on comparisons of the various phi matrices. The phi matrix specifies the composition of the latent categories. It has as many rows as there are concepts and as many columns as there are latent categories. A concept is said to belong to a particular latent category when its most substantial loading is in the column of that particular latent category. Some concepts have substantial loadings on several categories. These concepts are called ambiguous and, in this study, ambiguous concepts are indicated by secondary loadings in excess of .20.

Tables 3.2.5-3.2.8 and Figures 3.2.9 and 3.2.10 show the latent category composition of the pre- and post-instruction sorting data for the total class. As can be seen, four latent categories emerged from both the

Table 3.2.5

Concept Composition of Pre-Instruction Latent Categories [for the Total Class (N=284)]

Concept	Latent Category Number			
	1	2	3	4
nucleus	.65	-.02	.27	.04
ion	.61	-.15	.45	.04
electron	1.07	.04	-.07	-.01
proton	1.11	.02	-.05	-.04
neutron	1.12	.08	-.05	-.01
atomic weight	.01	.81	.12	.00
atomic number	.03	1.08	-.14	.02
oxidation number	-.09	.48	.70	-.06
radical	-.01	-.22	1.03	.14
shell or energy level	-.08	.01	1.21	-.08
element	.09	.08	-.05	.84
compound	-.02	-.01	-.02	1.07
mixture	-.06	-.02	.04	1.07

Table 3.2.6

Summary of the Entries in Phi
for the Pre-Instruction Sorting
Data for the Total Class (N=284)

Latent Category	Number of Concepts Assigned	Magnitude of Primary Loadings			Number of Secondary Loadings (greater than or equal to .20)
		90+	60-89	30-59	
1	5	3	2	0	2
2	2	1	1	0	0
3	3	2	1	0	1
4	3	2	1	0	0

Table 3.2.7

Concept Composition of Post-Instruction Latent Categories [for the Total Class (N=265)]

Concept	Latent Category Number			
	1	2	3	4
nucleus	.82	.03	.13	.00
electron	.84	-.03	.22	-.04
neutron	1.06	.00	-.06	.01
proton	1.06	.03	-.11	.03
atomic weight	.05	1.10	-.17	.02
atomic number	.03	1.14	-.17	.03
oxidation number	-.18	.78	.53	-.08
shell or energy level	.11	.38	.56	-.08
radical	-.16	.03	.87	.23
ion	.05	-.20	1.25	-.07
element	.03	.03	.10	.82
mixture	.00	.00	-.06	1.04
compound	.00	.00	-.01	1.06

Table 3.2.8

Summary of the Entries in Phi for the Post-Instruction Sorting Data for the Total Class (N=265)

Latent Category	Number of Concepts Assigned	Magnitude of Primary Loadings			Number of Secondary Loadings (greater than or equal to .20)
		90+	60-89	30-59	
1	4	2	2	0	1
2	3	2	1	0	1
3	3	1	1	1	2
4	3	2	1	0	0

Figure 3.2.9 Diagram of latent category composition of pre-instruction sorting data for the total class (N=284) [concepts with high secondary loadings are indicated by arrows, the head of the arrow indicating the category on which the concept has the high (greater than .20) secondary loading].

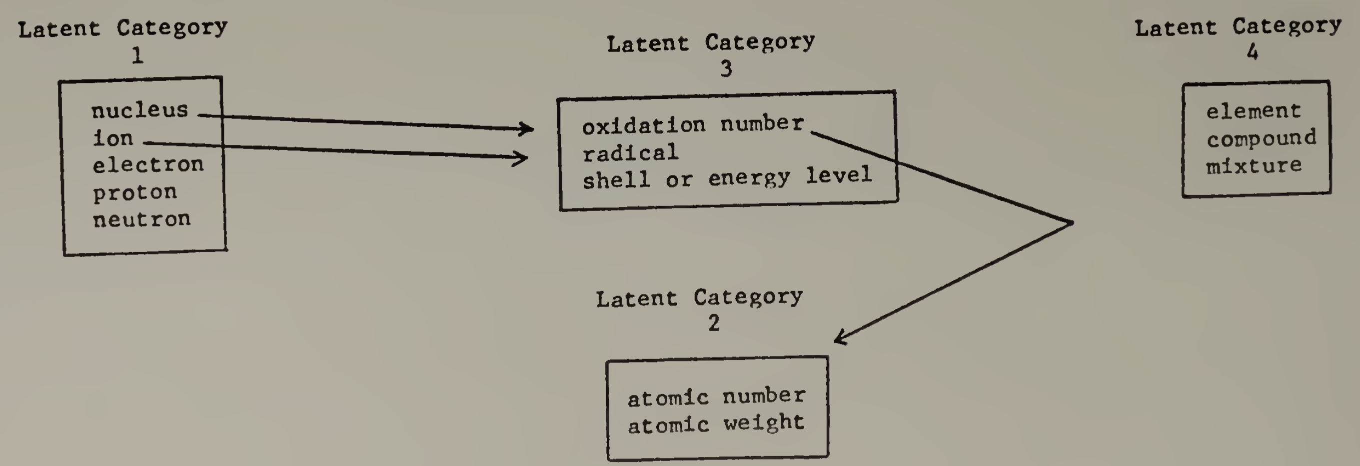
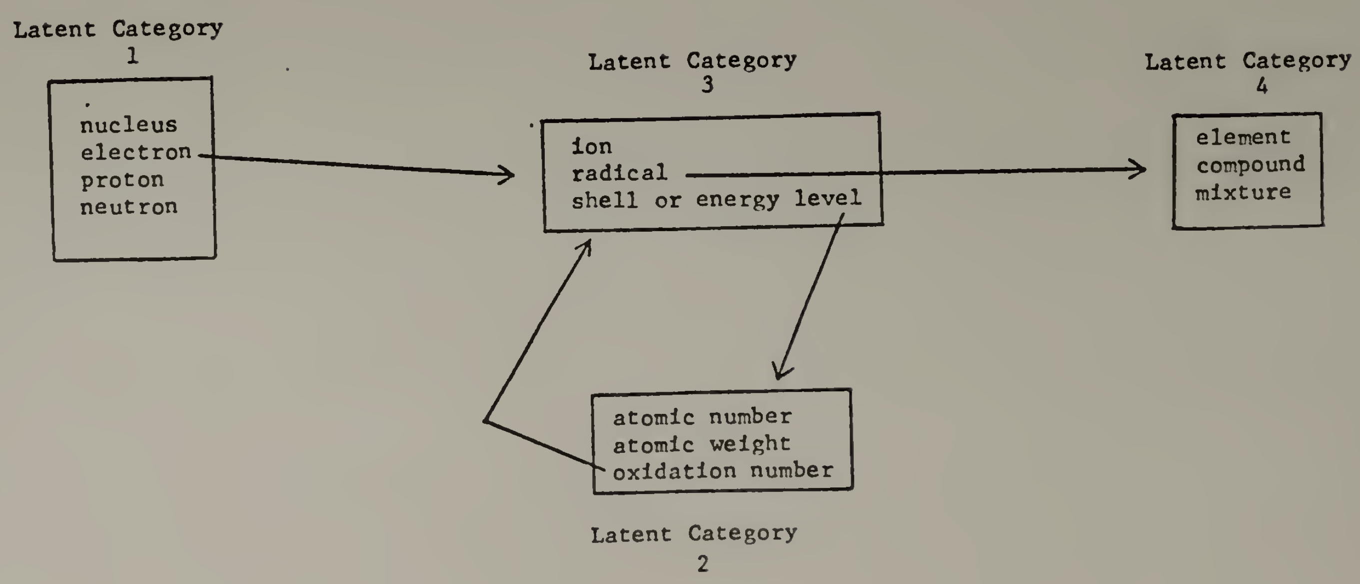


Figure 3.2.10 Diagram of latent category composition of post-instruction sorting data for the total class (N=265) [concepts with high secondary loadings are indicated by arrows, the head of the arrow indicating the category on which the concept has the high (greater than .20) secondary loading].



pre- and post-instruction sorting data. The concepts ion and oxidation number were the only concepts to be placed in different latent categories after instruction. Initially ion was found in latent category 1, a category that could be considered to represent the component parts of an atom (Figure 3.2.9). After instruction ion was placed in latent category 3, a category that appears to be characterized by the electrical nature of its concepts (Figure 3.2.10). The concept oxidation number was initially found in latent category 3. After instruction its highest loading was on latent category 2, the category defined by the two concepts atomic number and atomic weight.

One ambiguity was removed by instruction and several others were created. After instruction the concept nucleus no longer loaded substantially on latent category 3, the so-called "electrical" latent category (Figure 3.2.10). Figure 3.2.10 also reveals the new ambiguities present after instruction. Electron loaded substantially on latent category 3. The concept radical was associated with latent category 4, and the concept shell or energy level had a substantial secondary loading on latent category 2.

Tables 3.2.9 and 3.2.10 show the lower diagonal entries of the omega matrices for the pre- and post-instruction LPA analyses of the sorting data for the total class. A diagonal entry represents the probability that a pair of concepts from that latent category will be sorted into the same manifest category. Here the concern is with the off-diagonal elements, specifically the off-diagonal elements that are underlined. (An off-diagonal element was underlined when its magnitude reached the arbitrary level of .10). These entries can be used to indicate the most probable latent

Table 3.2.9

Probabilities of Latent Category Confusions
[Pre-Instruction Sorting Data
for the Total Class (N=284)]

Latent Category	Latent Category 1	2	3	4
1	.69			
2	.07	.92		
3	.12	.33	.34	
4	.08	.07	.11	.73

Table 3.2.10

Probabilities of Latent Category Confusions
[Post-Instruction Sorting Data
for the Total Class (N=265)]

Latent Category	Latent Category 1	2	3	4
1	.83			
2	.14	.71		
3	.20	.18	.38	
4	.03	.04	.14	.78

category confusions. That is, these off-diagonal elements give the probability that any particular pair of concepts from the two different latent categories will be put into the same manifest category. Using the arbitrary cutting point of .10 to signify a significant confusion, the most probable latent category confusions for the pre- and post-instruction sorting data for the total class were constructed. These appear in Figures 3.2.11 and 3.2.12. It can be seen that the pre- and post-instruction confusions are very similar. Yet after instruction more confusion seems to occur in the categorization of concepts from latent category 1 (<u>nucleus</u>, <u>electron</u>, <u>proton</u>, and <u>neutron</u>). Its confusion probabilities with latent categories 2 and 3 both increase. The category that is least isolated and, thus, most confused with the other categories is category 3, the so-called "electrical" category. The strongest confusion probability occurs between categories 2 (<u>atomic number</u> and <u>atomic weight</u>) and 3 [(<u>oxidation number</u>, <u>radical</u>, <u>shell or energy level</u>, and <u>ion</u> (in the pre-instruction case)]. Latent category 4 appears to be the most isolated of the latent categories.

(b) <u>Latent Partition Analyses for the High and Low Achievers</u>

The pre-instruction sorting data was partitioned into two groups. One group consisted of those students who scored on the top half of the Module Pretest and the other group consisted of those students who scored on the bottom half of the Module Pretest. The post-instruction sorting data was partitioned into two groups on the basis of Module Posttest scores. The top half of the Module Posttest scoring distribution was in one group while the bottom half of the distribution was in the other group.

The differences between the mean category cohesiveness values for the high and low achieving students were not clear-cut. An examination of

Figure 3.2.11 Latent categories and the more probable confusions [pre-instruction sort for the total class (N=284)].

Category	The More Probable Confusions
1	
2	1 ----- .12 ----- 3 ----- .11 ----- 4
3	.33 ¦
4	2

Figure 3.2.12 Latent categories and the more probable confusions [post-instruction sort for the total class (N=265)].

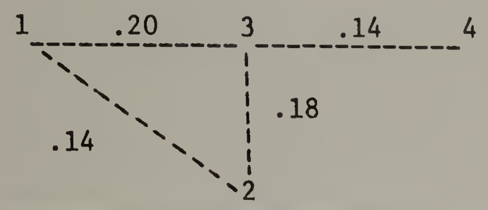
Category	The More Probable Confusions
1	
2	
3	
4	

Figure 3.2.13 shows that the mean category cohesiveness values tended to be of the same magnitude irrespective of student achievement level for the post-instruction sorting data. For the pre-instruction sorting data the high achieving group of students tended to have a higher mean category cohesiveness value (Figure 3.2.13).

Table 3.2.11 and Figure 3.2.14 show the composition of the latent categories and the ambiguities for the pre-instruction sorting data of the group of students who scored on the top half of the Module Pretest. Table 3.2.12 and Figure 3.2.15 show the latent category composition and the ambiguities of the group who scored on the bottom half of the Module Pretest. It can be seen that the high achieving students form four latent categories while the low achieving students form only three. The concepts ion, nucleus, and oxidation number are ambiguous for the high achievers. The concept radical is the only ambiguous concept for the low achieving students. It had about equal loadings on latent categories 2 and 3 (Figure 3.2.15).

