ABSTRACT

STABILITY OF SEMANTIC MEANING SPACE AND
CHANGE IN CONCEPT MEANING DURING
TEACHER TRAINING

By

Richard J. Stiggins

This study was designed to further explicate the nature of semantic meaning space as it relates to educational concepts presented during teacher training. It sought to test the potential of the semantic differential scaling procedure for systematically assessing a noncognitive outcome of teacher training. More specifically, research hypotheses were generated from four major questions:

1. What are the primary dimensions which characterized the semantic meaning space with respect to educational concepts?
2. What is the nature of the effect of teacher training on these dimensions? Do they remain constant from the beginning to the end of instruction?
3. Is there evidence that instruction influences the connotative meaning of individual concepts presented during teacher training?
4. Is there evidence that the instructor plays a role in these changes, if they occur?

Two additional questions addressed the potential role of the semantic differential in providing feedback of an evaluative nature for the planning of future instruction:

5. Are the connotative meanings of concepts and the changes in those meanings appropriate for teacher training?
6. Are there other relevant hypotheses which might be addressed by employing this scaling procedure?

The method devised for seeking answers to these questions was the administration of a pilot tested semantic differential instrument to 252 undergraduate education majors during the first and last weeks of their first teacher training course. Eleven concepts were rated on 15 scales. The concepts were chosen to represent technical educational psychology terms, concepts related to the interpersonal demands of teaching, and concepts in no way related to education. The first two groups included terms systematically presented in the course, and the latter group was included as a control. The scales were selected to tap three dimensions of meaning: evaluative (favorable-unfavorable), potency (weak-powerful), and activity (active-passive).

The results indicated that more than three dimensions of meaning were tapped. Unlimited maximum likelihood factor analysis revealed that the most parsimonious explanation of scale interrelationship was a four-factor solution: evaluative, personal evaluative (pleasant-unpleasant), leniency (severe-lenient), and potency (weak-powerful, active-passive).

These four latent constructs remained very stable over time. Multivariate analysis of variance of factor scores revealed that there were significant changes in the meanings of the instructional concepts within this stable frame of reference, but that there were no changes in the meanings of noninstructional terms. No differential effect of instructors was found. Further, it was speculated that the meanings and changes were appropriate for teachers and teacher training. However, additional refinement of the scaling procedure is required before other relevant hypotheses can be tested.

STABILITY OF SEMANTIC MEANING SPACE AND CHANGE IN CONCEPT MEANING DURING TEACHER TRAINING

By

Richard J. Stiggins

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services,
and Educational Psychology

1972

619463

ACKNOWLEDGMENTS

The selection, clarification, and execution of the research reported here was a process which was as true a representation of myself as it is possible for me to attain at this point in my professional development. I am, in turn, a reflection of interactions with many people who shared greatly in that development: Joe L. Byers, who shared and clarified so many things, including the completion of this dissertation; Robert L. Ebel, who shared the task of planning the course of my professional development; and Robert C. Craig, Frederick R. Ignatovich, and Willard Warrington, all of whom provided stimulus for growth. Their encouragement, support, criticism, questioning, patience, and acceptance are greatly appreciated. The nature of the contributions of each of these people was clearly apparent, even through the final steps of doctoral study. For this, I thank them.

I am also indebted to the staff and students of Education 200 for their valuable support and assistance.

Maintaining order in the face of total confusion can be a difficult and trying task. Only the wife of a doctoral candidate knows how difficult. Thank you, Nancy.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vii
LIST OF APPENDICES	ix
Chapter	
I. INTRODUCTION	1
Rationale for the Research	6
Component of Teacher Training to Be Assessed	10
Assessment Technique	12
Context of the Assessment	16
Questions to Be Addressed	18
II. REVIEW OF RELEVANT AND RELATED RESEARCH	22
The Semantic Differential Technique	23
The EPA Structure and the Language of Emotion	29
Psychometric Qualities of the Semantic Differential Technique	31
Applications of the Semantic Differential to Educational Training and Change	42
Conclusion	46
III. METHODOLOGY	48
Introduction	48
Instrument Development Pilot Test	50
Pre/Post Administration	55
Analysis	67
Conclusion	90
IV. RESULTS	92
Introduction	92
Scale Interrelationships and Factor Analysis	92
Tests of Factor Structure Invariance	115

Page

Test of Change in Concept Meaning. 118
Instructor Role in Meaning Change. 129
Conclusion 137

V. DISCUSSION 139

Introduction 139
Scale Interrelationships and
Factor Analysis. 139
The Question of Concept Meaning Change . . . 146
Instructor Differences in Change in
Meaning Intensity. 151
The Question of Other Testable Hypotheses. . 154
Conclusion 157

BIBLIOGRAPHY 159

APPENDICES 168

LIST OF TABLES

Table		Page
1.	Test-Retest Reliabilities in Applications of the Semantic Differential.	33
2.	Validity Coefficients for Semantic Differential Scores (s) and Thurstone Scale Scores (t).	36
3.	Results of the Chi Square Test of Goodness of Fit of the Three-Factor Model to the Variance-Covariance Matrix of Responses to Each Concept for Both Pretest and Posttest Observations.	93
4.	Rotated Factor Matrices for Four-Factor Solution of Pooled Correlation Matrices, Pre and Post.	111
5.	Results of All Statistical Tests Assessing The Degree of Factor Invariance Over Time When Comparing the Pooled (Over Concepts) Pretest and Posttest Factor Solution.	116
6.	Multivariate Analysis of Variance Results of Change in the Evaluative Dimension	120
7.	Multivariate Analysis of Variance Results of Change in the Potency Dimension.	121
8.	Multivariate Analysis of Variance Results of Change in the Leniency Dimension	122
9.	Multivariate Analysis of Variance Results of Change in the Personal Evaluative Dimension	123
10.	Multivariate Analysis of Variance Results of Change in the Overall Meaning Intensity. .	124
11.	Factor Score Means for Each of the Four Major Meaning Dimensions for Each Concept . .	128

Page

12. Results of Manova of Instructor Differences in Pretest Meaning and Change in Meaning on the Three Dimensions Where a Significant Change Was Detected. 134

13. Results of Analyses of Variance of Change in Meaning Over Instructors for Individual Concepts. 137

LIST OF FIGURES

Figure		Page
1.	Example of an Application of Cattel's Scree Test.	71
2.	Example of a Concept in Semantic Meaning Space	78
3.	Fit of the Factor Solutions for the Concept Myself as a Teacher	94
4.	Combined Data Assessing Fit of Factor Models for the Concept Nonverbal Behavior.	95
5.	Fit of the Factor Solutions for the Concept Myself.	96
6.	Fit of the Factor Solutions for the Concept Questioning and Listening Skills.	97
7.	Fit of Factor Models for the Concept Behavioral Objectives	98
8.	Fit of Factor Models for the Concept Reinforcement	99
9.	Fit of Factor Solutions for the Concept Respondent Learning	100
10.	Fit of Factor Solutions for the Concept Shaping	101
11.	Fit of Factor Solutions for the Concept Physician	102
12.	Fit of Factor Models for the Concept Religion.	103
13.	Fit of Factor Solutions for the Concept Marijuana	104
14.	Fit of Factor Solutions for the Concept Behavioral Objectives Employing 12 of the 15 Scales.	108

Page

15. Fit of Factor Solutions for the Pretest and Posttest Results Pooled Over Concepts. 110

16. Graphic Representation of Pretest and Posttest Meaning Assigned to IPL Concepts: Myself as a Teacher, Nonverbal Behavior, Questioning and Listening Skills, Myself. 130

17. Graphic Representation of Meaning Assigned to Carrel Concepts: Behavioral Objectives, Reinforcement, Respondent Learning, Shaping 131

18. Graphic Representation of Pretest and Posttest Meaning Assigned to Non-instructional Concepts: Physician, Religion, Marijuana 132

19. Sample of Data Representing Role of Instructor in Meaning Change. 136

LIST OF APPENDICES

Appendix Page

A. SUPPLEMENTAL GENERAL INFORMATION 169

B. PILOT TEST SUPPLEMENT. 173

C. PRE/POST ADMINISTRATION SUPPLEMENT 177

D. RESULTS SUPPLEMENT 180

CHAPTER I

INTRODUCTION

Among the many issues confronting educators in America, one of the most important is clearly a definition of the role which educational institutions are to play in our society. There are those who point to cognitive skills and the teaching and learning of subject matter competencies as the primary mission of these institutions, and there are those who would claim a primary role for affective goals and objectives. (For a discussion of these points of view, see Ebel, 1972.) With the increasing call for accountability with its many and varied definitions, there is a pervading need for evidence of success in academic ventures in the form of data. This call for accountability provides ample grounds for heated discussion between the two camps mentioned above, because those who support efforts in the area of affective growth are hard pressed, indeed, when confronted with the task of presenting evidence demonstrating the effectiveness of their endeavors. This is not the case, however, for those educators who back the cognitive mission. Achievement testing has developed into a sophisticated apparatus for assessing the outcome of student and instructor efforts. It is being pressed into use in national and state accountability

assessments. To deny the critical role of the cognitive mission and its accompanying system of evaluation would be difficult to defend. However, those who support the primacy of the cognitive endeavors hasten to admit that there are affective outcomes of any educational experience. One reason that these may be considered secondary by-products, is their abstract nature and the difficulties in quantification which accompanies evaluation of affective goals. Herein lies the problem of evidence demonstrating the success or failure of affective educational endeavors.

It is proposed here that this need not be the case in every instance of affective educational goals. Consider an example in the professional preparation of teachers. The traditional goal of training programs in education has been to produce professionals equipped with tools with which to perform the tasks demanded by the profession. Specific capabilities are taught to attain specific competencies because those experienced in the field have determined that these competencies are of assistance in developing an effective learning environment. In other words, those concerned with the professional preparation of teachers advocate the use of the pedagogical tools and skills included in the teacher training curriculum. This is, of course, the case in any professional preparation sequence. But, unlike other fields, when the teacher has completed the sequence, he proceeds into an academic situation where, aside from global restrictions and limits suggested by his professional

judgment, he has great freedom to choose the tools and procedures to be applied. Each teacher is responsible for the learning environment in which students dwell in any given classroom.

The determination of, or selection of, tools which will be applied is a critical process. But what is the process? Is it enough to know that when a teacher is sent into a work situation, lacking experience, he knows what tools are at his disposal and how to use them? The answer to the latter question is clearly, "no." Knowledge of behavior is a necessary but not a sufficient condition for the manifestation of the learned behavior when practicing autonomously in the classroom. In addition to knowledge, there is at least one other dimension that is critical to the manifestation of the desired behavior. That is the value that the teacher trainee places in each potential procedure. This may be described as a positive affective disposition toward the use of the required or desired skill or tool. For example, knowledge of the mechanics and theory of achievement testing will not necessarily lead to its use in the classroom. The knowledge must be supplemented by a motivation to employ this procedure. And so it is with each of the tools, concepts, and procedures the teacher takes from undergraduate and graduate training. Motivation, as well as knowledge, will determine what will be employed (and to what end) after the last final examination.

But when educators endeavor to evaluate the effectiveness of instruction in teacher training, the primary emphasis (indeed the only emphasis) is placed on assessment of cognitive outcomes. This leads directly back to the problem described at the outset, which cited the relatively more advanced state of the science of measuring cognitive outcome when compared with the instrumentation and methodology available to assess the affective (motivational) concerns. The research reported here is an attempt to begin dealing more systematically with this affective component of teacher training.

In order to avoid confusion, there is an issue implied in an argument of this nature which must be clarified and discussed before proceeding. This clarification begins by stating that the above argument does not demonstrate a position favoring the primacy of affective outcomes. Cognitive outcomes are critical and the teacher trainee should continue to be held accountable for cognitive growth in the same manner in which he has been traditionally. However, the responsibility for the development of the motivation or predisposition to employ pedagogical skills and concepts cannot and should not be left to the student alone. Where, then, should this responsibility reside?

The simple answer to this question is that the responsibility for the motivational predisposition lies primarily with the teacher, and for good reason.

Consider Glaser's (1962) conceptualization of instruction which characterized the pedagogical task as planning instruction, carrying out teaching strategies, measuring achievement, and allowing the results to feed back into instructional planning and strategies. If the results aren't as desired in the cognitive realm, one (or both) of two occurrences has been manifested: either the instruction was inadequate or the student has failed in his responsibility to study. In either case, valuable information is provided by the testing. Now apply the same model to measurement of the motivational predisposition defined earlier. It was stated that knowledge is a necessary but not a sufficient condition for later manifestation of the desired behavior. Consequently, assume for the moment that learning has occurred and that students know how and when to apply the tools, concepts, and procedures taught. If a measure is taken and it is determined that the students do not react favorably toward the concepts, do not see them as powerful, or able to play an active role in the learning environment, to what may these reactions be attributed? Clearly one of two things has occurred. Either the concepts presented do not command these attributes in their own right or the instruction has failed to demonstrate that the concepts are to be so valued. In either case, the major responsibility for this goes to the person who selected the concepts and/or planned the instruction. Again, in either case, valuable information has been gathered for the feedback loop

and the planning of future instruction. It should be the inherent responsibility of the instructor to use such information to make adjustments. It is difficult to conceive of a dedicated instructor who would not use such information, but the problem appears in the fact that such measures are almost never taken. When instructional evaluations are made, they are generally evaluations of the instructor and not the content presented. Courses proceed for term after term with students passing exams with flying colors, much to the satisfaction of their instructors, and then, when it becomes time to apply the concepts, tools, and procedures which have been learned, they choose not to do so.