Tables 3.2.13 and 3.2.14 and Figures 3.2.16 and 3.2.17 present similar information for the post-instruction sorting data partitioned on the basis of Module Posttest scores. Two of the ambiguous concepts for the high achieving students came from latent category 2 (the "electrical" latent category). These were the concept radical which has a high secondary loading on latent category 4 and the concept shell or energy level which has a high secondary loading on latent category 2. Electron, a concept from latent category 1, was also ambiguous as it had a high secondary loading on latent category 3. The concept oxidation number was very ambiguous to this group of students. It loaded almost equally on latent categories 2 and 3.

A greater variety of ambiguities was present in the sorting data of the

Figure 3.2.13 Mean category cohesiveness for chemical concepts before and after instruction according to achievement level.

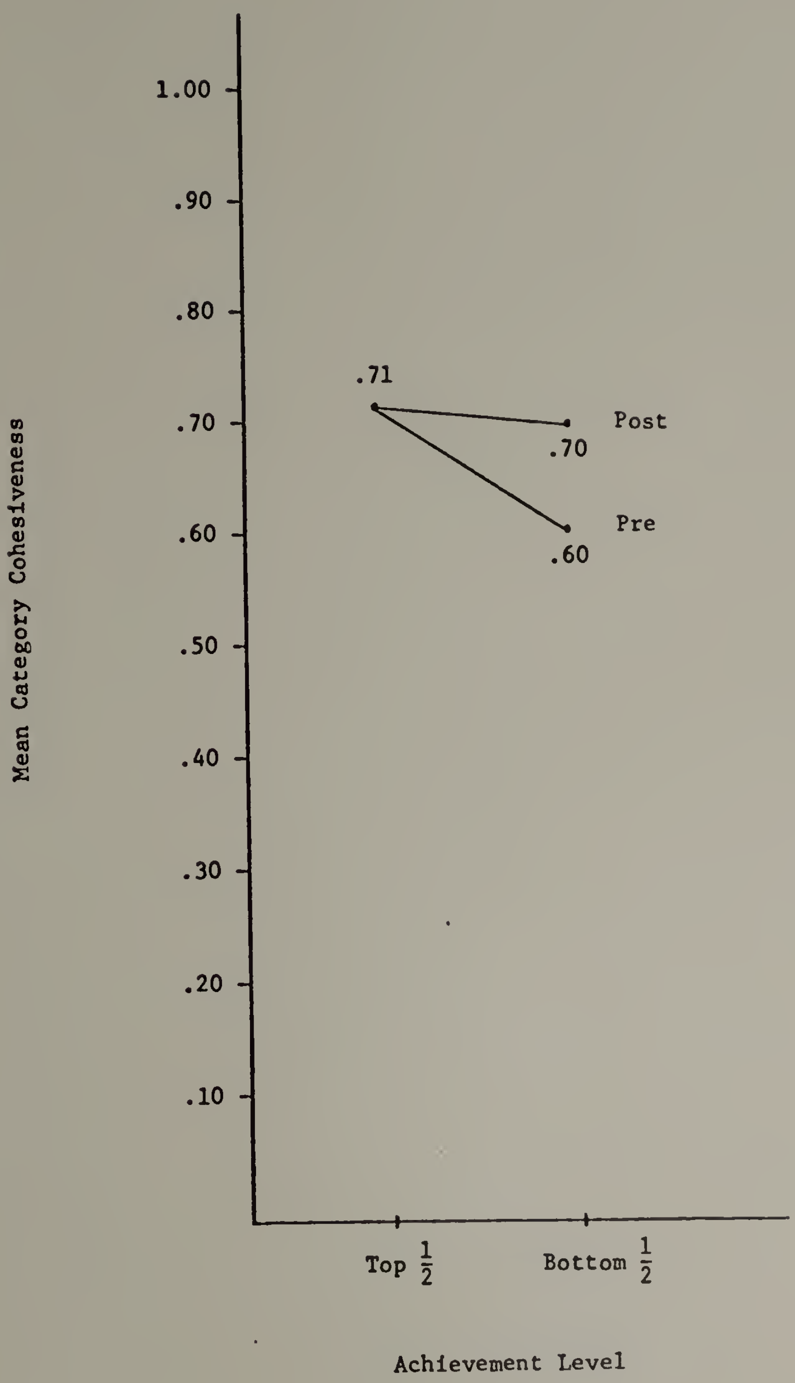


Table 3.2.11

Concept Composition of Pre-Instruction Latent Categories [for the Group Who Scored in the Top Half on the Module Pretest (N=119)]

Concept	Latent Category Number			
	1	2	3	4
nucleus	.54	-.18	.48	.09
ion	.65	-.04	.30	.07
electron	1.05	.00	.00	-.03
proton	1.13	.05	-.11	-.04
neutron	1.11	.04	-.12	-.02
atomic weight	.03	.89	.03	.00
atomic number	.03	1.00	-.07	.02
oxidation number	-.11	.44	.78	-.05
radical	-.01	-.23	1.05	.13
shell or energy level	-.08	.05	1.18	-.10
element	.11	.10	-.07	.84
compound	-.04	-.02	.01	1.08
mixture	-.05	-.01	.01	1.07

Figure 3.2.14 Diagram of latent category composition of pre-instruction sorting data for those students who scored in the top half on the module pretest (N=119) [concepts with high secondary loadings are indicated by arrows, the head of the arrow indicating the category on which the concept has the high (greater than .20) secondary loading].

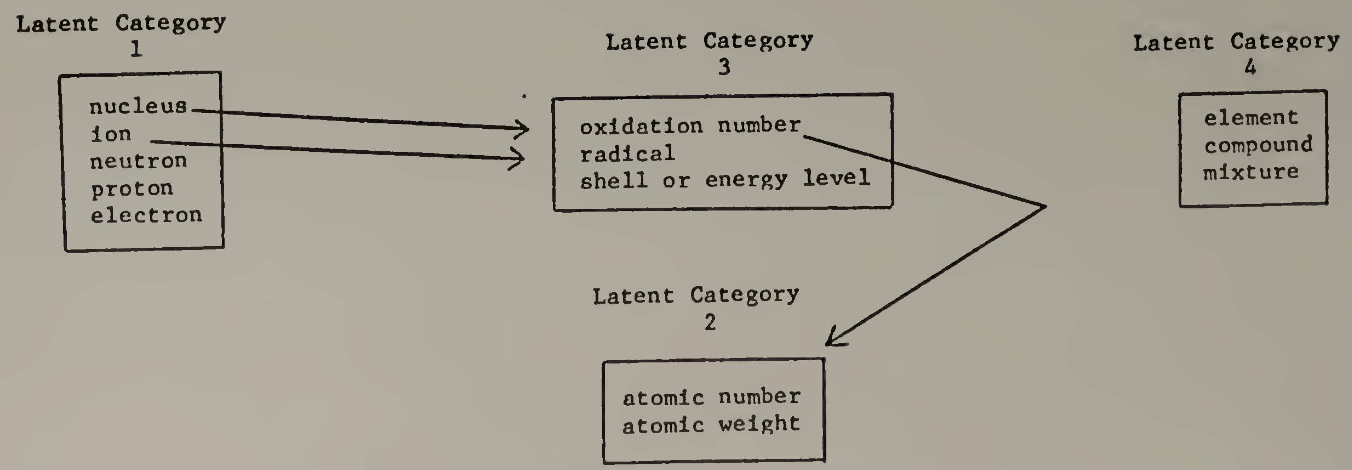


Table 3.2.12

Concept Composition of Pre-Instruction Latent Categories [for the Group Who Scored in the Bottom Half on the Module Pretest (N=109)]

Concept	Latent Category Number		
	1	2	3
nucleus	.72	.10	.05
ion	.77	.05	.06
electron	1.12	.00	-.02
proton	1.15	-.06	-.03
neutron	1.14	-.05	-.03
atomic weight	-.01	1.10	-.02
atomic number	-.03	1.22	-.04
oxidation number	-.03	1.07	-.02
radical	.11	.29	.29
shell or energy level	-.08	.64	.06
element	.10	.04	.82
compound	-.02	-.05	1.14
mixture	-.06	-.02	1.15

Figure 3.2.15 Diagram of latent category composition of pre-instruction sorting data for those students who scored in the bottom half on the module pretest (N=109) [concepts with high secondary loadings are indicated by arrows, the head of the arrow indicating the category on which the concept has the high (greater than .20) secondary loading].

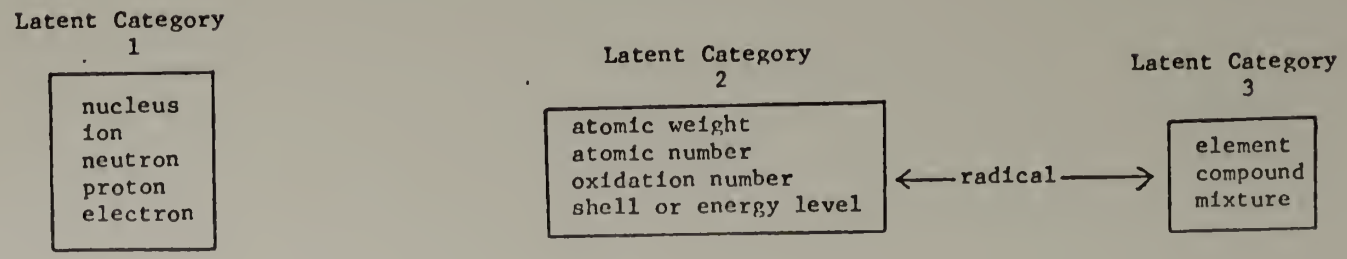


Table 3.2.13

Concept Composition of Post-Instruction Latent Categories [for the Group of Students Who Scored in the Top Half on the Module Posttest (N=108)]

Concept	Latent Category Number			
	1	2	3	4
nucleus	.90	.03	.02	.02
electron	.83	-.04	.24	-.05
neutron	1.06	.00	-.01	.01
proton	1.01	.02	-.07	.01
atomic number	.00	1.10	-.14	.03
atomic weight	.05	1.07	-.16	.03
shell or energy level	.05	.37	.63	-.10
radical	-.07	-.08	.79	.32
ion	.03	-.21	1.26	-.06
oxidation number	-.15	.66	.67	-.09
element	.03	.08	.01	.87
mixture	.00	.00	-.06	1.06
compound	-.01	-.01	.03	1.05

Table 3.2.14

Concept Composition of Post-Instruction Latent Categories [for the Group of Students Who Scored in the Bottom Half on the Module Posttest (N=93)]

Concept	Latent Category Number			
	1	2	3	4
nucleus	.64	.00	.47	-.07
electron	.96	-.01	.09	-.02
neutron	1.03	.00	.00	.00
proton	1.12	.04	-.24	.07
atomic number	.04	1.06	-.12	.04
atomic weight	.03	1.08	-.10	.01
oxidation number	-.11	.85	.28	-.04
shell or energy level	.12	.34	.66	-.10
radical	-.27	.01	1.21	.07
ion	.22	-.20	1.01	-.07
element	-.02	-.02	.27	.77
mixture	.01	.01	-.04	1.00
compound	.04	.02	-.05	1.02

Figure 3.2.16 Diagram of latent category composition of post-instruction sorting data for those students who scored in the top half on the module posttest (N=108) [concepts with high secondary loadings are indicated by arrows, the head of the arrow indicating the category on which the concept has the high (greater than .20) secondary loading].