In short, what is advocated here is that the same model for the evaluation and reworking of instructional sequences be applied to both cognitive and affective concerns in teacher training so that the training sequence will have maximum impact on the student both cognitively and affectively. A means for accomplishing this is developed here.

Rationale for the Research

One need not search far in the education literature, and more specifically the research literature on education, to find support for the type of research reported here. A brief summary is presented below in the form of comments on the need for the systematic and scholarly evaluation of

educational endeavors and the need for research on teacher training.

The Need for Evaluative Research

The need for the development of instrumentation and methodology with which to carry out evaluation of educational programs is restated throughout the literature, particularly in such volumes as the American Educational Research Association Monograph Series on Curriculum Evaluation, the 1969 National Society for the Study of Education Yearbook edited by Ralph W. Tyler, Educational Evaluation: New Roles, New Means, and the Handbook on Formative and Summative Evaluation of Student Learning by Bloom, Hastings, and Madaus (1971). The current status of the concern for the scholarly advancement of evaluation as a discipline is demonstrated by the 1970 creation of a new division of the American Educational Research Association, Division H., Evaluation of Educational Programs.

Those who have been concerned with providing orientation and direction in the endeavor of exploring and developing new methodologies and instrumentation, particularly Stake (1967), Gagné (1967), Provus (1969), Stufflebeam (1968), and Scriven (1967), all repeat the same themes and admonitions. Evaluation should go beyond assessment of cognitive outcomes to a consideration of the antecedent variables, to the goals of a program, to the transactions which characterize

a particular program, and even more importantly, to the critical interactions among these variables.

The nature of the research which will result from these types of considerations is applied research in a very true sense and can be of great assistance in the planning of instruction. After discussing the nature and role of such evaluative endeavors, Ahman summarizes:

> Curriculum developers and educators are tempted to de-emphasize evaluation because of the complex and sometimes ill-defined methodological problems present. . . . Evaluation is a secondary activity in the development of curricula, but still one which needs to receive a major share of attention from the principle developers. As Scriven points out, the stakes are high (1967, p. 89).

The Need for Research on Teacher Training

There can be little doubt that extensive research on teacher training must be high on the list of priorities of educational psychologists in the coming decade. Gage introduces his Handbook of Research on Teaching (1963) with an indication of the reason for this:

> In recent decades, such research [on teaching] has lost touch with the behavioral sciences. It has not drawn enough nourishment from the theoretical and methodological developments in psychology, sociology and anthropology. Nor has it provided those disciplines with return stimulation, as it did in its earlier period. To remedy this condition--to bring research on teaching into more fruitful contact with the behavioral sciences--is the purpose of this Handbook (Preface).

Research on teaching can, of course, take place at many levels: at the point of selection of undergraduate

teacher trainees, during undergraduate and graduate level training, at the time of selection to fill teaching positions, and while the teacher is working in the field. Strong arguments can be constructed for the importance of research at all levels, because of the role this inquiry could play in the learning environments provided by teachers.

For the purpose of this research, however, it is necessary to briefly discuss the rationale for research at one of these levels, that of undergraduate training and the formative evaluation of particular programs therein.

Undergraduate teacher training programs are critical in the professional socialization of future educators, primarily because of the nature of the discipline. An effective teacher is, at best, very difficult to define. This lack of an acceptable criterion toward which to strive presents immense problems to those concerned with professional preparation of teachers. First, the selection of a potentially successful teacher in any systematic way is very difficult. Second, there are few criteria against which to judge curricular adequacy. And, finally, accurate judgment of a student's "successful" completion of a training program is very difficult.

Part of the reason for these difficulties may be that educators have concentrated too heavily on cognitive assessment at the expense of the truly critical affective assessment of teacher training. If the dependent variable

being examined here is fruitful, at least the last two difficulties mentioned above will be able to be systematically addressed. New instrumentation will be available for the evaluation and determination of instructional adequacy and this may lead to some insight into the realm of determination of success in a more clearly defined program.

Component of Teacher Training to Be Assessed

Having discussed the rationale for the evaluative research on teacher training programs, it is now appropriate to return to the specifics of the research reported. It has been argued that motivational and cognitive outcomes of teacher training are important for the production of effective teachers. The motivational component has been evaluated less effectively than has the cognitive. The procedure discussed below is aimed at an important first step in the remediation of this state of affairs.

Affect has been defined in numerous ways in education and psychology. Krathwohl, Bloom, and Masia (1964) in their Taxonomy of Educational Objectives (Handbook 2, Affective Domain) list such behaviors as awareness, willingness to receive, acquiescence, willingness and satisfaction in responding, etc. Psychological research in the affective area has been characterized by phenomenal fields (Combs and Snygg), needs (Murray), values (Allport), and attempts have been made to measure interests (Strong) as well as global personality characteristics (MMPI, Rorschach, TAT).

Those who advocate abandonment of efforts to contend with affective components of education as described above find much support for these global concepts with high levels of abstraction and accompanying measurement difficulties. Little that is systematically helpful in solving evaluative problems in education can be derived from these conceptualizations. Gage (1963) points out that a more effective strategy might be:

> Rather than seek criteria of effectiveness of teachers in many, varied facets of their role, we may have better success with criteria of effectiveness in small, specifically defined aspects of the role. Many scientific problems have eventually been solved by being analyzed into smaller problems, whose variables are less complex (p. 120).

If global definitions of affect have been ineffective contributors to the educational evaluation endeavors, then as Gage suggests, perhaps a measure of the specific affect described in the introduction can be of assistance. The affect described was concerned with the favorable or unfavorable attitude of future teachers toward the concepts and tools presented in training. In addition, it was defined as the degree to which the student is able to attribute power to the tools at his disposal and the extent to which these tools are able to play an active role in the learning environment if they are employed there.

In short, one may be concerned with two measures of outcomes of instruction in teacher training: cognitive and affective. This research is concerned with the affective. There are numerous directions in which to pursue affect.

This research defines it as motivational predisposition to employ the knowledge which has been acquired. Motivation can be defined in numerous ways, but the one of interest here is the degree favorability, power, and activity which the teacher trainee perceives in the concepts presented. It is hypothesized that if these meanings associated with the concepts can be measured, then it will be possible to correlate the measures with actual classroom behavior to determine the validity of this definition of motivation. It is the "if" clause above which is the primary interest of this research. That is, an attempt is made to operationalize a possible measure of motivational predisposition for later validation.

These possible indicators of "affect" can be, and are, directly influenced by instruction and should therefore be evaluated and should serve as stimulus for change in instruction if it is indicated. This level of specificity of definition of affect may provide a smaller, more visible target of the type recommended by Gage.

Assessment Technique

In our endeavor to measure these motivational predispositions, we can find assistance in the linguistic realm, specifically from the work of Osgood and his associates (1957) in the development of the semantic differential scaling technique. This technique was developed in order to provide a systematic procedure for assessing the connotative

meaning of concepts or objects and is a combination of psychometric scaling and linguistic assessment. This combination yields an index of the particular affect being measured here, which are the motivational predispositions.

Very simply, this technique involves the use of bipolar adjectives, usually separated by a 7-point scale, which the respondent employs to modify a concept. The manner in which he uses each scale reveals the connotational meaning he ascribes to the concept. For example, if the subject is describing the meaning of the psychological concept reinforcement on a scale which has its ends anchored by the adjectives weak and powerful, and he sees the concept as very powerful, his responses would be very close to the powerful end of the scale. This same concept could be rated on numerous other adjective scales and the ratings would reflect the connotative meaning that the concept reinforcement has for each respondent.

There is a fundamental distinction to be made between the meaning of a concept measured by the semantic differential and the meaning of that same concept as measured by an achievement test. This distinction is critical to the definition of affect being presented here. Any learned concept has a large domain of meaning associates which are the result of experiences with that concept. Part of that domain is its denotative meaning and this part is a reflection of the substantive definitional meaning of the concept. For example, one who possesses knowledge of pedagogical

skills as a result of teacher training possesses the denotative meaning of the skills. This speaks to the issue of how, where, and when to employ the tools and skills and is traditionally assessed by an achievement test. This denotative meaning is supplemented by additional associations made with the concepts which reflect the meaning of a word apart from its explicit description or definition. Concepts commonly trigger an emotional or affective reaction which is every bit as much a part of the meaning of a concept as its definition. It is this type of emotive association or meaning which is being tapped by the semantic differential scaling technique. In this study, only a part of that emotional or affective reaction is of interest. That subset of reactions consists of the aspects of connotative meaning which appear to portray the motivational predisposition to use the concepts at some later time. In other words, there is a need to know if the concepts taught in teacher training trigger favorable reactions, command connotations of power, and possess the qualities necessary to be active contributors in a learning environment. If these can be measured, then later research can validate the assumption implicit in this argument: that these measures are in fact indicators of later classroom behavior. Such an assumption would appear to have some merit because of the validity studies done in areas other than teacher training, which demonstrate that SD measures of connotative meaning are related to behavior. However, discussion of these will be

postponed until Chapter II, where the validity of the technique is discussed.

In their discussion of the Measurement of Meaning, Osgood, et al. (1957) present a model based on a classical conditioning paradigm which is designed to explain the manner in which connotative meanings become attached to a concept. A complete discussion of this model will appear in Chapter II. However, some of the major points are worthy of note here. Language and linguistic elements play a critical role in the attachment process, because language is the primary vehicle for the transmission of complex meaning. In each learner, there resides an elaborate system for decoding verbal input and integrating into a structure of perceptions (cognitive and emotional) which has been created through experience. The purpose of instruction (i.e., teacher training) is to bring about new cognitive and affective (motivational) associations by using language and relying on an integration of these symbols with past experience. It follows from this that language can be profitably employed to assess the associations after instruction. For example, when one wishes to measure the nature and quality of subject matter connections, one uses language to establish a set of controlled conditions (test items) to take a "reading" of the associations which have been made. It is argued here that this same strategy can be used to assess motivational associations by establishing a set of controlled

conditions to read out connotative meanings using language as the vehicle.

This is the intent of the semantic differential technique. As Osgood points out (1957, pp. 19-20), there are several levels at which one may make such an assessment. For example, it is possible to provide a stimulus and allow the respondent total freedom in his verbal response as in a projective psychometric test. However, this is most difficult to interpret. He then proposes that limits be placed on the linguistic elements to be used by the respondent to describe the stimulus, leading to the semantic differential format described earlier. The respondent to an instrument employing this technique is simply asked to describe his associations with a concept on a limited number of descriptive scales which are selected to tap the meaning alternative of interest to the psychometrician. By so prescribing the elements of the description, the researcher is able to control for individual differences in vocabulary and general grammatical and verbal facility.

Such a technique would, therefore, seem to have the potential to supply useful information about the impact of instruction on students and the meanings they come to associate with concepts presented.

Context of the Assessment

Three points have been made thus far: there was a need to further explicate the process of becoming a teacher

through evaluative or applied research; a primary phase of such research should be to supplement cognitive outcomes of teacher training with information concerning affective outcomes; and, an assessment technique was available which has the potential for supplying such information. Given a teacher training context, a simple application of the technique would reveal the utility of gathering such information and the role such affective data could play in the planning of instruction. Such a context was supplied by the initial course in the teacher training sequence at Michigan State University, which provides the student with a first look at the task and personal demands of teaching.

The course is structured into two phases which run concurrently and which deal with the task demands as well as the personal demands of the profession. Task demands are defined as the knowledge and skills required to establish and maintain an effective learning environment (Henderson, et al., 1972). This is the substantive content presentative phase where principles of planning instruction (i.e., reinforcement, shaping, respondent learning) are presented. Instruction on the nature and application of these tools is carried out in individual study carrels via cassette tapes, slide presentations, and instructional film loops.

The other phase of the course deals with the personal and interpersonal demands of teaching such as the role that questioning and listening skills can play in a learning environment or the importance of nonverbal behavior

in providing feedback to the teacher. These concepts are confronted in the context of an interpersonal process laboratory (IPL) of no more than 15 students under the supervision of a graduate teaching assistant. The students attempt to develop in themselves the capability of dealing effectively with the interpersonal relationships which will be demanded of them as teachers.

Emphasis in the IPL's is on self-growth, while in the study carrels it is on helping others to grow. Concepts, tools, and procedures, which will be of assistance in practice, are systematically presented, and it is within this context that the meanings associated with these concepts were assessed.