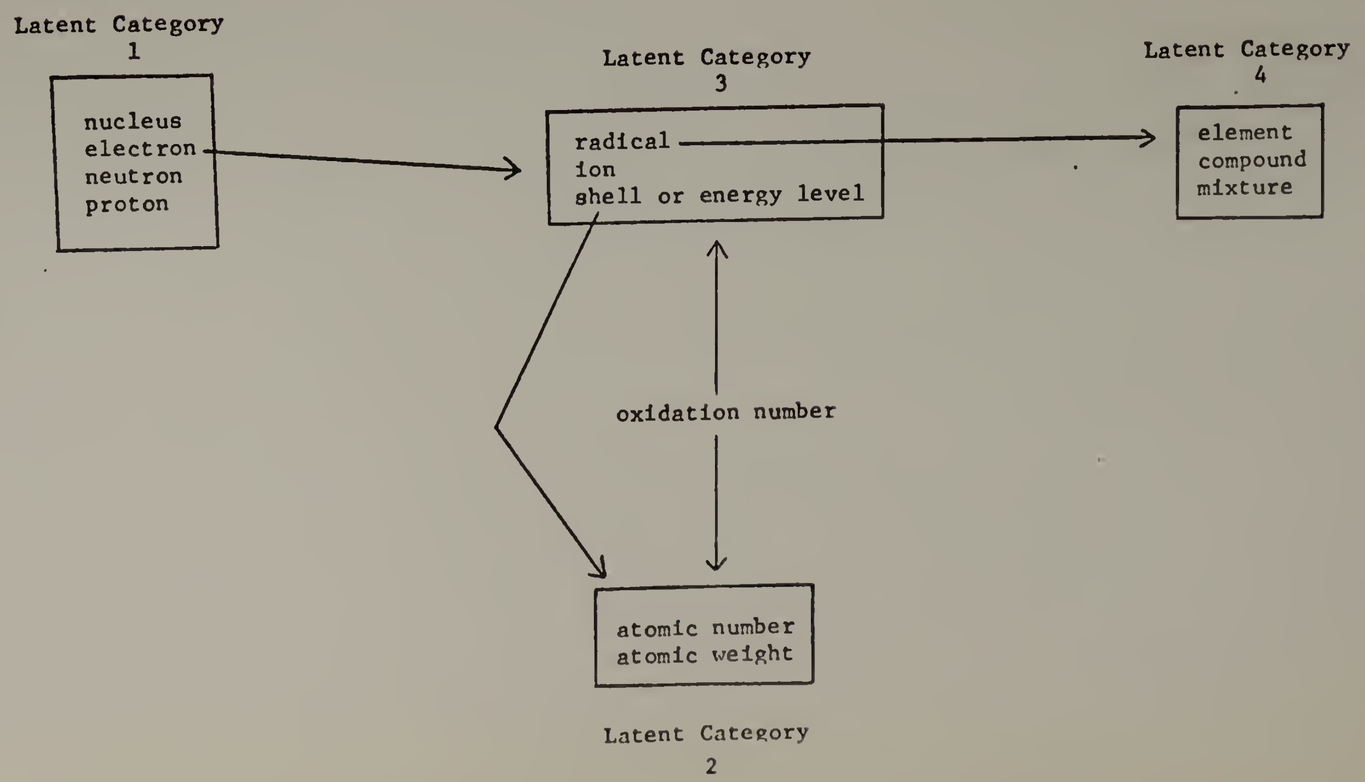
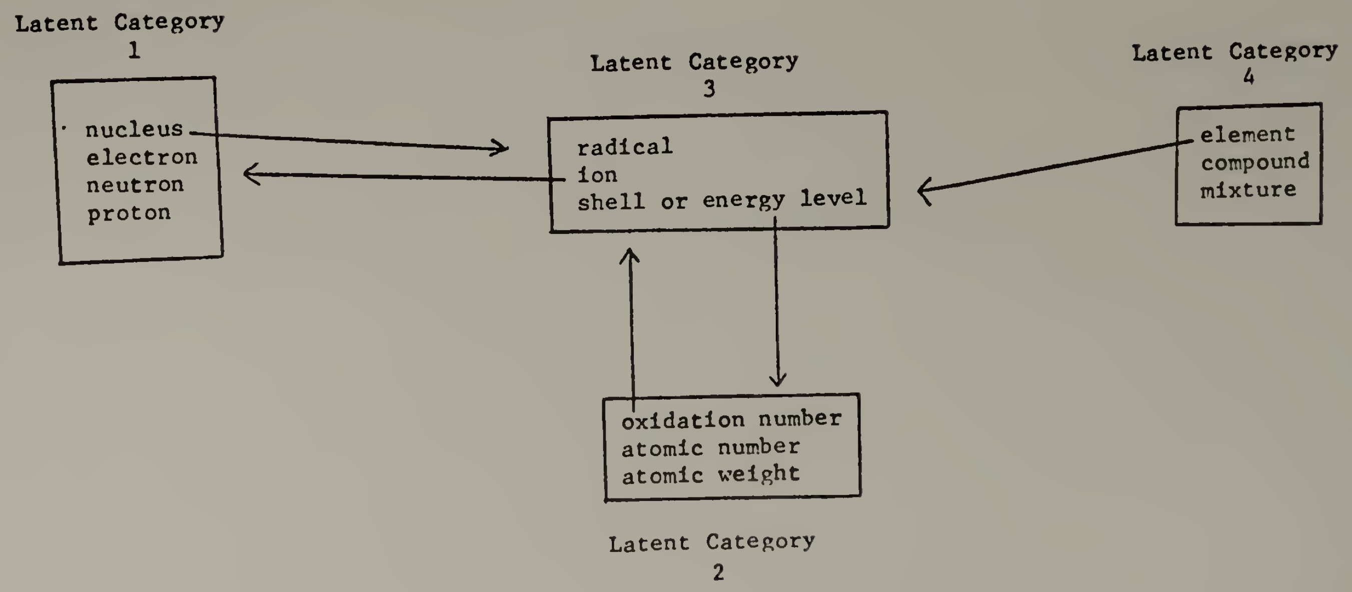


Figure 3.2.17 Diagram of latent category composition of post-instruction sorting data for those students who scored in the bottom half on the module posttest (N=93) [concepts with high secondary loadings are indicated by arrows, the head of the arrow indicating the category on which the concept has the high (greater than .20) secondary loading].



low achievers on the Module Posttest (Figure 3.2.17). As with the high achievers, several ambiguities involved the "electrical" concepts (shell or energy level and oxidation number) and were directed between latent categories 2 and 3. The ambiguous concept from latent category 1 for the high achievers was electron; for the low achievers it was nucleus. With the high achieving students, latent category 4 was completely isolated. This was not the case with the low achievers. The concept element was ambiguous to them, having a high secondary loading on latent category 3.

It appears that the high achieving students have focused their ambiguities around "electrical" concepts (electron from latent category 1, shell or energy level and radical from latent category 2, and oxidation number, which has substantial loadings on both categories 2 and 3). The low achieving students also have problems with concepts like shell or energy level and oxidation number, although to a lesser extent. Yet their ambiguities are not limited to "electrical" concepts. They experience ambiguities with "non-electrical" concepts from latent categories 1 and 4 (nucleus and element).

(c) General Comments

In addition to the analyses reported latent partition analyses were carried out on each of the instructional treatment groups. Because differences between treatment groups were slight, these analyses were not included. Lack of a criterion population on which to base a criterion partitioning also limited these analyses. Without a common criterion on which to focus, comparisons between groups lack qualitative value.

The alert reader is perhaps wondering why the omega matrices were not interpreted for the high and low achieving students. This decision was completely arbitrary. It was felt that the role of the omega matrix in interpreting LPA data could best be illustrated with the total class data. The addition of this type of interpretation for the high and low achieving students would only tend to divert the focus from the main analyses.

The previous sections have shown how data analyzed by the LPA model can be useful to the understanding of instructional effects. The interpretations of the phi and omega matrices permit the determination of the groupings of concepts into categories, the degree of definition of the categories, and the possible confusions between the categories. A representation of the cognitive structure results from these interpretations.

An important area for future research would be to obtain a criterion solution. The sorts of a population of chemical experts would provide criterion interpretations of the phi and omega matrices. If a measure of goodness of fit was developed, the students' solution could be compared to that of the experts. This would allow the determination of the effect of instruction on the cognitive space of the students. With such a criterion solution groups or clusters of students could be isolated and their discrepancies from the criterion solution could be compared.

3.3 Generalizability

Generalizability of the ATI Study

The intent of the discussion in section 1.3 was to caution the reader of ATI studies against the over-generalization of results. While progress may be slightly delayed, failure to show restraint can result in the unintelligible mass of information that the ATI literature has become.

The ATI aspect of this study represented action research rather than basic or experimental research. It was an on-the-job type of problem solving whose benefits would have been accurable to the school in which the research took place.

From their extensive literature review, Cronbach and Snow (1969) extracted several tentative conclusions. Using these conclusions in the context of process analysis is the proper way to design treatments for an ATI study. In the area of aptitude variables, Cronbach and Snow (1969) suggest pitting a measure of general ability against various measures of specific abilities.

It is acknowledged that these are the guidelines that one should use in designing an ATI study. In fact, the design of this study did parallel these guidelines in many respects. The important point, however, is that such factors as time, money, available tests, and the school situation were primarily responsible for the selection of aptitude measures and the design of treatments. If any interactions were found, the appropriate aptitude measures and treatments could easily have been implemented by the school system. Yet from the standpoint of basic research, Cronbach and Snow's guidelines could have been more closely followed, and more time and money could have been spent in treatment designs.

The existence of different learning styles has to be labelled as elusive and tenuous. Conflicting results have been reported in the ATI literature. Perhaps the actual causative interactive variables are "hidden" in the overall instructional treatment, the treatment acting as a moderator variable. What is needed is more action research in individual schools. If significant interactions in particular programs are found, the various instructional methods can be implemented. Then if we are dealing with "hidden" interactive variables, such variables could be more easily diagnosed in an on-going program as comparison and continual observation would be possible.

After initially downgrading this ATI study because it may have been lacking in optimal research design, it is now felt that it may have been of more eventual theoretical worth because of its on-the-job nature. At the very least, we would not be purposely adding more possible conflicting information to an already tenuous ATI theory.

Generalizability of the Evaluative Study

In assessing the generalizability of the evaluative component (the utilization of the various psychometric models to monitor instruction) of this study several aspects must be considered.

First of all one must remember that the students in the study were from a middle class area. Thus, strictly speaking, generalizability of any results to other high school students is not possible.

Assignments of students to instructional treatments were random within each class period. This should have controlled for extraneous independent variables. Yet this brings to light another aspect. How representative are the variables? Specifically how representative was the ninth grade science instruction? The individual instructional treatments were unique to this study and differences that were found among them should not be attributed to any specific features (such as audiotapes or programmed material) possessed by the treatments. Are we positive that all of the salient variables on which the treatments differ have been isolated? The answer is no, and until we are no statements can be made about the generalizability of any results to specific types of treatments in general.

Another factor limiting the generalizability of the results was the context effect that Torgerson (1965) speaks about, was acting with the stimuli. That is, the overall variation in the sorting and similarity data is relative to the particular stimulus set selected.

One example was encountered while completing the Student Sorting Task. While doing the sorting of the concepts it was realized that different categorizations would have been formed if the sorter had focused on the component parts of an atom rather than on the concepts atomic weight and atomic number. While either sort has theoretical foundations, this problem would not have been encountered if the concepts of atomic weight and atomic number were not included. (Yet if they were not included the concept domain of the instructional module would not have been adequately spanned.)

While completing the Similarity Judgment Questionnaire the concept pair, proton-electron, was encountered. Some judges rated the concepts as being moderately similar. Yet it was realized that if the concept pair was seen removed from the questionnaire the concepts would be considered as almost complete opposites. A proton represents a positively charged particle while an electron represents a negatively charged particle. Why were the concepts rated as being moderately similar? They were rated as moderately similar because the surrounding concept pairs (such as atomic weight-oxidation number and mixture-energy level) represented completely dissimilar concepts. At least it could be said that both protons and electrons are several of the particles of which an atom is composed.

Thus the resulting latent categorizations and concept spaces appear to be unique to this particular set of concepts. The addition or removal of several concepts may alter the results. Yet this should not affect the usefulness of these techniques for evaluating instruction. Rather, extreme care must be used in selecting the appropriate stimuli (concepts). The stimuli selected must adequately represent the content domain covered by the instruction. If they do then generalization would not really be a problem because one would not want to generalize his results to other, less adequate, groups of concepts.

3.4 Student Questionnaire Results

An important consideration in the evaluation of instructional treatments concerns how well the treatments are accepted by the students. The questionnaire used in our study represented an attempt at determining how the students in the ninth grade science program felt about instructional materials in general, and, specifically, the instructional materials used in "The Structure of Matter" module.

After initially indicating which instructional treatment they were in during "The Structure of Matter Module", the students responded to twelve statements concerning instructional materials by circling one of five alternatives. The alternatives were STRONGLY AGREE, AGREE, UNDECIDED, DISAGREE, AND STRONGLY DISAGREE.

Using a code of "1" to represent a STRONGLY AGREE response, a "2" to represent an AGREE response, and so on up to a "5" for a STRONGLY DISAGREE response, it was then possible to compute the means and standard deviations of the student judgments for the twelve statements. This was done separately for the students in each of the four instructional treatments and again for the total group of students. The results are reported in Table 3.4.1.

A chi-square analysis of the student responses to the twelve statements was conducted to determine if the four instructional treatment groups differed in their response patterns. Significant differences were found among the groups for three of the statements. These statements and the percentage of responses in each category of the statements for each instructional treatment group are reported in Table 3.4.2. The results for the remaining statements (which had nonsignificant discrepancies between the observed and expected response frequencies) are reported in Table 3.4.3.