Questions to Be Addressed

According to the opening argument, from the viewpoint of those coordinating and planning instruction, such tools, concepts, and procedures as behavioral objectives, reinforcement, shaping, respondent learning, nonverbal behavior, and listening and questioning skills (among others) can play an important role in the development of an effective learning environment. But, it is insufficient to know that students have mastered objectives which stipulate that they know how and when to use these learning facilitators. It is argued here that, as a result of instruction, they ought to ascribe value to them in the form of a positive disposition toward their use. In addition, students should

perceive the learning facilitators as powerful and invest them with the capability of playing an active role in the learning environment. In order to adequately assess the worth of the semantic differential (SD) technique as a potentially relevant, dependent measure in the study of this affect, answers to the following questions will be sought:

Is there a systematic set of dimensions which individuals consistently employ as a frame of reference on which to rate the connotative meaning of educational concepts? The answer to this question will be inferred from the answers to two more specific questions: Are the primary dimensions which characterize this frame of reference Evaluation, Potency, and Activity, as might be aniticpated from prior SD research? Do these factors remain evident from the beginning of instruction to the end of instruction?

Does instruction have a systematic effect on the connotative meaning of the individual concept taught? The answer to this question is conditioned to some extent on the stability of the frame of reference. This gives rise to two related questions: If there is instability in factor structure, what is the nature of the post-instruction structure in relation to that which existed prior to instruction? If there is stability, that is the same factors characterize the post-instruction responses as represent the pre-instruction responses, do meanings of concepts change within the frame of reference? This is the critical question

because the answer sheds light on the influence of instruction on the meanings of specific concepts taught.

There are a number of influences which could contribute to a systematic change in the connotative meanings of concepts. Among the most important of these is the instructor's role in the change, and it is here once again that a test of the utility of the SD technique becomes apparent. The test is, of course, the degree to which instructor differences can be detected and this test will be carried out.

Another direction from which to approach the question of the relevance of this technique is to assess the degree to which it is able to supply useful information for the planning of future instruction. For example, a relevant contribution would be made if SD ratings revealed inappropriate meanings which are not influenced or changed to appropriate meanings by instruction. In this case, the appropriateness of the meanings would be defined by those who are responsible for planning the instruction. The results of this research will be interpreted with respect to this appropriateness issue.

The final assessment of the potential of this technique of measuring motivational predisposition is to seek other relevant hypotheses which might be addressed with such measurements. Some examples of these will be discussed in later chapters.

Each of these questions is systematically addressed in the research in order to assess the validity and utility of the SD in rating affective components of a teacher training experience.

CHAPTER II

REVIEW OF RELEVANT AND RELATED RESEARCH

There is a large body of research on the semantic differential (SD) both in terms of a psychometric technique and a dependent measure of meaning in theoretical and applied research. A small part of that research literature is reviewed here with the intent of being brief, concise, and yet including the essential components of prior thinking as it relates to the research reported here. These essential components include a detailed discussion of the development of the technique from its objective (measuring meaning) through its final format, the relationship between the SD and other research on the language of emotional or affective reactions, and an extensive presentation on the psychometric qualities of the technique. From the general discussion of the technique, its development and nature, the presentation proceeds to a higher level of specificity with a review of applications of the semantic differential in the educational domain and particularly teacher training including review of the limited research on changes in meaning due to an intervening treatment. The juxtaposition of these latter two reviews leads to the research design and methodology reported in the concluding chapters.

The Semantic Differential Technique

Chapter I included a concise discussion of the Osgood conceptualization of the role new experience and language play in the perceptions of the world of stimuli. This section will expand that discussion and add a brief review of the progression of research that led to the semantic differential technique. From this it will be possible to move to a definition of semantic meaning space and extract rationale for subspaces in meaning and the tailoring of the technique for specific applications. The conclusion of the section will be a more detailed presentation of the format generally employed when measuring the descriptive associations made with concepts via a semantic differential.

From among the various definitions of the term "meaning," Osgood chooses one which relates meaning to a "representational mediating process" (1957, pp. 5-9). This is a psycholinguistic conceptualization which tends to become more clearly understood when its elements are considered separately. The primary vehicle of communication is language, which is composed of linguistic elements. These elements, according to Osgood, have been created by man to communicate his thinking or inner states of consciousness. Linguistic elements used in combination therefore <u>represent</u> an operating system of thought in the organism. For example, the elements on this page are symbols which represent a certain pattern of ideas or concepts. Given these elements as stimuli, the reader is employing

them as mediators in the process of integrating these thoughts with his previous experience; that is, comparing what is said here with what is already known, both of which are represented by language. This is a process of encoding and decoding linguistic elements. The process is not what Osgood would term meaning, but is the combination of linguistic elements which are encoded to characterize a thought which is the manifest meaning given to a concept or idea. Consequently, it is this linguistic encoding of the words chosen to communicate a thought which he endeavors to assess systematically on the assumption that it is possible to use this information to draw conclusions about the "states of" or "events in" the language user. In an attempt to avoid confusion about this conceptualization of meaning and the role it can play in psychological research, Osgood adds a cautious limitation:

> Although it may be trivial in one sense to insist that all discriminable events in messages must ultimately be correlated with discriminable events in language users, this must be the case if we are to avoid mysticism in our interpretation of language behavior. When a language user comes out with sequences of linguistic responses which are ordered both as to structural and semantic characteristics, we must assume that there is some ordered, selective system operating within the organism. Ultimately it is the job of the psycholinguist to make a science out of the correlations between message events and states of the organism. In our work on what we have been calling "meaning," we have mapped only a small region of this complex set of correlations, and that rather sketchily (p. 321).

The objective toward which Osgood and his associates intended to move, therefore, was to use linguistics to

measure states of nature in individuals; or more specifically, the "connotative, emotive or metaphorical meaning" as opposed to the "denotative, designative, or referential meaning" (p. 321). The progression of their research endeavor led to the semantic differential technique representing these states.

There are, of course, a number of ways to measure such emotive reactions. Osgood (1957, pp. 10-18) reviews a number of these and comes to the conclusion that a psychometric scaling procedure is more appropriate. One alternative is physiological measures of emotional reactions: action potentials in striate musculature, salivary responses, and galvanic skin responses. Other methods of measuring meaning have come from learning studies (semantic generalization, transfer, and interference research), perceptual methods, and word association research. However, these are seen as being of dubious validity and each involves cumbersome and tedious procedures. Instead, Osgood suggests that a simple though infrequently used scaling procedure is the most plausible option, because it controls for individual differences in vocabulary and verbal facility. Such a procedure can lead to valid inferences about the psychological state of the respondent.

Each step in this research sequence is discussed by Osgood, et al. (1957) in Measurement of Meaning but cannot be totally reproduced here. However, the essence of the progression is outlined below. Osgood found that, when an

individual was asked his reactions to a previously encoded (familiar) stimulus concept, he typically responded with a series of adjectives and that the variety of the adjectives employed differed between individuals with groups of society. He therefore concluded that adjectives would be the most efficient descriptors but that it was necessary to exercise more control over the measurement of the reactions. This control derived from the scaling work of Mosier (1941) which used a bipolar sequence separated by an 11-point scale designed to tap a favorable-unfavorable attitude. The limitation noted in this work was that the measure was unidimensional, while the adjectives employed previously had appeared to be multidimensional.

Consequently, Osgood and his associates employed various sets of bipolar adjective pairs and coupled them with concepts to assess the manner in which undergraduates used the adjective pairs to modify the concepts. It became apparent that there were three primary dimensions on which individuals tended to rate the concepts. These were an evaluative dimension (characterized by such bipolar pairs as good-bad), which assessed the favorable or unfavorable nature of the concept in the perceptions of the rator; a potency dimension (strong-weak), which reflected the perceived power of the concept; and an activity dimension assessing the perceived action of the stimulus concept (active-passive).

From this research two major points emerge which are relevant considerations in the research reported here. The first is the frame of reference commonly formed when individuals are asked to rate meanings of concepts; that is the evaluative, potency, and activity dimensions. These three dimensions of emotive response have been revealed in other research in the language of emotion and will be discussed in greater length in the next section. The second major point that emerges is in the nature of the instrument itself and its potential for assessing linguistic associations by providing a controlled set of conditions within which the respondent must operate to make known his reactions to concepts and ideas. The potential is realized when one considers the number of combinations of concepts and descriptors it is possible to generate in one linguistic system. In fact, research to date has only tapped small pockets of information about these meanings and associations. Osgood and his associates began with the general case assessing well-known concepts familiar to all respondents. The conclusions they have been able to draw from such data provide only the broad framework of the system used to encode ideas and reactions. This is at best very incomplete by Osgood's own admission (see quotation above).

However, in advocating continued development and application of the technique, he stipulates that new, more specific domains be probed simply by selecting content-specific concepts and exploring the interrelationships with

various descriptors. In 1961, R. G. Smith concluded:

> . . . the dimensions of any special subject matter area must be individually determined even with areas as closely related as those of general speech to the theater arts, since there are both factor and scale variations in significant amounts. This necessitates, for any special area of investigation in which the semantic differential is to be used, a specific factor analysis to determine the important factors and the scales which measure them (p. 1).

The intent of the research reported here was to move in just such a direction in the realm of education; to assess the dimensionality and to seek to quantify meanings ascribed to instructionally relevant concepts, tools, and procedures; in order to use such information to advantage in planning instruction. Both Osgood (1957) and Heise (1969) present procedures to accomplish this and these were applied and are discussed in Chapter III.

For this reason the format of the instrumentation employed was that of a SD technique tailored to assess the domain of educational concepts. More specifically, the student was given a response sheet with a concept listed at the top and a series of bipolar adjectives, each separated by a traditional 7-point scale. The middle response option was labeled <u>neutral</u> and the extremes in both directions were labeled <u>slightly</u>, <u>very</u>, and <u>extremely</u>, resulting in the following standard SD format:

BEHAVIORAL OBJECTIVES

	extremely	very	slightly	neutral	slightly	very	extremely	
good	_____:	____:	_____:	_____:	_____:	____:	_____:	bad
strong	_____:	____:	_____:	_____:	_____:	____:	_____:	weak
active	_____:	____:	_____:	_____:	_____:	____:	_____:	passive

etc.

The EPA Structure and the Language of Emotion

The relevance of the evaluative, potency, and activity dimensions to this research was defined in Chapter I in terms of the qualities a teacher should ascribe to the instructional concepts, tools, and procedures presented during teacher training. The question of the application of acquired knowledge is addressed directly by the degree to which instruction and training have been able to demonstrate that the procedures presented are good, powerful, and active contributors to an effective learning environment.

However, there is an additional facet to the relevance of this tri-dimension in assessing the qualities ascribed to specific concepts. In the previous section it was reported that Osgood and his associates found these to be the basic dimensions or frame of reference in the general language realm. In other research as far back as the works of Wundt (1905) dimensions similar to these have been consistently reported. Wundt mentions pleasantness-unpleasantness, tension-relief, and excitement-quiet as the three dimensions of emotion. Schlosberg (1954) altered the terminology but the dimensions remain essentially unchanged: pleasant-unpleasant, rejection-attraction, and activation-sleep. Nowlis and Nowlis (1956) termed the dimension hedonic tone, social orientation, and level of activation. It was at this point that Osgood (1957) began to discuss the EPA structure and this was followed by others: Block (1957): pleasantness, relevant interpersonal relatedness, and level

of activation; Davitz (1969): hedonic tone, relatedness, and activation. There have been others who have cited various components of the three dimensions as being characteristic of man's communication in the affective domain. (See Davitz, 1969, p. 132 for a comprehensive listing.)

It becomes evident from this that if man is asked to describe his affective reactions regarding a familiar stimulus he will probably describe it in all or some subset of three components: whether or not he reacts favorably, how powerful he perceives it to be, and its level of activity. This has even been shown to be true across language and cultural differences (Kumata and Schramm, 1956; Tanaka, Ogama, and Osgood, 1963; Michon, 1960; Suci, 1960). This has brought Osgood to the conclusion that there is a common factor which he terms "semantic space" which has led man to develop linguistic elements which focus on three primary areas and add additional subdimensions and Miron (1969) to term the three dimensions as "universal features of human semantic systems."

The need to define and tailor an instrument to tap these three dimensions should now be evident. Such a procedure is advisable because of the nature of the information sought in this particular study and because it is the common frame of reference in describing the types of linguistic associations which are of interest.

Psychometric Qualities of the Semantic Differential Technique

The intent of the argument throughout Chapter I and to this point in Chapter II has been to clearly demonstrate the relevance of the research carried out and the relevance of the semantic differential technique to that research. However, in addition to relevance, the SD technique of quantifying psychological variables has been shown to have the psychometric qualities demanded of scaling techniques. The qualities on which one traditionally judges the adequacy of a particular instrument have been listed as objectivity, reliability, sensitivity, validity of metric assumptions, comparability, validity, and utility (Remmers, 1963; Osgood, et al., 1957). Each of these is defined and discussed at length below to demonstrate the status of the semantic differential as a measurement tool.

Objectivity

"A method is objective to the extent that the operations of measurement and the means of arriving at conclusions can be made explicit" (Osgood, et al., 1957). The criterion of objectivity is that the instrument produces "verifiably reproducable data regardless of the rator" (Remmers, 1963, p. 330). The essence of the defense on this quality is that, given the same concepts and scales plus constant instructions, any two researchers must come up with the same response patterns. Osgood observes:

> It may be argued that the data with which we deal in semantic measurement are essentially subjective--introspections about meanings on the part of subjects--and that all we have done is to objectify expressions of these subjective states. This is entirely true, but it is not a criticism of the method. Objectivity concerns the role of the observer, not the observed. Our procedures completely eliminate the idiosyncrasies of the investigator in arriving at the final index of meaning, and this is the essence of objectivity (pp. 125-126).

Reliability

The quality of reliability is defined as the ability of an instrument to yield the same variables over repeated measures within tolerable error (Remmer, 1963). There have been numerous studies reported dealing with the reliability of SD instrumentation in the test retest sense. These are summarized in Table 1.

Coefficients of this magnitude speak directly to the issue of reliability of the results of SD applications over repeated administrations. The issue of internal consistency is, of course, one which must be addressed for each discussion with each new set of data. However, Oles and Bolvin (1971) report coefficient alpha reliabilities of individual scales on the evaluative dimension ranging from .86 to .92. Though this is frequently not dealt with in SD scaling, construction of such instrumentation requires internal consistency within dimension of meaning measured. There will be a further discussion of this is Chapter III.

Table 1.--Test-retest reliabilities in applications of the semantic differential.

Study	Interval	Variable	r
Osgood (1957)	immediate test retest	actual scale scores	.85
Tannenbaum (1953)	5 weeks	factor scores (avg. of scale ratings)	.87 to .93
Devista & Dick (1966)	1 month	mean factor scores (avg. over concepts)	.73 to .94
Norman (1959)	1 month	scale means	.96
Miron (1961)	immediate	mean factor scores	.98 or more
Rowan (1954)	immediate	factor scores	.98 or more

Sensitivity

In order to be considered sensitive, an instrument or measurement technique "should yield as fine distinctions as are typically made in communicating about the object of investigation" (Remmer, 1963, p. 330). This quality is clearly demonstrated in the semantic differential technique from two separate directions, depending on what is defined as the object of investigation.

If the object is to employ linguistic elements to make generalizations about specific operating systems in the user of those elements (i.e., the speaker), an alternative to the semantic differential would be a projective test. In this case, as mentioned in Chapter I, the control offered by the SD for idiosyncrasies in vocabulary or differences in verbal ability makes it more sensitive to (or more capable of detecting) common aspects of mental operations across individuals.

If, on the other hand, one chooses to define the object of investigation as the measurement of affect, which is generally equivalent in the literature to defining it as an attitude measure, the technique must be shown to be as sensitive as other measures of attitude. Two studies have been carried out which address this issue.

Before discussing these, however, there is an important distinction to be made which speaks to the issue of the sensitivity of the semantic differential. The measurement of attitude has been traditionally defined as the

scaling of a learned favorable or unfavorable predisposition toward an object (Allport, 1967). It must therefore be concluded that, if one is concerned with attitude as a measure of affect, he is concerned with only one of the dimensions which can be measured by the SD techniques, the evaluative dimension. The other two primary dimensions of interest here, potency and activity, are critical components of affect which expand the definition of affect and the information which the technique can provide to those concerned with non-cognitive outcomes of education. The dimensions are, therefore, claimed as added sensitivity for the semantic differential over more narrow attitudinal definitions of affect.

However, even within the limited definition, if one compares the evaluative dimension of the semantic differential with other attitude scaling procedures, the criterion for ascribing the quality of sensitivity to the semantic differential is met. That is, it has been shown to be as sensitive as the Thurstone and Guttman procedures.

Table 2 reports part of the data discussed by Osgood concerning the correlation between responses to a Thurstone procedure and the semantic differential technique for three attitude objects measured twice. Osgood reports that these six coefficients exceed .90.

In another study reported by Osgood (p. 194) the SD was compared to the Guttman procedure (N=28 on 14 items) and subjects were found to be rank ordered with respect to attitude with fairly high consistency (rho=.78, $p<.01$).

Table 2.--Validity coefficients for semantic differential scores (s) and Thurstone scale scores (t).[a]

Object	$r_{s_1t_1}$	$r_{s_2t_2}$
The Church	.74	.76
Capital Punishment	.81	.77
The Negro	.82	.81

[a]After Osgood, et al., 1957, p. 194.

Consequently, it would seem that a microscope at least as powerful as those traditionally used can be focused on affect by employing the SD technique. Sensitivity can be added to the extent that one is able to measure dimensions of affect outside the common attitude definition.

Validity of Metric Assumptions

As psychometricians set about the task of developing instruments which have potential applications as dependent variables in the realm of behavioral science research, they make certain assumptions about the variable of interest and the technique they use to measure it. It is essential that these be valid assumptions in order to adequately operationalize the variable in such a manner as to allow inferences to be made about the state of the variable. The SD requires three such assumptions. The scales are (1) assumed to equal interval scales, (2) anchored at each end by polar opposites, and (3) passing through a common origin (Messick, 1967; Heise, 1969).

Messick tested the first two assumptions by means of a comparison of the bipolar scales and scales constructed by the method of equal appearing intervals. He came to the conclusion that, considering

> . . . an approximate equality of corresponding interval lengths from scale to scale and a similar placement of origin across scales, it seems reasonable to conclude that the scale properties implied by the SD procedures have some basis other than mere assumptions (1967, p. 206).

Greatest assurance of the equality in intervals can be secured, according to Cliff (1969), Howe (1962, 1966a, 1966b) by employing adverbial quantifiers extremely, very, slightly, and neutral to define the points on the scale. These add credence to the assumption that the seven points separating the poles are equidistantly spaced (Heise, 1969).

The third metric assumption, bipolarity, has been examined by Green and Goldfried (1965) and Bentler (1969). The first study hypothesized that, if adjectives are indeed bipolar, there should be a negative correlation between them when rated individually on a scale. The test of this hypothesis was carried out and some bipolarity was indicated. A second criterion was employed to further test this point, which stipulated that if adjectives are indeed opposites, a factor analysis of their individual ratings should be equal but opposite in sign. Again some bipolarity was revealed. However, neither test led Green and Goldfried to state conclusively that the assumption is valid. Bentler (1969) found approximate bipolarity in linguistic contrasts frequently used in SD research, as did Mordkoff (1963, 1965).

It must be concluded from this that, at worst, one makes little error in most applications of the semantic differential assuming bipolarity, equal internal properties, and a constant origin (Heise, 1969).

Comparability

According to Osgood (1957), comparability speaks to the issue of the range of situations over which the measurement technique is equally valid. As mentioned in the discussion of the relation of EPA structure to the language of emotion, the technique of measuring affect reactions via a semantic meaning space appears to be equally applicable across languages and cultures outside that on which it was initially tested. In addition, Heise (1969) cites the consistency of scale interrelationships within factors. However, there has been some evidence of concept scale interactions where different concepts cause the same scales to load on different factors when comparing widely different populations. If this discrepancy can be controlled by developing an instrument with stable scale loadings, the quality of comparability with the domain of immediate interest (within which the instrument development work was done), the quality of comparability, or validity across applications within that area can be safely assumed.

Validity and Utility

Thus far, the two psychometric qualities which have not been discussed are validity and utility, and for good

reason. There is substantial evidence in support of the psychometric stature of the semantic differential procedure on four of the first five counts, but reliability, validity, and utility must be demonstrated on each new application. Reliability has already been discussed. Utility is defined as the ability to efficiently yield information relevant to contemporary and practical issues. Chapter I presented the argument that the technique should have the potential to make such a contribution, but the purpose of this research is to attempt a verification of this capability in the realm of program evaluation in teacher training. Tied very closely with this issue is the conceptualization of validity or the extent to which scores on an instrument "correlate with scores on some criterion of that which it is supposed to measure" (Osgood, 1957, p. 140). The technique must demonstrate validity to be useful when applied to current problems.

First, it is necessary to discuss the criterion to which the semantic differential must be compared to be considered valid. It will then be possible to assess the validity and generalize to the utility.

In discussing the sensitivity of the technique, correlations were reported between the semantic differential evaluative scales and other attitude scaling procedures. When corrected to allow for the unreliability of the instruments, the validity coefficients exceeded .90. In a certain sense this demonstrates a high degree of validity, but it is

a limited type of empirical validity (Brown, 1970), only examining part of the total instrument. In order to truly speak to the issue of utility, this conceptualization must be expanded to include validity in a predictive sense. Such predictive studies have been carried out in numerous areas of investigation outside the educational domain and to a lesser extent within education.

From areas outside of education, the SD procedure has been tailored (i.e., scales and concepts selected) to predict movement toward mental health employing independent ratings of clinical psychologists as the criterion (Dingman, Paulson, Eyman, and Miller, 1962; Endler, 1961). In addition, it has been shown to be predictive of the professional status of counselors (professional versus novice) in relation to their empathetic responses to clients (Greenberg, 1970); to predict racial and ethnic status of students (McNeil, 1967); and various other personality characteristics (Rentz, et al., 1968; Griggs, 1959). In these areas and others, specifically tailored instruments can be said to be valid measures of the variable tapped by the behavior criterion.

Within the domain of education, there has been less work done attempting to assess the relationship between the SD procedure and other relevant psychological constructs. Spino (1959) and Kubiniec (1970) have attempted to predict academic achievement on the basis of connotative meaning ascribed to constructs, but both studies yielded inconclusive

results. Cook (1959) was somewhat more successful in this endeavor, as he was able to demonstrate some increase in the ability to predict ACE scores by adding SD ratings on the concepts "myself as a student" and "the ideal student." There has been other research involving applications of the technique to educationally relevant problems and these are reviewed in the next section, but they do not speak to the issue of validity per se. Perhaps part of the reason for this dearth of research is that there are few clearly defined and reliably measured constructs in the area of education (beyond academic achievement). In any case, the ultimate criterion to which any instrument purporting to measure a relevant construct in teacher training must be related to teacher behavior in the classroom. It is here that the full import of the validity-utility issue comes to the fore.

For example, if a group of students could be shown to ascribe favorable qualities as well as power and a high degree of activity to the concepts, tools, and procedures presented in teacher training and another group unfavorable, weak, and passive quality, then in order to demonstrate validity and utility of the measurement the two groups should behave differently when constructing their learning environment in the classroom. The first group should immediately extensively employ the tools and procedures presented, while the others should choose alternative procedures or at least delay implementation of the learned techniques. This is the

desired and ultimate goal of the research presented here, but due to the state of knowledge in educational applications of the SD technique it is necessary to take some preliminary exploratory steps first. Among these are the development of adequate instrumentation and trial applications to more controlled problems such as its sensitivity to changes due to instruction. In this way, the validity and utility of the semantic differential procedures in the educational realm will be assessed and developed.

Applications of the Semantic Differential to Educational Training and Change

Previous research relevant to educational pursuits involving the semantic differential can be found, but investigations in the area of teacher training are not numerous, nor do they supply much information about this particular pedagogical task. In 1964, Kerlinger suggested that the techniques developed by Osgood may have potential applications in teacher training:

> A related possibility is the study of the attitudes and semantic spaces of teacher trainees. What effect does teacher training have on the educational semantic space of teacher trainees? What effect does actual teaching experience have on the semantic space of teachers? And, if a change takes place, do concomitant changes in educational attitudes take place? (p. 579)

These critical questions related to the measurement of connotative meaning have not as yet been systematically and thoroughly addressed, and must be pursued prior to the ultimate validity test mentioned above.

The work that has been carried out that might be in some way related is exemplified by the work of Hoffman (1967), who only peripherally dealt with the measurement of teacher attitudes by employing SD ratings on the way to developing some methodological and psychometric procedure in the area of measuring connotative meaning. Feshbach and Beigel (1968) attempted to relate teachers' self-perceptions and their perceptions of an ideal child in order to draw some conclusions about the role of the degree similarity between the personalities of teachers and pupils in determining the success of classroom interactions, learning, and achievement. Wittrock (1962) employed the semantic differential procedures to assess the nature of the connotative meanings of "Public School Teachers" and Public School Children" in an attempt to further explicate the nature of the educational semantic space. This was a follow-up to the study by Husek and Wittrock (1962), in the same domain. Educational meaning space was the subject of further research by Walberg in 1967. These investigators were attempting to define exactly what the nature of the meaning of teacher was and how it related to education. In a very real sense, however, these were only developmental explorations like the one being presented here.