Table 3.4.1

Means and Standard Deviations of Questionnaire Responses for the Four Instructional Treatment Groups and for the Total Group of Students

STATEMENT NUMBER	TREATMENT 1 (Teaching)		TREATMENT 2 (Media)		TREATMENT 3 (Reading)		TREATMENT 4 (P.I.)		TOTAL GROUP	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
1	2.7	1.00	2.3	.75	2.6	.89	2.1	.59	2.4	.83
2	3.6	.72	3.9	.66	3.9	.57	3.8	.72	3.8	.67
3	3.8	.76	3.3	.94	3.3	.95	3.6	.92	3.5	.93
4	2.4	.76	3.0	1.00	3.0	1.10	2.3	.97	2.7	1.10
5	2.3	.98	2.9	1.20	2.7	1.10	3.0	1.10	2.8	1.10
6	3.0	.96	3.0	1.10	2.8	.94	3.4	.94	3.0	1.00
7	2.5	1.00	2.3	.89	2.4	.89	2.0	.64	2.3	.87
8	2.3	1.10	1.9	.82	2.2	1.10	2.0	.88	2.1	.95
9	3.5	.87	3.1	1.10	3.1	1.20	3.2	1.10	3.2	1.10
10	2.0	.71	2.0	.66	1.8	.69	2.0	.77	2.0	.71
11	3.7	.89	3.7	.94	3.8	1.10	3.7	.90	3.7	.95
12	2.3	.69	2.2	.79	2.3	.78	2.3	.78	2.3	.77

Table 3.4.2

Percentage of Responses in Each Instructional Treatment Group to Questionnaire Statements for Which Inter-Group Response Patterns Differed Significantly

Statement	Instructional Treatment Group 1[a] (N=25)	2[b] (N=44)	3[c] (N=49)	4[d] (N=45)
1. The instructional materials in the "Structure of Matter" module were at about the correct level of difficulty.				
(a) STRONGLY AGREE	0	9	4	9
(b) AGREE	67	53	59	71
(c) UNDECIDED	8	31	14	18
(d) DISAGREE	17	7	23	2
(e) STRONGLY DISAGREE	8	0	0	0
Number of Omits = 3				
$\chi^2 = 29.87$ with 12 df				
3. The instructional materials in the "Structure of Matter" module were too hard.				
(a) STRONGLY AGREE	0	5	0	0
(b) AGREE	8	16	31	14
(c) UNDECIDED	12.5	26	16	27
(d) DISAGREE	67	51	49	43
(e) STRONGLY DISAGREE	12.5	2	4	16
Number of Omits = 3				
$\chi^2 = 22.15$ with 12 df				
4. I liked the instructional materials in the "Structure of Matter" module about as well as instructional materials can be liked.				
(a) STRONGLY AGREE	8	4.5	4	20
(b) AGREE	56	33	41	44
(c) UNDECIDED	28	23	20	20
(d) DISAGREE	8	35	25	16
(e) STRONGLY DISAGREE	0	4.5	10	0
Number of Omits = 1				
$\chi^2 = 24.14$ with 12 df				

a Treatment consisting of teacher lectures.
b Treatment consisting primarily of audiotapes and worksheets.
c Treatment consisting of a reading handout.
d Treatment consisting of programmed instruction booklets.

Table 3.4.3

Percentage of Responses in Each Instructional Treatment Group to Questionnaire Statements for Which Inter-Group Response Patterns Did Not Differ Significantly

Statement	Instructional Treatment Group			
	1[a] (N=25)	2[b] (N=44)	3[c] (N=49)	4[d] (N=45)
2. The instructional materials in the "Structure of Matter" module were too easy.				
(a) STRONGLY AGREE	4	0	0	0
(b) AGREE	0	5	4	5
(c) UNDECIDED	29	14	8	23
(d) DISAGREE	67	70	80	59
(e) STRONGLY DISAGREE	0	11	8	13
Number of Omits = 3				
χ^2 = 17.06 with 12 df				
5. Students learn better from "live" teachers than from films, tapes, programmed booklets, or textbooks.				
(a) STRONGLY AGREE	20	11.5	12	9
(b) AGREE	48	33	33	22
(c) UNDECIDED	16	23	31	35
(d) DISAGREE	16	21	18	27
(e) STRONGLY DISAGREE	0	11.5	6	7
Number of Omits = 1				
χ^2 = 12.03 with 12 df				
6. Devices like films, tapes, programmed booklets, and textbooks remove an essential element from education-humanity.				
(a) STRONGLY AGREE	4	9	10	0
(b) AGREE	28	28	23	20
(c) UNDECIDED	36	26	47	29
(d) DISAGREE	28	30	18	40
(e) STRONGLY DISAGREE	4	7	2	11
Number of Omits = 1				
χ^2 = 16.52 with 12 df				

Table 3.4.3 (continued)

Percentage of Responses in Each Instructional Treatment Group to Questionnaire Statements for Which Inter-Group Response Patterns Did Not Differ Significantly

Statement	Instructional Treatment Group			
	1[a] (N=25)	2[b] (N=44)	3[c] (N=49)	4[d] (N=45)
7. Devices like films, tapes, programmed booklets, and textbooks make school more interesting as they provide alternatives to the traditional teacher-led class.				
(a) STRONGLY AGREE	8	12	12	20
(b) AGREE	60	60	51	67
(c) UNDECIDED	16	21	23	11
(d) DISAGREE	8	2	14	2
(e) STRONGLY DISAGREE	8	5	0	0
Number of Omits = 2				
χ^2 = 18.04 with 12 df				
8. Instruction is much more interesting when various types of materials like films, tapes, books, and teachers are used in combination.				
(a) STRONGLY AGREE	20	32	31	31
(b) AGREE	52	48	41	44
(c) UNDECIDED	12	16	10	18
(d) DISAGREE	12	4	18	7
(e) STRONGLY DISAGREE	4	0	0	0
Number of Omits = 0				
χ^2 = 13.18 with 12 df				

Table 3.4.3 (continued)

Percentage of Responses in Each Instructional Treatment Group to Questionnaire Statements for Which Inter-Group Response Patterns Did Not Differ Significantly

Statement	Instructional Treatment Group 1[a] (N=25)	2[b] (N=44)	3[c] (N=49)	4[d] (N=45)
9. It is easier and less confusing to have only one type of instructional mode. That is, I would sooner be in a reading mode or a programmed instruction mode than in some combination of these modes.				
(a) STRONGLY AGREE	0	7	4	4
(b) AGREE	12	29	38	24.5
(c) UNDECIDED	36	14	17	24.5
(d) DISAGREE	40	43	29	38
(e) STRONGLY DISAGREE	12	7	12	9
Number of Omits = 1				
χ^2 = 12.55 with 12 df				
10. If alternative types of instruction exist, students like to choose by themselves the type of instruction that they will use.				
(a) STRONGLY AGREE	20	18	31	24
(b) AGREE	64	64	57	56
(c) UNDECIDED	12	16	10	16
(d) DISAGREE	4	2	2	4
(e) STRONGLY DISAGREE	0	0	0	0
Number of Omits = 0				
χ^2 = 3.44 with 12 df				

Table 3.4.3 (continued)

Percentage of Responses in Each Instructional Treatment Group to Questionnaire Statements for Which Inter-Group Response Patterns Did Not Differ Significantly

Statement	Instructional Treatment Group 1[a] (N=25)	2[b] (N=44)	3[c] (N=49)	4[d] (N=45)
11. The alternative types of instruction exist, students like teachers to tell them which type of instruction to use.				
(a) STRONGLY AGREE	0	0	2	0
(b) AGREE	16	18	12	11
(c) UNDECIDED	8	11	20	26
(d) DISAGREE	64	57	37	45
(e) STRONGLY DISAGREE	12	14	29	18
Number of Omits = 0				
χ^2 = 14.99 with 12 df				
12. If alternative types of instruction exist, students like to consult with teachers to determine which type of instruction they should use.				
(a) STRONGLY AGREE	4	16	6	9
(b) AGREE	68	54	64	56
(c) UNDECIDED	20	23	22	26
(d) DISAGREE	8	7	6	9
(e) STRONGLY DISAGREE	0	0	2	0
Number of Omits = 2				
χ^2 = 6.91 with 12 df				

a. Treatment consisting of teacher lectures.
b. Treatment consisting primarily of audiotapes and worksheets.
c. Treatment consisting of a reading handout.
d. Treatment consisting of programmed instruction booklets.

From the mean responses to statements 1, 2, and 3 it can be seen that the students in all of the treatment modes generally felt the materials to be of the appropriate level of difficulty. They tended to disagree with statements inferring that the instructional materials in "The Structure of Matter" module were too easy or too hard.

A chi-square analysis, however, revealed that there were significant differences in response frequencies among the treatment modes to statements 1 and 3. Using this information coupled with an analysis of the individual means, it would appear that the students in treatment 1 (teacher mode) were the least sure that the instructional materials were of the correct difficulty level. Of all of the treatment groups they tended to disagree most strongly that the instructional materials were too hard, and they tended to be the most undecided that the materials were too easy. Thus there seems to be a slight tendency for the students in treatment 1 (teacher mode) to regard the materials in this treatment as being too easy.

A chi-square analysis of the responses to statement 4 indicated that the treatment groups differed significantly. An examination of the mean responses shows that treatments 1 and 4 (teaching and programmed instruction) were liked better than treatments 2 and 3 (media and reading). The students in treatments 1 and 4 tended to agree that the instructional materials in "The Structure of Matter" module were liked about as well as instructional materials can be liked. The students in treatments 2 and 3 were undecided about this statement.

While the treatment groups did not differ significantly in their response frequencies on the next eight statements, the mean response values seem to suggest the following:

1. Students were undecided or tended to agree slightly that they learn better from "live" teachers than from films, tapes, programmed instruction, or textbooks.

2. Students were undecided as to whether devices like films, tapes, programmed booklets, and textbooks removed an essential element from education-humanity.

3. Students tended to agree that devices like films, tapes, programmed booklets, and textbooks made school more interesting since they provided alternatives to the traditional teacher-led class.

4. Students agreed that it is much more interesting when various types of materials like films, tapes, books, and teachers are used in combination.

5. Students were undecided or tended to slightly disagree with the statement that it is easier and less confusing to have only one type of instructional mode.

6. Students agreed that if alternative types of instruction existed, they would like to choose by themselves the type of instruction they will use.

7. Students disagreed with situations where teachers tell them the type of instructional alternative to use.

8. Students tended to agree that it was desirable to have teachers consult with them when selecting the type of instructional alternative they will use.

In summary, the ninth grade science students seemed to feel that a diversity of instructional treatments was better than only one type of instruction. These alternative treatments are best utilized in combination. When faced with choosing among alternative instructional treatments, students prefer to choose by themselves the type of treatment they will use. As a compromise, they seem to be willing to consult with teachers in choosing their treatment. By no means do they want to be told what treatment to use.

The final two questions on the Student Questionnaire asked the students to comment on the "Structure of Matter" module and suggest ways in which the module could be improved. Individual student needs in instructional treatments were revealed by their responses to these questions. Selected comments pertinent to the specific treatments are stated below.

Treatment 1 - This was the treatment that consisted of a series teacher lectures. For this module the three Jamesville-DeWitt science teachers altered their usual procedure of teaching one to a class by having all three teachers lecture on various segments of the module to all lecture sections. Some interesting comments were the following.

"I think people should be allowed to pick which treatment group they want to be in." This same student also said, "I liked this module better than the others, because I can't stand reading or listening to tapes, and I had lectures, and I learn much more that way."

"For the teacher instruction module I think one teacher should teach it instead of a different teacher everyday." This same student also said, "I liked this module very much. It's the first module that I completely understood everything before taking the test."

"This group I was in was very interesting. I certainly learned from it, but I think it could be made more enjoyable by more live experiments and models."

"It would be fun to do more experiments." This student also said, "I like the different teachers instead of the same one every day. It makes for unbiased feelings."

Treatment 2 - This is the treatment that consisted primarily of the audio-tapes. A selection of student comments are stated below.

"I feel that in the media section there wasn't enough variety in the materials used. More variety would have been better." This same student also said, "For about the first time this year I feel that I've really learned something that will stick. That's because it was very organized and everything feel into place and made sense."

"I did the media and in some tapes they don't explain very well and I needed a seminar but couldn't get it. The teacher explained a few things to me, but I think a seminar should be included in all of the modes."

One student just said, "Select your own mode."

"Just don't stick to one mode. It becomes very boring. We should be able to consult and work together with other students." This same student also said, "I feel that if you ruin the first LAP Mastery Test, instead of having to go back to the same material to learn what you didn't get the first time, you should have other variations in the way to learn the second time. Not the same tape or whatever. Usually if you can't get it the first time, its not going to do much better the second time -- so you need another way."

Several students pointed out that the technical quality of this treatment was not equal to that of the other treatments. The quality of the reproduced audiotapes was poor due to an initial recording volume that was too high. This led to some interpretational problems.

Treatment 3 - This treatment was primarily composed of a reading handout. Selected student comments appeared as below.

"I didn't like the reading part. I just had to memorize the facts and then I forgot them. I needed something (like a worksheet) to let me use what I learned and then I wouldn't have forgotten them so easily." This same student also said, "I wish I could have picked my way of instruction. I hate to read and I got stuck in the reading mode. When I start to read my mind wanders off, then I have to go back again and read it all over. I know I would have done better if I got to pick my own group."

"I think the module on the "Structure of Matter" was the best and most organized module all year. It was interesting, easy enough to understand, and I learned a lot from it."

"I think you should use a combination of all these. With the reading I found I had to go to the teacher when I couldn't understand something." This same student also said, "The group I was in just depended on how good a reader you are."

"While you couldn't see the pictures in the reading handout (The author knew that several nonessential pictures were not reproducable. All essential pictures were made reproducable.), I like the direction and guidance you get with this. I could also work at my own rate and refer back to the handout."

Treatment 4 - This treatment consisted of five programmed instruction booklets. Selected comments are as follows.

"This was the first module where I only missed 4 questions on the module test. All the others I missed 7 and 8. So I think the Programmed Instruction Books are the best learning I've had all year!"

"I learned more in this module than any of the others. I'm not sure whether this was because the material was easy or because of the way it was taught."

"In the booklets you go over too much too fast. You don't explain things very clearly. I had to go to the teacher to have him explain all of booklet 4 to me. Yet even with all the problems, I've learned more in this module than in practically any other."

"I think the programmed books were the best of everything. I learned more from them than from anything else in science this year."

"It was a worthwhile unit. I could go at my own speed." "Programmed instruction was good and I learned more from it than from anything else all year."

While these are only selected responses from the group of students who actually replied to the last two questions on the questionnaire they are representative. Several tentative conclusions can be extracted from these responses. First of all, one can easily tell that the most popular treatments were the treatment consisting of the programmed booklets and the treatment made up of the teacher lectures. This qualitative difference may have been due to the technical quality of the treatments.

More importantly, as indicated by student responses, no one treatment is preferred by all of the students. As one student stated, "I hate to read" and "my mind wanders" when I read, yet "I got stuck in the reading mode." Another student liked the treatment which consisted of the series of teacher lectures because "I can't stand reading or listening to tapes, and I had lectures, and I learn much more that way."

However there is at least some evidence to suggest that students should <u>not</u> have the opportunity to select their own mode of instruction, at least if maximization of student learning is the criterion. Cronbach and Snow (1969) have shown that students are poor judges of the treatments that best suit them. What are needed are more action research studies (such as this one) that match learner characteristics to instructional treatments; <u>instructional treatments which the particular school system wants to implement</u>. In lieu of these studies, a diversity of available treatments appears superior to a single treatment. By having some say in choosing their mode of instruction, students are apt to be more satisfied than if they were forced into a specific treatment. One would expect that satisfied students would perform better than dissatisfied ones.

A final theme seems to be prevalent in the student responses. Selected students in all of the treatment groups said that they learned more in this particular module than they did in any previous module. They attributed this to the organized presentation of the specific treatments and implied that other modules weren't as well organized.

Chapter IV

Conclusions and Implications for Further Research

4.1 ATI Study

Limitations

There were several obvious limitations to this section of the research. One limitation was the lack of control over the execution of the instructional treatments. As was pointed out previously, treatment 2 was not completely effective because of technical problems with the audiotapes. From responses to the Student Questionnaire, it appeared that the treatments were not separated as completely as would have been desirable experimentally. While this extremely flexible nature is the actual method of operation for the Jamesville-DeWitt science program, it severely limits any inferences that can be made from the data.

Another limitation is the time lag between several of the aptitude measures and the implementation of the treatments. Specifically, the various SRA scores were collected two years before the actual study. While the reliability of these test scores is high, such a time lag reduces their usefulness.

General Comments

The discussion presented in section 3.1 provides some evidence on how to assign students to the treatment groups on the basis of Letter Sets and Test Anxiety Scale for Children scores. Yet many perplexing questions remain. Specifically why did only one anxiety measure produce a significant interaction and why didn't the interactions hold up across the two posttesting situations? What are the reasons for the tenuousness of ATI's. Robert Glaser (1972) feels that the lack of empirical evidence to substantiate interactions is due to faulty aptitude measures. He states that present aptitude tests are not producing dimensions for measuring those individual differences that interact with different ways of learning. The aptitude measures do not appear to relate to the processes of learning that have been under investigation. Glaser also feels that the treatments investigated in the ATI studies were not generated by any systematic analysis of the kinds of psychological processes called upon in particular instructional methods, and individual differences were not assessed in terms of these processes.

Glaser discusses what new aptitudes should be measured. He feels that a fruitful approach is the conceptualization of individual difference variables in terms of the process constructs of contemporary theories of learning, development, and human performance. As an example he refers to the work of William Rohwer. Rohwer has been concerned with the process of "mental elaboration". This process refers to the fact that individuals recode or transform materials presented to them by elaborating the content. As an example, it is known that it is easier to remember the words "boy" and "horse" when the learner provides himself with or is provided with some visual or verbal relationship between the words. This relationship could be a picture of a boy on a horse or it could be a sentence connecting the two words (Glaser, 1972). Glaser points out that Rohwer's work suggests that individual differences related to children's backgrounds influence the way in which they carry out cognitive processes such as "mental elaboration". Since this process facilitates learning, Rohwer feels that it would be fruitful to train particular children in various elaborative techniques of learning. Although not specifically stated, Rohwer must feel that certain techniques of learning elaboration would be most suitable for children with certain backgrounds.

What does all of this mean to an ATI study such as that conducted during this research? It means that commercially available aptitude tests will be practically useless. Specific aptitude measures will have to be tailored to the particular interaction under investigation. Specific treatments will have to be analyzed and developed with great precision. They will have to differ by measured amounts on the variables of interest. The treatments will have to be carried out precisely as prescribed by the investigators.

Where does this take ATI research? In this investigator's opinion it takes it out of the classroom and into the psychological laboratory. Can anyone think of a school system where they will allow treatments differing only in the degree of "mental elaboration" to be implemented? (Can anyone imagine the researcher who has spent very considerable time to develop such treatments, allowing the treatments to be tested in schools where cooperation and implementation is not always of the highest degree?) Will schools allow aptitude measures that assess the background of a student in the area of "mental elaboration"? It is felt that at this time neither the school nor the researcher would benefit from such studies being conducted in a school setting. From the school's standpoint present curricula and schedules would have to be interrupted. From the researcher's standpoint the appropriate control would be hard to achieve.

It was refreshing to read Glaser's article after conducting this ATI study. It will be remembered that treatment names were avoided and formal ATI hypotheses were not stated. Reasons given were that these processes led to the reading of qualities into our treatments that they were not known to possess. This was an unsophisticated way of saying that our treatments were not generated by any systematic analysis of the learning processes involved and our aptitude measures were not selected to measure differences in the learning processes.

If another ATI study is ever conducted there is a basic change that would be made. Four global treatments wouldn't be chosen in a search for ATI's in a module of science instruction. Rather the focus would be on a particular cognitive process and the treatments and aptitude measures would be designed around that process. In implementing the study close control would be manditory. The treatments would have to be carried out exactly as specified. The change would be from a macroscopic to a microscopic view of the learning process.

4.2 Evaluative Study

Latent Partition Analysis

The results of the latent partition analyses reveal several interesting findings. Two concepts were placed in different latent categories after instruction. One ambiguity was removed by instruction and three others were created. The latent category composition of the high achievers differed from that of the low achievers. A greater variety of ambiguities was present in the sorting data of the low achievers. The high achieving students focused their ambiguities on the "electrical" concepts. The low achieving students had problems with the "electrical" concepts, but to a lesser extent. In addition, they experienced ambiguities with two additional "non-electrical" concepts. Finally, differences between pre- and post-instruction categorizations within each treatment group were found.

This data would appear to be useful to the classroom teacher. The concepts that shifted categories after instruction and the ambiguities removed by instruction may point out those areas where instruction was most intense. Ambiguities created by instruction may point out areas that need further clarification. The teacher could examine the latent category composition before and after instruction and decide on the particular concept interrelationships that should be stressed in future assignments.

The differences between the high and low achieving students may be especially illuminating. The different ambiguities between the two groups may point out areas that require additional clarification for the low achievers.

Several limitations of this phase of the study should be pointed out. The packet of concepts in the Student Sorting Task consisted of a fixed order of the concepts in both the pre- and post-instruction administrations. To avoid any groupings of the concepts solely on the basis of the order in which they appeared, the order of the concepts should probably have been arranged randomly within each packet.

A final limitation concerns itself with the particular subject matter of the instructional module. The students had obviously been introduced to many of the concepts in earlier science courses. Perhaps more dramatic changes would have been evident if the students had no prior exposure to the concepts. It now appears that a more suitable area of instructional content could have been chosen for our investigation.

Individual Differences Model for Multidimensional Scaling

The individual differences model for multidimensional scaling revealed that instruction had the effect of producing an additional interpretable point of view. The effect of instruction was also to differentially effect sensitivity to differences existing among the concepts. Certain concept pairs were judged to be more dissimilar after instruction, while others were judged to be more similar. (Used in conjunction with the LPA data, these results may enable the teachers to change the emphasis of their programs.) Instruction appears to have provided the high achievers with more of a basis for organizing the concepts than it did for the low achievers. Both the four and three dimensional points of view for the high achievers were interpretable, but the five dimensional point of view wasn't (due to a poor fit).

There is both a major and a minor limitation to this phase of the study. The minor limitation has been discussed under the previous section. If a content area had been chosen about which less had been known, greater instructional differences would have been expected.

The major limitation concerns the nature of the Similarity Judgment Questionnaire. Using 13 concepts, 78 concept pairs had to be included in the questionnaire. In addition, to gather reliability data another 13 pairs were added to the original 78. This resulted in an exceedingly long and boring instrument. To alleviate this tedium the number of concepts should have been reduced from the 13 used in this study.

General Comments

It was originally hoped that a substantial number of science experts would take the Student Sorting Task and the Similarity Judgment Questionnaire. From their responses a criterion latent categorization and a criterion concept space could have been formulated. However, for the present, this phase of the research had to be discontinued. Only 12 expert responses were received. Owing to its importance, it was felt best to hold off until a substantial criterion population could be accumulated.

The importance of this aspect of the study became more apparent after analyzing the data. While group differences were indicated by both the latent partition and points of view analyses, it was difficult to evaluate the differences without reference to some type of criterion. This problem was magnified by the number of analyses conducted. Without a common reference point the many analyses tended to become confusing. Thus the next step is clearly delineated. Only with its completion will we really be able to judge the effectiveness of these two psychometric models to evaluate instruction.

References

Baker, F.B. Computer-based instructional management systems: a first look. Review of Educational Research, 1971, 41, 51-70.

Berliner, D.C., & Cahen, L.S. Trait-treatment interactions and learning. In F.N. Kerlinger (Ed.), Review of Research in Education. New York: F.E. Peacock Co., in press.

Bloom, B.S. et al. Taxonomy of educational objectives. Handbook 1: Cognitive domain, New York: David McKay Co., Inc., 1956.

Bracht, G.H. Experimental factors related to aptitude-treatment interactions. Review of Educational Research, 1970, 40, 627-646.

Briggs, L.J. Learner variables and educational media. Review of Educational Research, 1968, 38, 160-175.

Brown, W.F., & Holtzman, W.H. Survey of study habits and attitudes manual. New York: The Psychological Corporation, 1967.

Bunderson, C.V. Ability by treatment interactions in designing instruction for a hierarchical learning task. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, February 1969.

Burton, B.B., & Golbeck, R.A. The effect of response characteristics in multiple-choice alternatives on learning during programmed instruction. San Mateo, Ca.: American Institute for Research, February 1962.

Carroll, J.B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Castenada, A., McCandless, B.R., & Palermo, D.S. The children's form of the manifest anxiety scale. Child Development, 1956, 27, 317-326.

Cliff, N. The "idealized individual" interpretation of individual differences in multidimensional scaling. Psychometrika, 1968, 33, 225-232.

Crandall, V.C., Katkovsky, W., & Crandall, V.J. Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. Child Development, 1964, 35, 91-109.

Cronbach, L.J. How can instruction be adapted to individual differences? In R.M. Gagné (Ed.), Learning and individual differences. Columbus, Ohio: Charles E. Merrill Books, 1967.

Cronbach, L.J., & Snow, R.E. Individual differences in learning ability as a function of instructional variables. Final Report, March 1969, Stanford University, Contract OEC 4-6-061269-1217, U.S. Office of Education.

Deken, J.G., & Johnson, T.J. Treatment of similarity data by the method of item definition analysis. St. Ann, Mo.: CEMREL, Inc., 1971.

Eckart, C., & Young, G. The approximation of one matrix by another of lower rank. Psychometrika, 1936, 1, 211-218.