In order to bring the role of the technique to its full potential as seen by Kerlinger, the procedures have to be brought to bear on the question of change and this has only been done twice in teacher training under severely

limited conditions. The procedures have supplied useful change data in the area of counseling and psychotherapy. Both Endler (1961) and Dingman, et al. (1967) used evaluative dimension ratings of self-concepts to demonstrate the effects of therapy in abnormal populations. Hartley (1968) assessed investigated changes in perceptions of self and others on evaluative, potency, and activity dimensions as a result of group interactions processes and was able to demonstrate significant changes. Otis and Barrett (1967) carried out research on the role of educational and vocational counseling in bringing about changes in semantic differential ratings. Hypotheses were tested in evaluative, potency, and activity perceptions with complete success in predicting evaluative changes and very little success in revealing changes on the other two dimensions. Data on the change in ratings as the result of educational treatment or instruction are not nearly as extensive, particularly in the area of teacher training.

Some work has been carried out by Hoover and Schutz (1968) and by Walberg, Metzner, Todd, and Henry (1968). These came quite close to addressing the questions advanced by Kerlinger, but only in a limited manner. Hoover and Schultz measured changes in the favorable-unfavorable evaluative reactions of students regarding concepts which "represent on an indirect level the major value dimensions tapped by the course." This means in effect that the

concepts were not specifically dealt with in the course, but were peripherally related to a particular value bias presented. This, along with the limitation of measuring only evaluative reactions, is the major inadequacy of the study. The researchers would have enhanced the external validity of the study by choosing more dimensions to tap on concepts actually presented in instruction. It is worthy of note, however, that there were statistically significant changes in attitude toward 10 of the 13 concepts rated.

Walberg, et al. (1968) succeeded in overcoming one of the limitations mentioned in the previous study by revealing changes on more than one dimension over a 14-week period of practice teaching. The concepts in which changes were sought were "myself as a teacher," and the subjects were asked to rate how their students and their peers (teachers) would rate them. The purpose of the research was to measure differential effects of practice teaching in lower class ghetto schools versus affluent suburban schools on the self-concepts of teachers. Differential changes were reported as a result of the experiences, thus supporting the role of the semantic differential in this type of research. The major limitation from the point of view of the research being reported here was that the changes dealt only in self-concept. There are other concepts which might have been added, such as myself as a person, child, authority, education in general. These are all concepts addressed

systematically in this type of practice teaching experience and would add to the generalizability of the results.

In sum, the research applications of the SD technique which sought to map changes due to a teacher training experience have serious limitations when taken separately. One (Hoover and Schultz) measures numerous peripheral concepts on a single dimension, while the other (Walberg, et al.) rates one concept on numerous dimensions. The research described here combines these two research designs to measure numerous directly relevant concepts (including self-concept) on more than the evaluative dimension.

Conclusion

A small portion of the research literature on the SD and its applications has been reviewed here with the intent of demonstrating the relationship between the technique and similar measurement methods in the domain of emotion, the adequacy of the technique in terms of psychometrics and the past applications of the technique to teacher training experiences and the measurement of change therein. By coupling prior thinking on applications of the technique with the rationale developed in Chapter I, it has become apparent that one may systematically address the questions posed by Kerlinger, and an attempt has been made here to do so. This attempt revolved around assessing changes in semantic space and the meanings of concepts as the result of an educational experience where the concepts are systematically

pre

eva

in

was

dif

ins

tio

eva

presented during instruction and the bipolar scales tap the evaluative dimension as well as activity and potency factors in the students' reactions to these concepts. The research was carried out in the hope that the role of the semantic differential procedures in revealing the influence of instruction could be demonstrated, thus defining an additional dependent variable for the systematic and scholarly evaluation of educational programs.

CHAPTER III

METHODOLOGY

Introduction

At the conclusion of Chapter I, a series of questions was posed which, if answered, would supply some insight into the role the semantic differential would play in evaluative procedures and into the role of instruction in teacher training. Some of these questions spoke directly to the issue of change in semantic meaning space raised by Kerlinger in Chapter II:

1. Are the primary dimensions which characterize the frame of reference with respect to educationally relevant concepts dealt with in teacher training identifiable as evaluative, potency, and activity factors?
2. What is the nature of the effect of teacher training on these dimensions? Do they remain constant from the beginning of instruction to the end?

Other questions posed in Chapter I were intended to probe more deeply into the effect of a teacher training experience, going beyond the level of change in factor structure to change in meanings of concepts and the role that instruction plays at that level:

3. Is there evidence that instruction influences the connotative meaning of educationally relevant concepts dealt with during a training experience?
4. If there are such changes is there evidence that different instructors play differential roles in bringing them about?

Still other questions spoke directly to the role of the technique and its potential in program evaluation in the teacher training context. These cannot be addressed statistically, but are nevertheless crucial to this investigation of a new methodology:

5. Are the connotative meanings of specific concepts and the changes in those meanings appropriate, given the objectives of instruction?
6. Are there other obviously relevant examples of hypotheses which might be addressed via this type of scaling procedure?

These questions have been restated here because they form the framework for the methodological procedures to be described here by generating research and statistical hypotheses as well as tests for these hypotheses. By way of orientation, the chapter will begin with a description of an instrument development pilot test with its purpose, sample, procedures, and results. The discussion will then turn to the application of the SD in order to seek answers to the questions above, describing the sample, instrumentation, design, (including its limitations), procedures, and data reduction.

Finally, the specific research and statistical hypotheses will be stated with their respective tests.

Instrument Development Pilot Test

Purpose

There were two major purposes for administering a pilot test of the semantic differential prior to seeking information about changes in semantic meaning. The first was to assess the feasibility of gathering and analyzing such data on a large scale. This involved such subproblems as determining a means of communicating the interest of the research without cueing responses, development of instructions for administration of the instrument to large groups, development of data processing capabilities and the capabilities to analyze the response of a large number of students.

The second purpose for the pilot was to take some preliminary steps toward the development of instrumentation for measuring the dimensions of connotative meaning desired. It is necessary to mention at this point, however, that this type of development cannot be completed in one step. But the first step was deemed necessary prior to a full-scale administration, primarily because of the desire to have scales which measure three specific dimensions (evaluative, potency, and activity).

Pilot Test Sample

For these reasons, a group of 120 undergraduate education majors at Michigan State University responded to

a pilot semantic differential instrument (described below). The subjects were enrolled in Education 200 (which was described earlier as the first course in the teacher training sequence) during the Winter Term of 1972. Pilot instruments were given to a random sample of 200 of the 600 enrollees. Given the exploratory nature of this pilot, a relatively low response rate of 60 per cent did yield useful data.

Procedures

The instrument to which these students responded was made up of 17 concepts chosen from three content domains within the course and these were each rated on 25 sets of bipolar adjectives chosen from the research literature as representative of the evaluative potency and activity dimensions. The poles were separated by a 7-point scale with the points labeled "extremely, very, slightly, neutral, slightly, very, extremely." The three content areas represent the three facets of Education 200, with concepts being chosen from the individual study carrels which present the task demands of teaching, the small group experience which deals with the personal demands of teaching, and the large group presentation which aims at a gross emotional reaction to the field of education. The first group of concepts, therefore, was technical educational psychology concepts, the second was interpersonal concepts, and the third was general educational concepts. These are listed along with the bipolar scales in Appendix A.

These elements were assembled, one concept per page, on an optical scan answer sheet with a letter of introduction and instructions taken verbatim from Osgood (1957, pp. 82-84) to compose the test booklet (see Appendix B).

The booklets were distributed by the instructors during the last week of classes and were to be returned by the students one week later. When returned, the booklets were separated by concept. In the few cases where a concept was left unrated, that concept was eliminated. In a few cases a single scale was unrated, and a neutral response was entered so that the remainder of the ratings would be used.

Results of the Pilot Test

Regarding the first purpose, which questioned the feasibility of employing the SD scaling procedure, numerous inadequacies were noted. Personal interaction with respondents as booklets were returned revealed a generally negative reaction, due primarily to the inadequacy of the description of the purposes and relevance of the research presented in the letter. Additional contributions to the lack of enthusiasm on the part of the students came from unclear instructions and the length of time required to rate 17 concepts on 25 scales. Under these circumstances, it was difficult to maintain motivation at a sufficiently high level to insure thoroughly considered responses.

The data collection procedure involving the use of optical scanning for card punch was a very efficient means

of preparing the data for analysis. However, the nature of the response sheet employed (see Appendix B) necessitated choosing a response from 1 to 7 on the scale at the left and transposing that response to an opscan field at the right. This was a tedious process and was seen as an inadequacy in the procedure.

Therefore, the following decisions were made with respect to the alterations in procedures to be made in developing the instrument for step 2 in the research:

1. It is very important to more clearly and concisely define the purposes of the instrument to the respondent, since it may not be immediately apparent in the instrument.
2. The instructions taken verbatim from the SD research literature were totally inadequate. It was apparent that the instructions had to be rewritten to contribute to the student's understanding of the task and its purposes.
3. The instrument had to be reduced in length to maintain a sufficiently high motivation level throughout.
4. The development of a new response sheet was called for to facilitate responding by allowing the student to work a response within the scale field, i.e., between the adjective pairs.

Instrument Construction

The responses to the concepts on the 25 scales were summarized in a number of different ways to assure complete understanding of the qualitative and quantitative response patterns. This implies two complementary concerns. The first deals with the statistical nature of the data and calls for summary of central tendency and dispersion of responses as well as analysis of the interrelationships of responses. The second concern is with the summary of the semantic or linguistic relations between the concepts and scales. Both of these concerns, one empirical and one rational, were seen as essential contributors to the decisions to be made concerning the nature of the final instrument.

The quantitative analysis involved the computation of the means and standard deviations of each scale on each concept. Scales desirable for continuation to further research were those which had nonneutral mean values (Kerlinger, 1964, p. 520; Mitsos, 1961), and which showed wide variability (to allow for clear descriptions of factor structure). Two separate manipulations of these data led to a fairly clear-cut decision about which should be maintained.

First, for each concept the means and standard deviations were rank ordered and the concepts with the lowest sum of the two ranks were noted. Second, a mathematical combination was computed combining the means and standard deviation for each scale on a given concept by extracting

the

dev

wer

be

in

ag

an

se

in

ab

ca

a

s

c

w

n

s

t

the square root of the sum of the squared means and standard deviations for each scale. These combination values scales were then ranked and the highest ranks were noted. As might be expected, the same bipolar pairs of adjectives were noted in each of these two cases. There was a high degree of agreement in these rankings across concepts.

To supplement these data, a common procedure for analysis of SD data was employed. A series of factor analyses was performed on the data to determine how the scales interrelated. Such an analysis of the scales noted in step 1 above clearly revealed an evaluative dimension and gave indications of potency and activity dimensions. Again, there was a high degree of agreement among concepts.

The result of the statistical analysis was that 15 scales should be retained. The rational decision as to the concepts to be continued is presented in the discussion which follows.

The pilot test, therefore, was valuable in that it revealed the major inadequacies in the procedures and shed some light on the nature of the instrumentation required. These considerations were carried on into a full-scale application to seek answers to the change questions posed.

Pre/Post Administration

The second step in this research was taken during Spring Term 1972 at Michigan State University, when a second group of enrollees in Education 200 were asked to rate

concepts via the semantic differential procedure at the beginning and the end of the term. This section of the methodological discussion will present information concerning the sample tested, the instrument employed and the design chosen for application of the technique, and the procedures employed in data collection and data reduction.

Sample

Since some of the questions of interest refer to the influence of an educational experience on the semantic meaning of concepts taught, it was necessary to exclude any students who failed to complete that experience successfully. There were 412 enrollees in Education 200 during the term when the research was carried out, and of these 300 completed instruction successfully, receiving a passing grade. Of these 300, 252 or approximately 81 per cent completed the task of rating the concepts at the beginning and at the end of the term. These 252, then, comprised the sample which provided the information described here. The 19 per cent who completed instruction and who were not included lacked either the pretest or posttest rating and were about equally distributed between the two. Other than the always present issue of differential motivation, there is no obvious reason why the 81 per cent is a special or obviously biased group.

Instrumentation

The SD instruments to which these students responded at the beginning and at the end of instruction were identical. They were composed of 11 concepts and each was rated on 15 bipolar scales. Just as with the pilot test, the poles of each scale were separated by 7 points modified by the adverbial descriptors extremely, very, etc. The response sheets were modified to allow for response between the poles. For a sample see Appendix C.

The selection of scales was described in the pilot test section and resulted in the following bipolar scales representing the evaluative, potency, and activity dimensions:

EVALUATIVE (favorable to unfavorable)	POTENCY (weak to strong)	ACTIVITY (passive to active)
unfair-fair	worthless-valuable	relaxed-tense
bad-good	lenient-severe	passive-active
negative-positive	weak-powerful	insensitive-sensitive
unimportant-important	gentle-violent	still-moving
uninteresting-interesting		
unpleasant-pleasant		
unenjoyable-enjoyable		

However, the factor loadings in the pilot test analysis were not clear cut. Consequently, the question of identifiable dimensions representing the three listed above still existed and this led to a reassessment of the factor loadings on the pretest posttest administration. The nature of this reassessment was a test of the fit of orthogonal

dimension to the data and is described in detail in the hypotheses and results to follow.

The concepts to be rated on each of these scales were chosen to represent two of the areas tapped in the pilot.* These were concepts dealt with in the interpersonal process laboratory experience and in the individual study carrels.