Evans, G.T. The analysis of categorizing behavior. Psychometrika, 1970, 35, 367-392. (a)

Evans, G.T. Application of a model of categorizing behavior. Multivariate Behavioral Research, 1970, 5, 489-505. (b)

Flanagan, J.C. Functional education for the seventies. Phi Delta Kappan, 1967, 49, 27-32.

French, J., Ekstrom, R., & Price, L. Manual for kit of reference tests for cognitive factors. Princeton, New Jersey: Educational Testing Service, 1963.

Frymier, J.R. Development and validation of a motivation index. Theory into Practice, 1970, 9, 56-85.

Glaser, R. Adapting the elementary school curriculum to individual performance. In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1968.

Glaser, R. Individuals and learning: the new aptitudes. Educational Researcher, 1972, 1, 5-13.

Glaser, R., & Nitko, A.J. Measurement in learning and instruction. In R.L. Thorndike (Ed.), Educational measurement. Washington: American Council on Education, 1971, 625-670.

Green, B.F., Jr. The computer revolution in psychometrics. Psychometrika, 1966, 31, 437-445.

Gulliksen, H., & Wilks, S.S. Regression tests for several samples. Psychometrika, 1950, 15, 91-114.

Hall, A.E. Concept attainment: A psychometric approach. Dissertation Abstracts, 1964, 25, 644.

Hambleton, R.K. A report on the research and evaluation activities in the Jamesville-DeWitt individualized instruction program in ninth grade science. Final Report. Albany, New York: Bureau of School and Cultural Research, New York State Education Department, 1971.

Hambleton, R.K. A review of testing and decision-making procedures for selected individualized instructional programs. ACT Technical Bulletin. Iowa City, Iowa: American College Testing Program, 1973.

Hambleton, R.K., & Gorth, W.P. Criterion-referenced testing: Issues and applications. Center for Educational Research Technical Report No. 13. Amherst, Mass.: School of Education, University of Massachusetts, 1971.

Harasyn, P., & Precht, D. A FORTRAN IV program for latent partition analysis. Edmonton, Alberta: Division of Educational Research Services Computer Program Documentation, University of Alberta, 1971.

Hecht, A.R. A semantic differential evaluation of attitudinal outcomes of introductory physical science. A paper presented at the annual meeting of the American Educational Research Association, New York, 1971.

Hess, R.J., & Johnson, T.J. Changes in cognitive structure after a semester of college instruction. Paper presented at meeting of American Education Research Association, New York, 1971.

Hieronymus, A.N., & Stroud, J.B. Comparability of IQ scores on five widely used intelligence tests. Measurement and Evaluation in Guidance, 1969, 3, 135-140.

Johnson, T.J., Ashton, P., Deken, J., Kerr, N., & Robb, R. Cognitive and associative structure in economics. Paper presented at the annual meeting of the American Education Research Association, Minneapolis, 1970.

Jones, W. SRA Achievement Series. In O.K. Buros (Ed.), The Fifth Mental Measurements Yearbook. The Gryphon Press, Highland Park, New Jersey, 1959, 48-55.

Kerlinger, F.N. Foundations of behavioral research. New York: Holt, Rinehart and Winston, Inc., 1964.

Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 28, 1-27. (a)

Kruskal, J.B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 115-129. (b)

Milholland, J. Lorge-Thorndike Intelligence Tests. In O.K. Buros (Ed.), The Fifth Mental Measurements Yearbook. The Gryphon Press, Highland Park, New Jersey, 1959, 478-484.

Miller, D., Baker, J.F., Clasen, R.E., Conry, J.L., Conry, R.F., Pratt, A.D., Sheets, S.E., Wiley, D.E., & Wolfe, R.G. Elementary school teachers' viewpoints of teaching and learning. Final report, 1968, University of Wisconsin, Project 5-1015-2-12-1, U.S. Office of Education.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. The Measurement of meaning. Urbana: University of Illinois Press, 1957.

Pennell, R. Finding points of view in judgment data. Research Bulletin 71-8. Princeton, N.J.: Educational Testing Service, 1971.

Philips, B.N. Interpretive study 11: Anxiety and school related interventions. Research Report BSOR003-71. Albany, New York: New York State Education Department, 1971.

Potthoff, R.F. On the Johnson-Neyman technique and some extensions thereof. Psychometrika, 1964, 29, 241-256.

Root, R.T. A multidimensional scaling analysis of meaning. Dissertation Abstracts, 1963, 23, 2979.

Ross, J. A remark on Tucker and Messick's "points of view" analysis. Psychometrika, 1966, 31, 27-31.

Ross, R.T. Optimum orders for the presentation of pairs in the method of paired comparisons. Journal of Educational Psychology, 1934, 25, 375-382.

Sarason, I.G. Empirical findings and theoretical problems in the use of anxiety scales. Psychological Bulletin, 1960, 5, 403-415.

Sarason, S.B., & Mandler, G. Some correlates of test anxiety. Journal of Abnormal and Social Psychology, 1952, 4, 810-817.

Shepard, R.N. The analysis of proximities: Multidimensional scaling with an unknown distance function: 1 & 11. Psychometrika, 1962, 27, 125-140.

Shepard, R.N. A taxonomy of some multidimensional scaling methods. Paper presented at the Bell Laboratories' Multidimensional Scaling Workshop, Philadelphia, Pa., 1972.

Tallmadge, G.K., & Shearer, J.W. Interactive relationships among learner characteristics, types of learning, instructional methods, and subject matter variables. Experimental Publication System, 1970, 8, Ms. No. 291-346.

Thorpe, L.P., LeFever, D.W., & Naslund, R.A. SRA achievement series. Chicago: Science Research Associates, 1963.

Tittle, C.K. Lorge-Thorndike Intelligence Tests: Multilevel Edition. In O.K. Buros (Ed.), The Seventh Mental Measurements Yearbook. The Gryphon Press, Highland Park, New Jersey, 1972, 684-686.

Torgerson, W.S. Multidimensional scaling: 1. Theory and method. Psychometrika, 1952, 17, 401-419.

Torgerson, W.S. Theory and methods of scaling. New York: Wiley, 1958.

Torgerson, W.S. Multidimensional scaling of similarity. Psychometrika, 1965, 30, 379-393.

Traub, R.E., & Hambleton, R.K. The effect of instruction on the semantic space defined by statistical and psychometric concepts. Paper presented at the annual meeting of the American Educational Research Association, New York, 1971.

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

Wagenaar, W.A., & Padmos, P. Quantitative interpretation of stresses in Kruskal's multidimensional scaling technique. British Journal of Mathematical and Statistical Psychology, 1971, 24, 101-110.

Wiggins, N. Individual viewpoints of social desirability. Psychological Bulletin, 1966, 66, 68-73.

Wiggins, N. Individual differences in diagnostic judgments of psychosis and neurosis from the MMPI. Educational and Psychological Measurement, 1971, 31, 199-214.

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

Young, F.W. TORSCA-9, A FORTRAN IV program for nonmetric multidimensional scaling. Behavioral Science, 1968, 13, 343-344.

Young, F.W. Nonmetric multidimensional scaling: Recovery of metric information. Psychometrika, 1970, 35, 455-473.

Appendix A

Objective List and Test Descriptions
(All instruments except for the SSHA, SRA, and IQ tests are reproduced in supplemental report no. 1)

The Objectives for tne Module "Structure of Matter"

1.1 identify a substance as a compound by its characteristics.

1.2 identify a substance as an element by its characteristics.

1.3 identify a substance as a mixture by its characteristics.

2.1 determine certain characteristics of an atom (atomic number, atomic mass or weight, number of protons, electrons or neutrons) using the Periodic Table of the Elements.

2.2 select the correct Bohr atom diagram given the atomic number and atomic weight of an element.

2.3 given the Bohr model of an atom and the Periodic Table, identify the atom.

3.1 using the Periodic Table, classify a given element as a metal, a non-metal, or an amphoteric element.

3.2 using the Periodic Table, determine the most common oxidation or valence number of an element.

3.3 determine the change in the electron configuration number(s) when an atom becomes an ion.

3.4 determine the number of atoms in a molecule, given the chemical formula for the molecule.

3.5 select the correct chemical formula for a compound, given a Table of Radicals and a Periodic Table.

Module Pretest

The Module Pretest was developed to provide a preliminary assessment of the objectives of the "Structure of Matter" module. The test consists of 40 multiple-choice questions with four options per question. The first three objectives of the module were covered by 10 items. The fourth was covered by 11 items. The next two objectives were covered by two items per objective. The last five objectives were covered by three items each.

Module Posttest

The Module Posttest was developed to provide a final assessment of the objectives of the "Structure of Matter" module. The test consists of 40 multiple-choice questions with four options per question. As with the Module Pretest, the first three objectives of the module were covered by 10 items. The fourth objective was covered by 11 items. The next two objectives were covered by two items each. The last five objectives were covered by three items.

Junior Index of Motivation (JIM Scale)

The JIM Scale is an instrument designed to assess junior high school students' desire to learn in school or, in other words, to measure achievement motivation. It was predicated upon the assumption that whatever causes one to try to do good work in school comes primarily from within rather than from without, and that whatever this motivation or force is, it is probably rooted in one's personality structure, his value structure, and his curiosity (Frymier, 1970). The author points out that the individual items of the JIM Scale probe the students' value structure, their self-concept, and their openness to experience in such a way that a measure of motivation can be obtained. Specifically the JIM Scale consists of 80 statements. The student has to choose one of the following alternatives for each: +2 indicating slight support, agreement; +1 indicating strong support, agreement; +3 indicating slight opposition, disagreement; +4 indicating strong opposition, disagreement.

Frymier cites several studies that were used to validate the JIM Scale. In one study, students who were seen by their teachers as being highly motivated made significantly higher JIM Scale scores than students who were seen by their teachers as being low in motivation.

By using discrepancies in IQ and grade from an average IQ and grade, underachievers and overachievers were isolated in the population of interest. The author then showed that students who are overachievers will make significantly higher total JIM Scale scores than will students who are underachievers.

Frymier provided additional validity evidence by correlating JIM Scale scores with the scores from another measure of motivation (Farquhar's M-Scale). For grade 9 students the correlation was .44, while for grade 11 students it was .57. The JIM Scale also gave correlations of .44 for grade 9 and .50 for grade 11 students when correlated with ITED (Iowa Tests of Educational Development) composite scores.

In summary, the JIM Scale appears to be able to differentiate among students differing in level of motivation. It correlates to a considerable degree with another measure of academic motivation and with standardized achievement scores. Frymier feels that validity of the JIM Scale seems to have been substantiated to the point that it can be recommended as a research tool for use with groups.

Split-half reliabilities of .83 and test-retest reliabilities of .70 seem to indicate that the JIM Scale is internally consistent and dependable over time. Because of the long time interval between test administrations (10 months), and accepting the evidence for the validity of the JIM Scale, it appears that motivation itself is a fairly constant phenomemon over an extended period of time. Frymier states that the test-retest reliability coefficient of .70 implies that the motivation to learn in school persists from one year to the next.

In view of the importance of achievement motivation to school success we considered this to be a useful variable to assess in the study.

The Intellectual Achievement Responsibility Scale

Crandall, Katkovsky, and Crandall (1964) provide the rationale for the Intellectual Achievement Responsibility Scale. They state that individuals have been found to differ in the degree to which they believe that they are usually able to influence the outcome of situations. They may believe that their actions produce the reinforcements which follow their efforts, or they may feel that the rewards and punishments meted out to them are at the discretion of powerful others or are in the hands of luck or fate. In fact, the same reinforcement in the same situation may be perceived by one individual as within his own control and by another as outside his own influence. Crandall, Katkovsky, and Crandall think that these personal beliefs could be important determiners of the reinforcing effects of many experiences. If, for example, the individual is convinced that he has little control over the rewards and punishments he receives, then he has little reason to modify his behavior in an attempt to alter the probability that those events will occur. Rewards and punishments, then, will have lost much of their reinforcing value, since they will not be as effective in strengthening or weakening the individual's response.

The Intellectual Achievement Responsibility Scale (IAR) attempts to measure beliefs in internal versus external reinforcement responsibility. It is aimed at assessing children's beliefs in intellectual-academic achievement situations and limits the source of external control to those persons who most often come in face-to-face contact with a child, his parents, teachers, and peers (Crandall, Katkovsky, and Crandall, 1964). Specifically, the IAR scale is composed of 34 forced-choice items. Each item stem describes either a positive or a negative achievement experience which routinely occurs in children's daily lives. This stem is followed by one alternative stating that the event occurred because of the behavior of someone else in the child's immediate environment. A child's I+ score (indicating belief in internal responsibility for successes) is obtained by summing all positive events for which he assumes credit. A child's I- score (indicating belief in internal responsibility for failures) is obtained by summing all negative events for which he assumes credit. His total I score is the sum of his I+ and his I- subscores.