IPL CONCEPTS	CARREL CONCEPTS
MYSELF AS A TEACHER	BEHAVIORAL OBJECTIVES
NONVERBAL BEHAVIOR	REINFORCEMENT
QUESTIONING AND LISTENING SKILLS	RESPONDENT LEARNING
MYSELF	SHAPING

With the exception of reinforcement, each of these was tested in the pilot. The general educational concepts were not followed up because of the general and diffuse nature of the instruction which peripherally dealt with them. To carry these through would have resulted in the same type of inadequacy seen in the Hoffman (1967) study, which rated concepts only tangentially related to instruction, leading to very tentative conclusions about the role of the instruction.

In addition to these concepts, three others were added which are completely unrelated to instruction.

NONINSTRUCTIONAL CONCEPTS

PHYSICIAN
RELIGION
MARIJUANA

---

*For definition of concept, see Appendix A.

These were included to allow for some quasi-control in the hope that, in the absence of a control group of subjects, they might assist in making some tentative judgments about the role of instruction. This is admittedly a type of control which lacks rigor, especially due to the lack of independence manifested between the ratings of instructional and noninstructional concepts. However, since a control group was not feasible for this study and since this is an exploratory type of research, these concepts were seen as a potentially useful addition. The conclusions drawn involving this control will be made carefully and with the full realization that later research must necessarily allow for tighter control.

By selecting these initial elements for SD research of this variety, it is clear that the criteria for such selection prescribed by Osgood have been met. The suggested criteria for stimulus selection (concepts) are that the concepts should be relevant to and representative of the area of research interest but with anticipated individual differences, unitary in meaning (within acceptable limits), and familiar to all subjects. The scales (bipolar adjective pairs) should be relevant to the area of investigation and representative of the desired factor structure (generally three for each factor), with an eye toward semantic stability and "they should be linear polar opposites and pass through the origin" (Osgood, 1957, pp. 77-81).

Two final comments are necessary to make the description of the instrument complete and these refer to the limitation of length of the final task and the order of concepts and scales. Since this is an initial and an exploratory attempt to assess the utility of a methodology, the same concepts were used as stimuli for all subjects. It was determined that 17 concepts or 25 scales in the pilot test reduced motivation to unacceptably low levels. The combination of these two limitations made it necessary to assess a very limited number of concepts (11), thus raising a question concerning the generalizability of these results to other concepts. The 11 chosen were not selected by means of any random procedure to be systematically representative of any particular domain. Consequently, the reader should realize that such generalizations cannot be made. This limitation can only be removed by employing a large number of such stimulus concepts in ongoing research. This type of ongoing study has been carried out in the general application of the semantic summaries as the "Semantic Atlas" reported by Snider and Osgood (1969). However, in the specific field of education, this work is only beginning and must begin with limited investigations such as the one reported here.

Finally, the order in which the concepts and scales would appear in the booklet was achieved by assigning numbers to each and using a table of random numbers to generate the order. Then the direction of each scale was determined by beginning with the scales as they appeared earlier in

this section and randomly reversing them. These procedures are in accordance with standard research techniques and with the semantic differential methodology as described by Kane (1971).

Design

The design employed for assessing the capabilities of the semantic differential for detecting relevant change is a one group pretest-posttest procedure. This design, as described by Campbell and Stanley (1963), has some inherent weaknesses, some of which pose as threats to the internal validity, given the nature of the intended research.

The first potential source of internal invalidity is the fact that events other than instruction (defined as treatment) may be plausible rival explanations of any change that takes place. There are two reasons why this is not a serious weakness here. First, systematic events that occur to all subjects in a large group are unlikely, and second, other experiences related to the concepts which are chosen specifically for relevance to education courses, in particular Education 200, are highly unlikely since that is the first course taken in the teacher training program and the only course in education taken during the term.

Another potential threat to internal validity is the possibility that the psychological process of interest (semantic space) may vary systematically with changes in time or simply due to maturation. According to Campbell and

Stanley, this requires special attention in studies carried out over an extended period. Here, once again, there are two points which reduce this threat to internal validity. The first is derived from the developmental semantic differential research of Foster (1960), Matz (1963), and DiVesta (1966). These studies report some instability in semantic space in early education, followed by increasing stability through the teens. Therefore, as Darnall (1964) has concluded, there is every reason to conclude that maturation plays a role in this type of research. However, it seems appropriate to conclude that the growth curve of this psychological characteristic, like so many others, levels off in the late teens. This has not been tested empirically beyond the middle teens, so maturation may not be totally discarded as a threat in this design. But when one combines the early developmental research with the facts that the duration of this study is just over two months and that the respondents were nearly all college sophomores and juniors, the threat is certainly minimized.

A third potential alternative explanation of change is a testing effect which stipulates that merely responding to a semantic differential would serve as a stimulus for changing the characteristics of an individual's semantic space. This assertion has not been tested empirically and must therefore stand as a possible alternative explanation for any changes which do take place. However, the use of

the group of noninstructional concepts clarified this point (discussed in Chapter V).

There are four additional threats to the internal validity of this research, only one of which may present problems. Differential pre-post measures due to instrument decay are impossible, given identical instruments being employed at both times. Statistical regression is only a problem with selection of extreme groups from the population of interest. Since the population of interest is college students majoring in education, there is no reason to suspect that this is an extreme group in terms of semantic ratings. Selection bias is controlled by measuring as many of the students of interest as possible. But this may give rise to a problem due to mortality or the loss of subjects who drop out of the course during the term. Further, with a sample of this size, loss of subjects due to absenteeism and other administrative difficulties are important considerations. Efforts have been made to minimize this threat, by including only those who completed instruction and (which requires attendance) by measuring changes in as many of these as possible (81 per cent).

In general, it must be remembered that though the cause of change is important in this research, it is the instrumentation and methodological model which are of primary interest. If the dependent variable indicates potential, it will later be applied to true experimental studies which precisely control these problems.

Procedures

The instrument described above was administered to students enrolled in Education 200 during the first week of classes of the Spring Term of 1972 and again during the last week of classes, thus allowing the intervening treatment to be carried out over a period of nine weeks. The first test booklets assembled and passed out included a cover letter explaining the intent of the research as follows:

> As you proceed through the course, you will be exposed to numerous instructional concepts, tools and procedures which will, hopefully, be of assistance to you when you become a practicing teacher. It is important, therefore, that our instruction be effective in teaching the meaning and use of these tools and concepts. In order to assess the worth of our teaching procedures, periodic measurements will be made in the form of tests to determine how much you have learned. That is, there will be an attempt made to determine whether or not you know how and when to use the instructional concepts and tools presented in the course. This is one method of determining if our instruction is as effective as it might be.
>
> However, there is another influence which the instruction has on you as a student which is as important as how much you learn, and that is how you react emotionally to the concepts and tools as they are presented to you. This effect of instruction is almost never measured in our educational endeavors, but we feel that it could serve the useful purpose of providing information as we plan instruction for the future.
>
> The survey which you are asked to respond to there is an initial attempt to measure some of these non-cognitive outcomes of our instruction by asking you to describe your reactions to the concepts. It is an initial attempt because we plan to use your responses to adjust and refine these measurement techniques in order to develop a systematic means of measuring your reactions. For these reasons, we require your assistance.

The posttest booklets included basically the same description with adjustment in verb tense required to put the course in the past and a reminder that they had responded

to a similar instrument earlier in the term. For complete copies of both letters, the reader should refer to Appendix C.

The letter was followed by a set of instructions including a further explanation of the research and describing how to use the scales to modify the concepts. These were modifications of the Osgood instructions employed in the pilot test, and appeared on the basis of student reaction to more adequately describe the task. The instruction sheet is reproduced in its entirety on the following page, rather than in the appendix, to insure complete understanding of the task on the part of the reader.

The optical scan response sheets employed in this procedure were also modified after the pilot allowing the student to make his response on a scale located between the polar adjectives. For a sample answer sheet, the reader should refer to Appendix C.

The exact data collection procedures called for the pretest booklets to be given to the students at the first class meeting by the instructors, and for the booklets to be returned within one week. Approximately 10 students returned the booklets after the deadline, and these were eliminated from the study. It was felt that these students might be too far into instruction and might therefore bias the pretest results. The posttest booklets were handed out directly to the students during the last week of instruction, and they were asked to fill them out at that time under the

INSTRUCTIONS

The procedure we have chosen to measure your reactions to the instructional concepts and tools presented in Education 200 is to have you judge each against a set of descriptive scales. Consider the terms or phrases which appear at the top of each page as concept tools and procedures which you have at your disposal in your classroom when you start to teach, and as you respond, make your judgments on the basis of the potential you feel the tools have in your learning environment. On the top of each page of this booklet, you will find listed different instructional concepts and tools to be judged, and beneath each will be a set of scales. Please take time to think about and define each of the concepts in your own mind before you respond. Then rate each concept on each of the scales.

HERE IS HOW TO USE THE SCALES:

If you feel that the concept at the top of the page is very closely related to one end of the scale, for example, if you see NON-VERBAL BEHAVIOR as extremely active or passive, respond as follows:

Active	■	2	3	4	5	6	7	Passive
				or				
Active	1	2	3	4	5	6	■	Passive

If you feel that the term at the top is quite closely related to one end of the scale, as for example in the case where you feel that BEHAVIORAL OBJECTIVES are very strong or weak tools to use in a classroom, respond accordingly:

Strong	1	■	3	4	5	6	7	Weak
				or				
Strong	1	2	3	4	5	■	7	Weak

If, however, you see the tool as only slightly related to one end of the scale, for example, if you react to SHAPING as being only slightly positive or negative, then you should respond:

Positive	1	2	3	4	■	6	7	Negative
				or				
Positive	1	2	■	4	5	6	7	Negative

Finally, if you see the term as being neither one nor the other, as in the case of the concept SHAPING as being neither active nor passive, then make a neutral response:

Active	1	2	3	■	5	6	7	Passive

NOTE: Included among the instructional tools in this booklet are three terms which are unrelated to education. This is an intended part of the procedure and these terms should be rated in the same manner as the others.

REMEMBER THESE THINGS AS YOU RESPOND:

(1) Take a few seconds to think about the terms and the scales before responding. This should take about twenty minutes to complete.
(2) Make each response a separate and independent judgment.
(3) It is your first considered impression that we want. Please don't respond carelessly, but don't labor or puzzle over any scale.

supervision of the experimentor. The alteration in procedures was made to reduce the amount of class time taken for evaluation procedures.

Analysis

Data Reduction

The data collection procedures described above yielded very thorough data with every stimulus concept rated on all or most of the scales by the 252 respondents. In those few cases where a scale was left unmarked, a neutral response was entered. Of the 82,160 data points, less than 100 were left blank. This was seen as a necessary data completion procedure due to the critical role that the variance-covariance matrices would play in the analysis of the data. Computation of this type of summary data requires complete responses.

A machine card drop from the optical scan response sheets produced 22 data cards per subject coded by concept, time of measurement, and instructor, each containing the responses for a given concept pre and post. These raw responses were summarized by computation of means and the variance-covariance matrices for each concept at testing time one and two. With the large amount of data in this more manageable form, analysis was carried out.

Hypotheses and Tests:
Factor Structure

The research hypotheses tested are derived directly from the questions posed earlier. These reflect concerns in the area of factor structure and changes therein:

Research Hypothesis #1

> A three-factor model will fit the raw responses to each of the concepts rated on the 15 scales and these will be conceptually identifiable as evaluative, potency, and activity.

Research Hypothesis #2

> The frame of reference (factor structure) which characterizes or fits the interrelationships among concepts at the beginning of the term will fit the responses for the concepts at the end of the term.

The first research hypothesis was tested in two ways, one of which was quite conservative and the other quite liberal. These two decision models both sought to identify the correct factor model (i.e., number of factors) from the point of view of a statistically parsimonious explanation of the response interrelationships.

The first was an application of the unlimited maximum likelihood factor analytic (UMFLA) procedure developed by Jöreskog (1965) and adapted for use on the CDC 6500 Computer System at Michigan State University by Schmidt and Scheifley (1970).

According to this procedure:

> A correlation matrix of order P_o by P_o is factor analyzed with a K_o common, orthogonal factors, resulting [in a] maximum likelihood solution . . . represented

> by a factor matrix (of order P_o by K_o) of factor loadings and a diagonal matrix (of order P_o by P_o) of unique variances. The factor loadings of variables represent the regression weights in the regression of the variables on the common factors (i.e., latent constructs) (Schmidt and Schiefley, 1970, p. 7).

The computational procedure calls for the program to perform a series of maximum likelihood solutions using Kaiser's (1958) varimax method and after each solution is completed it is "examined from the point of view of fit using Lawley's (1963) chi-square test" (Schmidt and Schiefley, pp. 9-10). If the fit does not meet the probability level of a false rejection of true null hypothesis that the model fits the data (which is specified by the experimentor) then another factor is extracted and the test is repeated. This continues until a model of K common orthogonal factors is found which adequately explains the interrelationships among the variables.

The primary reason for selecting this factor analytic procedure was that the traditional principal components analysis assumes that all of the variance in responses is true variance (i.e., the subjects represent a population), while the UMFLA procedure does not make such an assumption. Rather, it assumes the subjects to be representative of a larger population and extracts an estimate of unique variance prior to beginning the factor analytic process. Therefore, the resulting estimates of factor loadings represent a more precise description of scale interrelationships and have an

available statistical test for assessing the fit of the various factor models (Morrison, 1967).

The second test of the appropriate factor model was the "scree test" described in detail by Cattell (1966). Though the statistical basis of this test has not been thoroughly described, it is a convenient test which characterizes important common factors as those associated with the largest latent roots. According to Cattell, in a graphic representation of the latent roots, "the curve falls in a curvilinear fashion and then becomes absolutely straight in a 'scree of' small factor debris" (p. 206). The appropriate factor model is that characterized by the latent roots which depart from the scree line (i.e., are the larger initial latent roots). For example, in Figure 1 (page 69) the three-factor solution would be appropriate and the latent roots beyond that would be considered unimportant.

The second research hypothesis stipulates that the factor structure which best explains the responses on the pretest is the same as the structure on the posttest. In effect, it was hypothesized that the respondents used the scales in the same way or to form the same frame of reference at the beginning and at the end of instruction. This hypothesis was tested and was examined at a number of levels.

Cattell (1966) and Cattell and Baggley (1956) report numerous tests of factor structure similarity and suggest that the appropriate test depends on the research interest and the level of measurement which the researcher

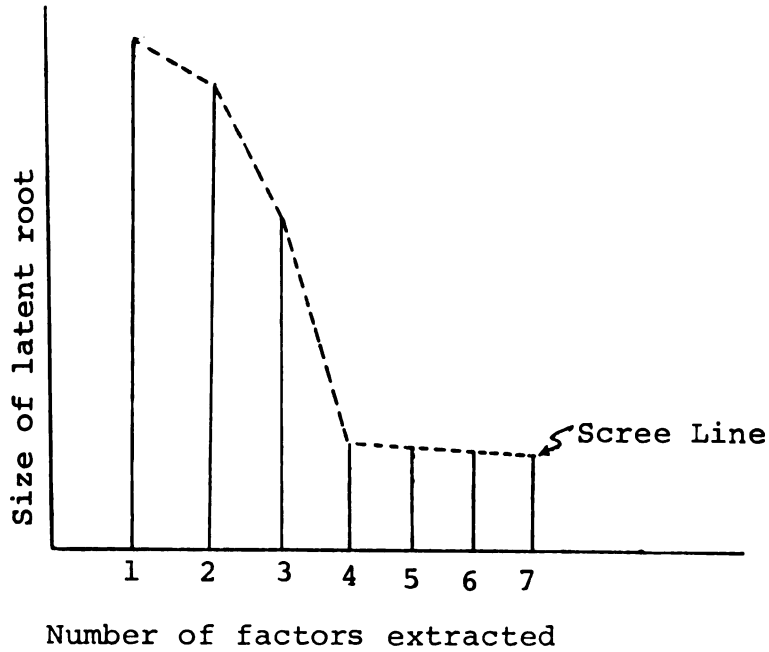

Figure 1.--Example of an application of Cattel's Scree Test.

is willing to assume. For example, the researcher may only be interested in comparing factor profiles or in levels of the factor loadings. In addition, he may be willing to assume that factor loadings are nominal, ordinal, interval, or ratio scales. Depending on the level of interest, he may choose from among nominal scale comparisons which include the Sign Test or the S test (Cattell, 1966). Assuming an ordinal scale, he may compare factor loadings via a non-parametric test such as Spearman's Rho or Kendall's Tau. If an interval scale is appropriate, a product-moment correlation will test factor similarity making maximum use of the information contained in the data, as will Burt's coefficient of congruence, assuming a ratio scale.

Each of these tests approaches the problem of comparing factor loadings in the same manner--that is, by comparing loadings from a distinguishable factor from Matrix A with a similar column of loadings from Matrix B. In the research under consideration here, for example, an identifiable factor from the pretest matrix of factor loadings was compared with a comparable factor in the posttest matrix. Each of the identifiable factors was compared with one test from each of the levels of measurement.

The specific tests carried out to assess the degree of factor invariance were:

1. Nominal Scale: Harris Test (Cattell, 1966)

Tests which have been suggested when a nominal level of measurement is assumed are: (1) the nonparametric Sign Test (Cattell, 1966), which compares factors on the basis of the degree of agreement between signs of the factor loadings using an exact probability distribution; (2) the S test (Cattell, 1966), which dichotomizes the factor loadings as salient (relatively high loadings) or hyperplane (low) and compares the same variable from two factor matrices via a 2x2 contingency table and a Chi Square test of independence; and (3) the Harris test, simply an improvement on the S test, which uses more information by categorizing loadings as positive salient, hyperplane, or negative salient. The same test of independence is applied.

The latter of the three was chosen to test the factor invariance in the present research because of its maximum use of nominal scale data. The critical value for the determination of factor loading salience applied to the data was that value suggested by Nunnally (1967): ±.50.

2. Ordinal Scale: Spearman Rho (Seigle, 1956)

This nonparametric test assesses the degree of relationship between factors by comparing ranks assigned to the loadings for the comparable factors from the matrices of interest. Again, the test was of no relationship between ranks. Where N is very small (less than 10) the sampling distribution of ρ is appropriate. In the case where N=11-30

(the present data compared 15 scales) the sampling distribution of t with n-2 degrees of freedom can be applied.

3. Interval Scale: Produce-Moment Correlation Coefficient

This is simply the sample correlation computed for pairs of comparable factors. These sample statistics can be tested (H_o: no relationship) via the sampling distribution of r.

4. Ratio Scale: Coefficient of Congruence (Cattell, 1966)

The primary reason for the computation of this index of factor invariance as stated by Cattell is that it counters the major weakness of the product-moment correlation which is that it "takes no account of difference of levels of the two patterns" (p. 196). Burt's coefficient of congruence, r_c, is computed as follows:

$$r_c = \frac{\Sigma XY}{\Sigma X^2 \Sigma Y^2}$$

Where X and Y are loadings of the same scales. Since these values are no deviations of loading as in the product-moment correlation, a high degree of congruence is only indicated when level and pattern of factor loadings are very similar.

Other tests which attack the problem from the same direction are Tucker's (1958) "coefficient of congruence" and the Wrigly-Neuhaus "degree of factorial similarity" (1955). The computation of these indices is virtually identical to the Burt Coefficient.

The major weakness of this measure is that to date there is no test of significance available, because the metric assumptions have not been explored in this domain (Cattell, 1966, p. 196).

5. Analysis of Covariance Structure: (Schmidt and Schiefley, 1971)

One additional test of factor stability was carried out which attacked the problem from a different direction. Rather than comparing each pair of comparable factor loadings separately, this test compared the entire matrix of factor loadings on the pretest with the variance-covariance matrix of the posttest responses.

This was the most rigorous of the tests of factor similarity employed to compare the pre with the posttest interrelationships. This procedure attempts, through an interactive process to find a fit between a full rotated maximum likelihood factor matrix from time 1 and a variance-covariance matrix derived from responses by the same subjects to the same scales at time 2. Having found the point of closest approximation, a Chi Square test of goodness of fit is carried out.

The reason for five tests of factor stability over time is best explained by Cattell:

> No completely satisfactory test of the goodness of fit . . . has been developed, but the available procedures for matching factors and testing for significance of the factor loadings may be applied (1966, p. 339).

In order to thoroughly test the hypothesis of factor stability, one test from each of the progressively more stringent measurement levels and the analysis of covariance structure were carried out. By executing this progression, it becomes possible to make a decision about stability on the basis of a full examination of the scale interrelationships.

Dependent Measures as a Composite of Individual Scales

The factor structure which best represented the data from the rational (psychological) and statistical points of view is thoroughly discussed as part of the results presented in Chapter IV. Briefly, however, the most parsimonious solution was one of four factors and this was found to be invariant (i.e., stable) over time.* Given that this set of latent variables characterized the student's frame of reference both before and after instruction, it was possible to proceed to the second phase of the research, that of working at changes in the response to the concepts within this set of variables.

Before discussing the analysis of the data with respect to the questions and research hypotheses about meaning change, however, there are two critical preliminary points to be clarified. Both deal with the dependent measure

*It will be necessary to give some indications about the results to facilitate discussion in this chapter and to allow for clear presentation of the complete results in Chapter IV.

employed in the following analysis. The first deals with the operational or statistical definition of "meaning," and the second is the manner in which the data were prepared to reflect this definition.

The operational definition of meaning was derived directly from the bipolar scales and from the formulation of meaning as originally put forth by Osgood:

> The point in space which serves us as an operational definition of meaning has two essential properties--direction from the origin and distance from the origin. We may identify the properties with the quality and intensity of meaning respectively. The direction from the origin depends on the alternative polar terms selected, and the distance depends on the extremeness of the scale position checked (1957, p. 26).

A simple example will make Osgood's operational definition of meaning perfectly clear. In this example, assume that one subject has been asked to rate one concept (i.e., atomic bomb) on the complete uncorrelated (orthogonal) bipolar scales and that S responded as follows:

				neutral				
bad:	X	:	:	:	:	:	:	good
weak:		:	:	:	:	:	X	powerful
passive:		:	:	:	:	:	X	active

By placing these responses on three perpendicular axes (orthogonal axes) in three-dimensional Euclidean space it is possible to graphically represent the meaning of this concept for S (see Figure 2) in the subject's semantic meaning space. In this case the quality and intensity of each reaction is projected as a vector in three-dimensional

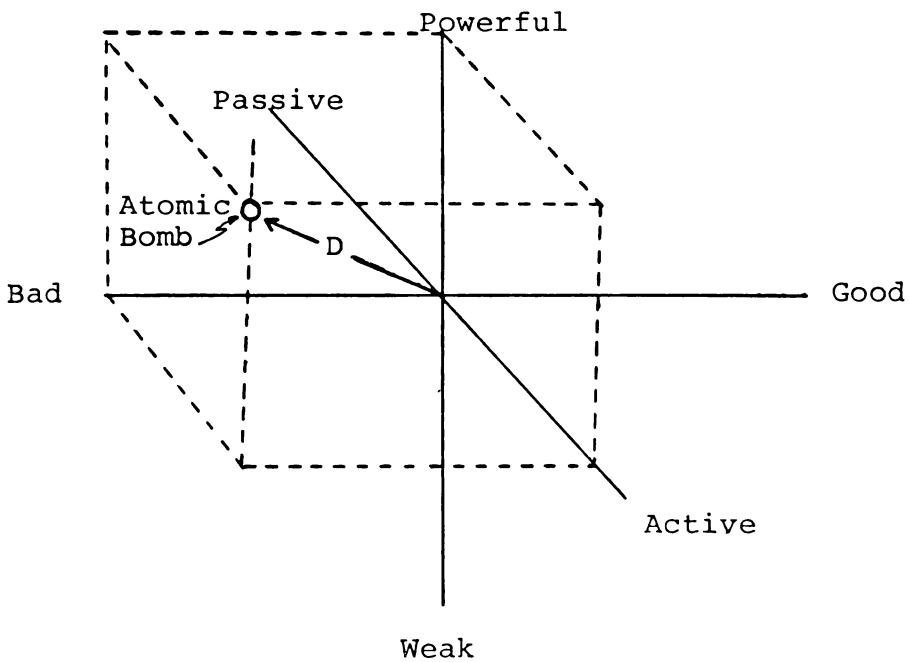

Figure 2.--Example of a concept in semantic meaning space.

space extending from a common origin in a direction selected by S and reflecting the distance specified by S. By assigning quantitative descriptors to the direction and distance options available to S, it becomes possible to mathematically represent and manipulate the responses.

Consequently, a change in the meaning of a concept, which was of interest in this research, might have been reflected as a change in either the quality or intensity of the response (or both). It is the nature of such a change which was the object of the analysis described below.

There are a number of different procedures available to reflect the change and examine its qualities. One is to attempt to interpret changes on individual scales of each concept. With a small number of scales and a few concepts, as in the example above, such a procedure might be most enlightening. However, with a large number of scales (i.e., 15 in this study) and numerous concepts (11), such a comparison of mean changes on each bipolar scale would be tedious at best and most difficult to interpret. Consequently, as in virtually all SD research, this was not considered a plausible analytic alternative.