The developmental sample consisted of 923 elementary and high school students and was drawn from five different schools so that it would be representative of children in diverse kinds of communities. Included were students from a consolidated country school, a village school, a small-city school, a medium-city school, and a college laboratory school.

Test-retest reliabilities for the younger children were .69 for total I, .66 for I+, and .74 for I-. For the ninth-grade students test-retest reliabilities were .65 for total I, .47 for I+, and .69 for I-. As measures of internal consistency split-half reliabilities were computed for the separate subscales. For a random sample of 130 of the younger children, the correlations were .54 for I+ and .57 for I-. For a similar sample of older children, the correlations were .60 for both the I+ and I- subscales.

The authors cite several other statistics to lend some additional support to the construct validity of children's beliefs in their control of reinforcements. Among these are the low correlations between the I+ and I- subscales. As for sex and age differences, I+, I-, and total I scores tend to increase only slightly with age and girls' scores tend to be somewhat higher than boys', especially from grade 6 upward. First-born children in the upper grades tend to give higher total I scores. Children's Social Desirability Scale scores correlate only slightly with IAR scores. IAR scores predict various achievement measures, especially course grades. The authors discuss all of these findings in the context of the theory developed around internal and external belief systems.

Anxiety Scales

Three different anxiety scales were included in the study: one to measure general anxiety, another to measure test anxiety, and the third to measure school anxiety. Of interest to us was whether there was an interaction between instructional mode and anxiety variables as has been demonstrated in several other studies.

Children's Manifest Anxiety Scale - The Children's Manifest Anxiety Scale (CMAS) was developed by Castenda, McCandless, and Palermo (1956). It is an adaptation of the Taylor Manifest Anxiety Scale appropriate for use with elementary school children. The scale consists of 42 anxiety items and 11 items which provide an index of the subject's tendency to falsify his responses. Phillips (1971) states that the anxiety items can be grouped into roughly the same five categories as those in the Manifest Anxiety Scale. Those categories are (1) physiological disorders, (2) general emotionality, (3) the direct admission of worry or nervousness, (4) physiological stress, and (5) self-consiousness and self-confidence. Phillips regards the CMAS as a measure of a generalized state of anxiety. It is one of the most popular measures of general anxiety in children.

One-week test-retest reliabilities were about .90 for the anxiety scale and about .70 for the lie scale. Intercorrelations between the anxiety scale and the lie scale clustered around the zero value. (This would be the desired correlation if one assumes that the tendency to falsify responses to the anxiety items could result in a high anxiety score as well as in a low one.) Sex differences were found on both scales. Girls scored significantly higher than boys. Lie scale scores were found to be associated with grade level. Sixth grade students had significantly lower lie scale scores than did fourth or fifth grade students.

Test Anxiety Scale for Children - When one considers the pervasive use of tests in our culture and the ways in which they determine the lives of the people who take them, it would be expected that the testing situation would engender an anxious reaction in more than a few individuals. To get at these anxious reactions Sarason et al.(1960) developed the Test Anxiety Scale for Children (TASC). The TASC is composed of 30 yes-no questions dealing with the child's feelings about his class performance, tests, and about how he compares with other members of his class. It is specifically intended for use with children in the elementary grades.

The developmental population consisted of 600 children in grades 2 through 5 in six elementary schools. Split-half reliability coefficients (corrected by the Spearman-Brown formula) ranged from .820 to .899 for grades 2 through 5. Test-retest coefficients ranged from .44 to .82 for grades 2 through 5. Sarason points out that a test-retest coefficient of .71 over all grades seems to indicate that the test-retest coefficient of .44 may be spuriously low.

As evidence of the predictive validity of the TASC, Sarason has correlated teachers' ratings of anxious behavior with TASC scores. Coefficients ranged from .21 to .31 for the four grades with a coefficient of .27 over all four grades. Sarason feels that these low validity coefficients may be more of a reflection of the invalidity of teachers' ratings than of any inadequacy in the TASC.

Various correlates with TASC scores seem to support the construct validity of the instrument. Sarason reports that TASC scores increased significantly with grade. Correlations between TASC scores and IQ were -.19, -.21, -.27, and -.28 respectively for grades 2 through 5. These small but significant correlations between TASC scores and IQ were anticipated on the basis of previous studies with college students (Sarason and Mandler, 1952). Anxiety theory would also point to this negative relationship since the test anxious reaction is thought to be primarily interfering in its effect. The correlation between _TASC_ scores and occupational level of the parents was .119 which is significant at the .001 level. A chi square analysis revealed that students who had either a reading, an arithmetic, or a behavior problem (diagnosis made by the teacher) obtained significantly higher _TASC_ scores than no-problem cases. Finally, Sarason _et al_. (1960) has reported several studies which indicate that pupils with high _TASC_ scores performed more poorly on test-like tasks than did children with low _TASC_ scores; these results did not hold for game-like tasks.

In factor analytic studies four factors were identified which held up across age and sex: (a) test anxiety; (b) somatic signs of anxiety; (c) negative self-evaluation; and (d) remote school concerns (Phillips, 1971).

All of these results seem to indicate that the _TASC_ reliability measures various aspects of anxious behavior in the school setting.

School Anxiety Scale - _The School Anxiety Scale_, developed by Phillips (1966), makes use of items from the _Test Anxiety Scale for Children_, the _Achievement Anxiety Scale_, the _Audience Anxiety Scale_, and other personality instruments (Phillips, 1971). It was designed to assess anxiety associated with a broader range of stressful school situations than is encompassed by the Sarason scales (such as the _Test Anxiety Scale for Children_). A factor analysis revealed four factors which roughly parallel those found for the _Test Anxiety Scale for Children_: (a) fear of taking tests; (b) physiological reactivity associated with a low tolerance for stres; (c) lack of confidence in meeting the expectations of others, particularly teachers; and (d) fear of negative evaluation by others, particularly in public performances. Fewer items in the _School Anxiety Scale_ load on the "test anxiety factor" than was the case for the _Test Anxiety Scale for Children_. Phillips reported that the _School Anxiety Scale_ correlates positively with the Proneness toward Neuroticism subscale of the _Children's Personality Questionnaire_.

Survey of Study Habits and Attitudes (SSHA)

This instrument was developed to measure study methods, motivation for studying, and certain attitudes toward scholastic activities which are important in the classroom. Obviously these represent important dimensions on which to measure students in our study. The purposes of the SSHA are (a) to identify students whose study habits and attitudes are different from those of students who earn high grades, (b) to aid in understanding students with academic difficulties, and (c) to provide a basis for helping such students improve their study habits and attitudes and thus more fully realize their best potentialities.

One form of the SSHA can be used with grades 7-12. It consists of 100 statements concerning study activities and attitudes. The student replies to each statement with one of the following answers: rarely, sometimes, frequently, generally, or almost always. The 100 statements were originally categorized by psychologists into four basic subscales. The particular subscales and subscores of the SSHA are as follows: Work Methods (use of effective study procedures, skill and efficiency in doing academic assignments) plus Delay Avoidance (promptness in completing assignments and ability to resist distractions) combine to yield a Study Habits score (a measure of academic behavior). Teacher Approval (feelings and opinions about teachers, their classroom behavior, and their methods) plus Education Acceptance (approval of educational objectives, practices and requirements) combine to yield a Study Habits score. The Study Habits score plus the Study Attitudes score combine to give a total Study Orientation score (an overall measure of study habits and attitudes).

Subscale intercorrelations ranged from .44 to .84 for men and from .27 to .76 for women, with medians of .53 and .39 respectively. Kuder-

Richardson Formula 8 estimates of internal consistency yielded coefficients for the four basic subscales ranging from .87 to .89. Test-retest correlations after a fourteen-week interval ranged from .83 to .88. The authors concluded that the four subscale scores are sufficiently stable through time to justify their use in predicting future behavior or in assessing the degree of change in study habits and attitudes after counseling (Brown and Holtzman, 1964).

Extensive validity evidence is presented by the authors. Validity coefficients reported on SSHA total scores with grade point averages ranged from .25 to .45.

Additional validity evidence showed that the partial correlation between SSHA total scores and grade point averages with scholastic aptitude held constant was highly significant, ranging from .41 to .47. According to Brown and Holtzman (1964) these results combined with others reported in the test manual clearly indicate the importance of the SSHA in providing measures of personal traits that are relevant to academic success but are not covered by scholastic aptitude tests.

In closing, it should be pointed out that the SSHA, being a self-report instrument, can be answered by students in such a manner that they will appear in a favorable light. Yet when it is responded to with honesty the SSHA appears to have the statistical qualities that enable it to provide information relevant to the academic success process.

Lorge-Thorndike Intelligence Test

The Lorge-Thorndike Intelligence Test yields a verbal IQ score, a non-verbal IQ score, and a total IQ score. The authors of the test feel that both verbal and non-verbal material test abstract intelligence, defined as the ability to work with ideas and relationships among ideas. They state that the following mental processes are descriptive of intelligent behavior and are sampled by their tests: (a) dealing with abstract and general concepts; (b) interpretation and use of symbols; (c) dealing with relationships among concepts and symbols; (d) flexibility in the organization of concepts and symbols; (e) utilizing one's experience in new patterns; and (f) utilizing "power" rather than speed in working with abstract materials.

The Lorge-Thorndike tests were restandardized in 1963 when they appeared in multi-level format. The restandardization population (grades 3-13) represented an appropriate stratified sample of American communities which were rated soci-economically as very high, high, average, low, and very low. Norms were developed for each group within these five levels.

Tittle (1972) states that the reliability of the test is appropriately assessed with the alternate forms method. The reliability coefficients range from .83 to .91 for the verbal battery and from .80 to .88 for the nonverbal battery. Because these correlations were computed on a population of a single grade, he feels that they are less apt to be spuriously high (as the correlations computed over all grades). Standard errors of measurement in terms of IQ points are given as an additional and highly desirable estimate of the tests' reliabilities. The data should allow the user to gain some understanding of the possible variability of scores, particularly at the extremes of the score distribution (Tittle, 1972).

Content, predictive, and construct validity are discussed in the technical manual. In the way of predictive validity, Tittle states that some correlations and related data are cited for the Lorge-Thorndike with achievement tests and school grades. The correlations with tests of achievement range from .60 to .70 with some up to .80. The nonverbal score typically provides the lower correlations within any set.

Studies of construct validity tend to include relationships with school achievement and with other tests that have acceptance as measures of intelligence. The most extensive study was conducted by Hieronymus and Stroud (1969). In their study of the comparability of IQ's obtained from the Lorge-Thorndike and four other intelligence tests administered to pupils in Iowa in grades 4, 7, and 10, they note that the correlations were quite variable, and in most cases below the reliabilities of the test, indicating that the tests were measuring somewhat different traits. The nonverbal IQ scores of the Lorge-Thorndike had lower correlations with the other intelligence test scores than the verbal. The Hieronymus and Stroud study also provides comparability data on the IQ's derived from the 1954 and 1963 editions of the Lorge-Thorndike. In their study the newer edition yielded slightly lower IQ's for the verbal score in grades 4, 7, and 10, and for the nonverbal in grades 7 and 10 (Tittle, 1972).

In summary, we include a quote of John Milholland (1959) speaking about the Lorge-Thorndike Intelligence Tests:

"These tests are admirable for the clarity with which objectives are stated and for the restraint exercised in the claims made for what they will do. They are frankly labeled intelligence tests, and we are told they are tests of abstract intelligence, defined as 'the ability to work with ideas and the relationships among ideas.'

The Lorge-Thorndike tests should be accorded a place among the best of our group intelligence tests. They are well designed, easily administered and scored, and, what is especially noteworthy, the uses recommended for them are reasonable and defensible."

SRA Achievement Series

Science Research Associates' Achievement Series measures the educational development of pupils in grades 2-9 in the following broad curricular areas: social studies, science, reading, arithmetic, language arts, and work-study skills. The tests are of the multiple-choice variety and are essentially power tests.

The Technical Report provides data on the makeup of the test booklets, the steps in the development of the tests in the series, and test characteristics, including score distributions and means and standard deviations for various grade levels. Kuder-Richardson (KR-20) reliabilities for the various tests range from the low .80's to the low .90's. These coefficients are indicative of generally high level structural quality and an acceptable level of consistency in test performance.

The product-moment intercorrelations among the various subtests generally run in the 0.50's and 0.60's. This seems to indicate that, while the separate tests are measuring several areas in common, each score is providing some unique information regarding educational achievement (Jones, 1959).

Predictive validity data is merely hinted at. In the Technical Report the authors state that "on the basis of the studies reported, it is quite evident that the SRA Achievement Series predicts high school achievement." Jones (1959) feels that the predictive validity of the series will require further investigation.

The specific SRA tests are as follows: It should be noted, however, that the Social Studies test will not be used in this study.

1. Social Studies. This test measures understanding and application of principles drawn from geography, history, government, and economics. There is only one test score.

2. Science. This test measures the pupil's knowledge and understanding of certain representative facts and principles of science. It stresses those concepts, generalizations, basic classifications, and cause-and-effect relationships customarily presented in elementary and junior high school science courses. There is only one test score.
3. Language Arts. This test measures a broad spectrum of skills in the use of language. There are separate scores for capitalization and punctuation, spelling, grammatical usage, and total language arts.
4. Arithmetic. This test measures various aspects of arithmetic achievement. There are separate scores for reasoning, concepts, computation, and total arithmetic.
5. Reading. This test uses complete stories to sample typical reading situations. There are separate scores for comprehension, vocabulary, and total reading.
6. Work-Study Skills. Test items in this area are based on materials typical of those found in textbooks, newspapers, and magazines used in the various curricular areas of elementary and junior high schools. There are separate scores for references, charts, and total work-study skills.

Mathematics Test

This test was taken from the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, and Price, 1963). It consists of 24, 5-choice work problems requiring arithmetic only. Factor analysis reveals that this test loads heavily on both numerical and general reasoning factors. The numerical factor is defined as the ability to manipulate numbers in arithmetical operations. The general reasoning factor is defined as the ability to solve a broad range of reasoning problems including those of a mathematical nature.

Letter Sets Test

This test was also taken from the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, and Price, 1963). It consisted of 30 questions. Each question was made up of five sets of four letters per set. The task is to find the rule which relates four of the sets to each other and to mark the one which does not fit the rule. Factor analysis reveals that this test loads heavily on an induction factor. Induction is defined as the associated abilities involved in the finding of general concepts that will fit sets of data. In other words, induction involves the forming and trying out of hypotheses.

[The descriptions of the Mathematics Test and Letter Sets Test were taken from the Manual for the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, and Price, 1963)]

Student Attitude Questionnaire

This questionnaire was designed to survey student attitudes towards the Jamesville-DeWitt science program (Hambleton, 1971). It consists of eight concepts which are relevant to the Jamesville-DeWitt individualized instructional program. Beneath each concept are eight bipolar adjective scales. The student is required to judge each particular concept in terms of these adjective scales. This procedure is the semantic differential technique which was developed by Osgood, Suci, and Tannenbaum (1957).

The particular concepts that were chosen were: teachers, science, student freedom in class, school, testing, individualized instruction, resource center, and enrichment program. These concepts were selected from a larger list of about thirty concepts because they seemed to be representative of the more important components of the science program

(Hambleton, 1971).

The adjective scales were evaluative in nature but could be divided into two categories, enjoyment and importance (Hecht, 1971). The enjoyment scales were unenjoyable-enjoyable, dull-exciting, boring-interesting, and unpleasant-pleasant. The importance scales were unimportant-important, useless-useful, worthless-worthwhile, and harmful-helpful.

Appendix B

Similarity Judgment Questionnaire and Student Sorting Task

Name:____ Date:______________

(Please Print)

SIMILARITY JUDGMENT QUESTIONNAIRE

The purpose of this questionnaire is to study how students perceive relationships among some important chemical concepts both before and after they have had instruction on them.

In the questionnaire you will encounter pairs of words. Your task is to examine the pair of words and make a judgment about the similarity of the concepts they represent. Then you should record your judgment by circling the appropriate number on the nine point rating scale which appears beside each pair of words. If you feel the words represent concepts which are very similar, circle one of the smaller numbers to the left end of the scale. If you feel that the words represent concepts which are very different, circle one of the larger numbers toward the right end of the rating scale.

To the extent that you can, try to spread your ratings over all nine points of the scale so as to represent nine levels of similarity. Also, try to base your judgments on the number of characteristics of each pair of concepts that are similar. In other words, be thoughtful in your approach to the task. Pairs of words that are similar in many characteristics should be rated more similar than pairs that are similar in few characteristics.

The concepts that appear in the questionnaire are the following:

proton	nucleus
ion	atomic number
neutron	atomic weight
element	mixture
radical	shell or energy level
compound	oxidation number
electron	

Read over the concepts to acquaint yourself with them. Then proceed to make the judgments requested in the questionnaire.

In a case where you are not familiar with the concepts use your best judgment.

Example		Very Similar		Moderately Similar		Moderately Different		Very Different	
electron - proton	1	2	3	(4)	5	6	7	8	9

In the example above the student thought electron and proton were moderately similar. If for example he wanted to indicate that they were very different he would circle number 9.

	Very Similar		Moderately Similar			Moderately Different		Very Different	
1. proton - ion	1	2	3	4	5	6	7	8	9
2. oxidation number - neutron	1	2	3	4	5	6	7	8	9
3. shell or energy level - element	1	2	3	4	5	6	7	8	9
4. mixture - radical	1	2	3	4	5	6	7	8	9
5. atomic weight - electron	1	2	3	4	5	6	7	8	9
6. atomic number - compound	1	2	3	4	5	6	7	8	9
7. nucleus - proton	1	2	3	4	5	6	7	8	9
8. neutron - ion	1	2	3	4	5	6	7	8	9
9. element - oxidation number	1	2	3	4	5	6	7	8	9
10. radical - shell or energy level	1	2	3	4	5	6	7	8	9
11. electron - mixture	1	2	3	4	5	6	7	8	9
12. compound - atomic weight	1	2	3	4	5	6	7	8	9
13. nucleus - atomic number	1	2	3	4	5	6	7	8	9
14. proton - neutron	1	2	3	4	5	6	7	8	9
15. ion - element	1	2	3	4	5	6	7	8	9
16. oxidation number - radical	1	2	3	4	5	6	7	8	9
17. shell or energy level - electron	1	2	3	4	5	6	7	8	9
18. mixture - compound	1	2	3	4	5	6	7	8	9
19. atomic weight - nucleus	1	2	3	4	5	6	7	8	9
20. atomic number - proton	1	2	3	4	5	6	7	8	9
21. element - neutron	1	2	3	4	5	6	7	8	9
22. radical - ion	1	2	3	4	5	6	7	8	9
23. electron - oxidation number	1	2	3	4	5	6	7	8	9
24. compound - shell or energy level	1	2	3	4	5	6	7	8	9
25. nucleus - mixture	1	2	3	4	5	6	7	8	9
26. atomic number - atomic weight	1	2	3	4	5	6	7	8	9
27. proton - element	1	2	3	4	5	6	7	8	9
28. neutron - radical	1	2	3	4	5	6	7	8	9
29. ion - electron	1	2	3	4	5	6	7	8	9
30. oxidation number - compound	1	2	3	4	5	6	7	8	9

		Very Similar		Moderately Similar			Moderately Different		Very Different	
31.	shell or energy level - nucleus	1	2	3	4	5	6	7	8	9
32.	mixture - atomic number	1	2	3	4	5	6	7	8	9
33.	atomic weight - proton	1	2	3	4	5	6	7	8	9
34.	radical - element	1	2	3	4	5	6	7	8	9
35.	electron - neutron	1	2	3	4	5	6	7	8	9
36.	compound - ion	1	2	3	4	5	6	7	8	9
37.	nucleus - oxidation number	1	2	3	4	5	6	7	8	9
38.	atomic number - shell or energy level	1	2	3	4	5	6	7	8	9
39.	atomic weight - mixture	1	2	3	4	5	6	7	8	9
40.	proton - radical	1	2	3	4	5	6	7	8	9
41.	element - electron	1	2	3	4	5	6	7	8	9
42.	neutron - compound	1	2	3	4	5	6	7	8	9
43.	ion - nucleus	1	2	3	4	5	6	7	8	9
44.	oxidation number - atomic number	1	2	3	4	5	6	7	8	9
45.	shell or energy level - atomic weight	1	2	3	4	5	6	7	8	9
46.	mixture - proton	1	2	3	4	5	6	7	8	9
47.	electron - radical	1	2	3	4	5	6	7	8	9
48.	compound - element	1	2	3	4	5	6	7	8	9
49.	nucleus - neutron	1	2	3	4	5	6	7	8	9
50.	atomic number - ion	1	2	3	4	5	6	7	8	9
51.	atomic weight - oxidation number	1	2	3	4	5	6	7	8	9
52.	mixture - shell or energy level	1	2	3	4	5	6	7	8	9
53.	proton - electron	1	2	3	4	5	6	7	8	9
54.	radical - compound	1	2	3	4	5	6	7	8	9
55.	element - nucleus	1	2	3	4	5	6	7	8	9
56.	neutron - atomic number	1	2	3	4	5	6	7	8	9
57.	ion - atomic weight	1	2	3	4	5	6	7	8	9
58.	oxidation number - mixture	1	2	3	4	5	6	7	8	9
59.	shell or energy level - proton	1	2	3	4	5	6	7	8	9
60.	compound - electron	1	2	3	4	5	6	7	8	9

		Very Similar		Moderately Similar		Moderately Different		Very Different	
61. nucleus - radical	1	2	3	4	5	6	7	8	9
62. atomic number - element	1	2	3	4	5	6	7	8	9
63. atomic weight - neutron	1	2	3	4	5	6	7	8	9
64. mixture - ion	1	2	3	4	5	6	7	8	9
65. shell or energy level - oxidation number	1	2	3	4	5	6	7	8	9
66. proton - compound	1	2	3	4	5	6	7	8	9
67. electron - nucleus	1	2	3	4	5	6	7	8	9
68. radical - atomic number	1	2	3	4	5	6	7	8	9
69. element - atomic weight	1	2	3	4	5	6	7	8	9
70. neutron - mixture	1	2	3	4	5	6	7	8	9
71. ion - shell or energy level	1	2	3	4	5	6	7	8	9
72. oxidation number - proton	1	2	3	4	5	6	7	8	9
73. compound - nucleus	1	2	3	4	5	6	7	8	9
74. electron - atomic number	1	2	3	4	5	6	7	8	9
75. radical - atomic weight	1	2	3	4	5	6	7	8	9
76. element - mixture	1	2	3	4	5	6	7	8	9
77. neutron - shell or energy level	1	2	3	4	5	6	7	8	9
78. ion - oxidation number	1	2	3	4	5	6	7	8	9
79. ion - proton	1	2	3	4	5	6	7	8	9
80. ion - neutron	1	2	3	4	5	6	7	8	9
81. atomic number - nucleus	1	2	3	4	5	6	7	3	9
82. neutron - element	1	2	3	4	5	6	7	8	9
83. atomic weight - atomic number	1	2	3	4	5	6	7	8	9
84. element - radical	1	2	3	4	5	6	7	8	9
85. mixture - atomic weight	1	2	3	4	5	6	7	8	9
86. shell or energy level - mixture	1	2	3	4	5	6	7	8	9
87. compound - radical	1	2	3	4	5	6	7	8	9
88. electron - compound	1	2	3	4	5	6	7	8	9
89. oxidation number - shell or energy level	1	2	3	4	5	6	7	8	9
90. nucleus - electron	1	2	3	4	5	6	7	8	9
91. proton - oxidation number	1	2	3	4	5	6	7	8	9

NAME: ______________________ DATE: ______________

STUDENT SORTING TASK

The purpose of this task is to aid us in seeing how students perceive relationships among some chemical concepts before and after completion of the appropriate module of instruction.

A packet containing 13 cards is attached to this page. On each card is printed the name of a chemical concept. Your task is to examine the concepts and put those concepts that appear similar into separate piles. You may sort the concepts into any number of piles. Just remember that all the concepts that you put together into a pile should be considered by you as having similar characteristics. The concepts that you put in a different pile should be considered by you to be different from those in other piles.

Read over the concepts to acquaint yourself with them. Then proceed to sort the concepts into as many piles as you feel necessary.

When you have finished your sorting, record your results at the bottom of this page.

To distinguish the piles, assign them different numbers. If you happen to have four piles then number them 1, 2, 3, and 4. It doesn't matter how you assign the numbers so long as each pile has a different number. In reporting your results, indicate the numbers on the cards in the different piles.

For example, if you formed four piles and had cards numbered 4, 5, 7, and 10 in one of the piles; cards numbered 1, 9, 12, and 13 in another pile; cards numbered 2 and 11 in another pile; and cards numbered 3, 6 and 8 in still another pile you would record your results as follows:

Pile 1	Pile 2	Pile 3	Pile 4
4,5,7,10	1,9,12,13	2,11	3 6,8,

Any questions? Now go ahead and remove the cards attached to this sheet of paper and begin the task of sorting.

RESULTS

proton 1	ion 2	neutron 3
element 4	radical 5	compound 6
electron 7	nucleus 8	atomic number 9
atomic weight 10	mixture 11	shell or energy level 12
xidation number 13		

Note: The 13 concepts on this page were used in the sorting task. The concepts were presented to the students in a packet of 13 pieces of paper with one concept appearing per page.