A more plausible alternative which is frequently chosen by SD researchers is to attempt to find an equitable means of summarizing the individual responses into some psychologically meaningful--yet more manageable--form. This is frequently done by assessing the interrelationships of the scales, averaging over those which seem to be eliciting

the same information (high correlations), and separating those groups of items which appear to have little in common (low correlations). The analytic procedure of choice in this instance is factor analysis in which the correlation matrix is rotated to orthogonal factors maximizing the distinction of high and low intercorrelations. The equitable data reduction is then carried out by averaging over those bipolar scales which demonstrate the highest factor loadings on a given orthogonal factor. These scores then have the advantage of being more manageable, but there is a condition placed on their use. This is an acceptable procedure when, and only when, the researcher is able to make rational sense out of the high and low factor loadings; is able to explain, in psychological terms, why the scales should be averaged over, and what the averaged composite score means. Such a procedure has been extensively applied in SD research. Clear examples can be found in the work of Jakobovits and Osgood (1967), Miron (1961), Walbert, et al. (1968), Feshback and Beigel (1968), and Hoover and Schutz (1968). It is not uncommon to find researchers relying on the extensively tested EPA structure of semantic differential research to form these simple averaged composite scores on the basis of an a priori selection of scales to represent this latent structure (Hartley, 1968).

To offset the advantage of easier interpretation of these composite scores, there is a major disadvantage in the procedure, which arises from the assumption implicit in

the computational and analytic steps carried out. In discussing this assumption, it is critical to bear in mind that raw scores with the highest factor loadings on each factor are averaged. This assumes that the intercorrelations of the scales and the factors they represent are unity; that is, perfectly correlated. Further, it assumes that the intercorrelations between scales and the factors which they don't represent are zero. This is tantamount to performing a complete and detailed analysis of scale interrelationships and then discarding 75 per cent of the information gained. It is at the very least a sacrifice of precision, but beyond that it is the assumption that there is a set of latent parameters which totally and perfectly explain the response variance and that the researcher has discovered these and tapped them with his scales. It is safe to conclude that this is simply never true. The extent to which the scales are not perfectly intercorrelated with the latent factor they represent will be reflected in the internal inconsistency (unreliability) of the composite scores. In addition, the extent to which the correlation of scales with other factors is not zero will be reflected in the lack of independence of the composite scores across factors. And yet research such as that carried out by Jakobovits and Osgood (1967), Miron (1961), and Osgood (1957, p. 87) operates under such an assumption and seems to be the rule in semantic differential research rather than the exception.

The errors which might possibly be made with such a procedure can be avoided and maximum precision and interpretability can be maintained by using all of the information gained by the factor analysis in the computation of the composite scores. Rather than examining the factor loadings and changing the highest to 1 and the lowest to 0, the factor loadings as they are computed can be used as regression weights to allow each scale to make its due contribution to each factor (Thurstone, 1947). If the factor loading matrix is rotated to orthogonal factors, the standardized raw scores can be multiplied by their respective regression weights (loadings) and summed within factors to yield completely independent factor scores. In addition, it is possible that a scale might contribute valuable information to more than one factor and do so for sound psychological reasons. Such a bifactor item (Nunnally, 1967) would be represented by moderately high factor loadings on two factors. If the loadings are altered to 1 or 0, the total contribution of that item to the meaningful psychological interpretation of the data will be sacrificed.

Consequently, in order to avoid the misinterpretations which might result from assuming that a factor structure or set of latent constructs perfectly explain the data, and to avoid the equally hazardous errors in interpretation which might result from analyzing individual scales, composite scores were generated using the rotated four-factor solution factor loadings as regression weights, thus allowing

for ease of interpretation, maximum internal consistency, and total independence of factor scores.

The computational formula for such scores in the cases where traditional principle components factor analysis has been carried out is given in matrix form by Thurstone (1947, p. 68):

$$N\overset{q}{\left[\quad\right]} = N\overset{k}{\left[\quad\right]}X \qquad q\overset{k}{\left[\quad\right]}$$

Factor Score Matrix	Raw Score Matrix	Factor Structure Matrix

Where N = number of subjects,

q = number of factors, and

k = number of items (or scales).

This formula is reproduced in an alternative form by Cattell (1966):

$$Z_f = Z_v V_{fe}$$

Where Z_f = matrix of factor standard scores,

Z_v = variable standard scores, and

V_{fe} = factor estimation matrix.

This same type of composite factor score is possible in the case where maximum likelihood factor analysis has been applied to the problem of assessing scale interrelationships. The fundamental difference between the two-factor

analytic procedures must be taken into consideration, however, in the estimation of factor scores. This is reflected in the equation for estimation of factor scores for the maximum likelihood case reported by Harmen (1960):

$$Y = X S^{-1} \hat{\Lambda}$$

Where Y is the matrix of factor standard scores, X is the matrix of variable scores, and S^{-1} is the inverse of the variance-covariance matrix. The first two components, Y and X, are the same matrices as those represented in Thurstone's and Cattell's formulae. However, S^{-1} is included by Harmen to correct the score matrix in accordance with the estimate of error partialled out at the beginning of the maximum likelihood procedure. The reader will recall that classical factor analysis operates under the assumption that the data represent a population and contain all true variance. The maximum likelihood procedure, on the other hand, is an estimation procedure computed on a sample of finite size. As $N \to \infty$, the two procedures become identical and no correction for error is required in later procedure. However, in the present case, the correction was applied.

An equivalent form of the Harmen equation is presented by Morrison (1967):

$$Y = X \hat{\psi}^{-1} \hat{\Lambda} (I + \hat{\Lambda}' \hat{\psi}^{-1} \hat{\Lambda})^{-1}$$

Where X and $\hat{\Lambda}$ are the same as above, $\hat{\psi}$ is the diagonal matrix of unique variances, and I is an identity matrix.

For the present purposes, the Harmen formulation was applied simply because of its conceptual similarity to the more familiar classical analysis formula and because of the ease of computation, once the inverse of the variance-covariance matrix is computed.

The result of these computations for the present data, which included the rotated four-factor estimation matrix, was four orthogonal factor standard scores representing four conceptually clear latent constructs for each subject on each concept, pretest and posttest. Because these scores were in standard score form ($\sim N, \overline{X} = 0, \delta = 1$) and were, therefore, not easily interpreted in terms of the original seven-point scale, a grand mean was added to each score (placing the 0 mean back at its original location) and the standardization procedure which resulted in $\delta = 1$ (i.e., dividing by the raw score δ) was reversed by multiplying by the original raw score standard deviation. As a result, there were four factor scores pre and post on each concept which approximated the original metric. This is critical because the original metric contains a valuable bit of information: the location of a neutral point against which to measure the intensity and quality of the original response.

The operational definition of meaning is, therefore, the same as it was in the simple example presented at the

outset of this section, except the three bipolar scales used in that example to describe "atomic bomb" become four orthogonal factor scores which describe educational concepts. The same geometrical interpretation would apply if it were possible to picture four dimensions. The change in meaning of any concept can be easily derived from this, simply by computing the difference between comparable factor scores pre and post. This value for each factor served as the operational definition of change in meaning.

An additional meaning index which can be computed from semantic differential data which is labeled as an indicator of how meaningful a concept is (Osgood, 1957) and is a composite of the factor scores. This again can be clearly represented in terms of a geometrical interpretation. Again, since the four-dimension case can not be represented, return to the simple example of the atomic bomb rated on three scales (refer back to Figure 2). If the distance marked off on each axis represents how meaningful the concept is to the person on that scale by reflecting intensity of meaning as compared with a neutral point of no meaning for that concept on that scale, then the distance from the point in three-dimensional space where the concept is placed to the origin reflects the overall intensity of that meaning on all scales. In three space that distance is computed according to the standard distance between two points formula where one of the points is (0,0,0). The formula is:

$$D = \sqrt{a^2 + b^2 + c^2}$$

Where a, b, and c are vector lengths on each axis. This can be easily generalized to the four space of the present data by

$$D = \sqrt{a^2 + b^2 + c^2 + d^2}$$

and yields an overall index of the intensity of meaning of a concept.

This value was computed for each concept and subject both for the pretest and posttest data. A change in the intensity of the meaning of the 11 concepts was then operationally defined as the difference between pre and post.

The foregoing discussion has presented the conceptual and the computational aspects of an operational definition of meaning as Osgood originally conceptualized them, with some alterations in the mathematical basis. The resulting dependent measures were 10 scores for each subject: four factor scores and a distance measure both pre and post. With the data in this form it was possible to systematically address the questions and research hypotheses concerning change in concept meaning.

Hypotheses and Tests:
Meaning Change

Two of the questions stated in the introduction to this chapter form the basis for research hypotheses concerning the changes which take place in the meanings of instructional and noninstructional concepts while a teacher proceeds through undergraduate preparation or, in this case, the first course in the professional training sequence. Analysis of such changes is, of course, conditioned on having variables measures which are comparable both at the beginning and at the end of instruction. This implies stable factor structure. The stability of the factor structure in the data reported here will be described in detail in the next chapter. Briefly, however, there was a very high degree of stability found from pretest to posttest observations. It was possible, therefore, to proceed to the examination of changes of concept meaning within that invariant structure. This examination revolved around two research hypotheses, with the second only testable given an affirming result on the first.

Research Hypothesis #3

> There will be a statistically significant change in the meaning of instructionally relevant concepts from pretest to posttest which will not be reproduced in the noninstructional concepts.

The test implied in this hypothesis was carried out separately for each of the orthogonal factor scores and the distance measure. Each was a one factor one level design with the change in factor score (post minus pre) for

each of the 11 concepts as multiple dependent measures on the same subject. The statistical test was of the null hypothesis:

$$H_o: \quad \begin{bmatrix} \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ \mu_{11} \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

or that the vector of mean changes was not different from a vector of zeroes, was carried out with a multivariate analysis of variance. Evidence that a significant change had in fact taken place would be a multivariate F-ratio, which is improbable assuming the null hypothesis. Given a significant overall change, the contribution of the changes in individual concepts (i.e., instructional and noninstructional) to the overall is assessed primarily by means of a step down F and the individual univariate F-ratios which assess the null hypothesis:

$$H_1: \quad \mu_x = 0$$

Given a significant change in meaning on any one factor, it becomes possible to look more deeply for the influences which brought about such a change. This led to the second research hypothesis concerning change.

Research Hypothesis #4

> The instructor with whom the student was associated played a critical role in the changes in meaning which were manifested, and thus role will be demonstrated by differences in the changes for different instructors.

Once again, the results of the overall change are presented in Chapter IV. But briefly, there were significant changes in meaning for instructional, but not non-instructional concepts in three of the four factors and the distance measure. Therefore, an analysis of instructor influence was indicated. This was carried out by means of a one way fixed effects multivariate analysis of variance with instructor as the independent variable and three dependent variables per subject. The three were the average changes in (1) IPL concepts, (2) carrel concepts, and (3) noninstructional concepts. The null hypothesis was:

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}_1 = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}_2 = \ldots\ldots = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}_{19}$$

and if it is rejected there is evidence of differential changes across instructors. Once again the step down and univariate F's provide information as to the location of the most important contributions to the differences. This analysis, given a significant main effect, can be followed by <u>post hoc</u> tests to determine more precisely the nature of the instructor influence.

Conclusion

These four research hypotheses were subjected to tests based on the designs described. They represent all but two of the questions posed earlier. The remaining

questions will remain unconsidered until the discussion of results is presented, since neither requires a statistical analysis to determine the answer. The answer to the first can be implied from the results, but must be ultimately answered by those responsible for instruction: Are the meaning changes and the final meanings assigned to concepts after instruction appropriate for education majors given the course objectives? The second unanswered question, which speaks to the issue of the psychometric generalizability of the SD technique, calls on the creative act of generating additional applications: Are there other relevant hypotheses on which this technique might be brought to bear?

Before presenting this discussion, however, the results of the tests of factor structure, factor invariance, and meaning change will be presented.

CHAPTER IV

RESULTS

Introduction

By way of orientation, this chapter will present results in the same order as the hypotheses were presented in the previous chapter. The final decision as to the appropriate factor structure was made as a result of a progression of factor analyses summarizing the data in various ways. These will be described in sequence. The decision concerning the invariance or stability of the factor structure over time was also made as a result of a series of ever more stringent statistical tests. These will be presented in sequence. The tests of hypotheses concerning meaning change will be presented via tables of summary data and MANOVA tables. These will be concise presentations followed by a concise summary of the various decisions.

Scale Interrelationships and Factor Analysis

A complete exploration of the interrelationships of the scales began with a separate maximum likelihood factor analysis of the responses to each concept both pre and post. The purpose of each separate analysis was to test the hypothesis that a three-factor solution was the best explanation of the responses. Table 3 reports the results

of these tests and they clearly demonstrate that the research hypothesis of the three-factor solution is inappropriate. However, the analysis need not stop at that point, because these analyses contain information concerning the factor model which is the most appropriate. Note in Figures 3-13 that in nearly every case the five- or six-factor model explains enough of the response variance to yield a good fit using Lawley's Chi Square Test.

Table 3.--Results of the Chi Square test of goodness of fit of the three-factor model to the variance-covariance matrix of responses to each concept for both pretest and posttest observations.[a]

Concept	Pretest	Posttest
MYSELF AS A TEACHER	$\chi^2 = 93.5$	$\chi^2 = 81.8$[b]
NONVERBAL BEHAVIOR	154.4	136.9
QUESTIONING AND LISTENING SKILLS	99.4	106.5
MYSELF	152.9	166.7
BEHAVIORAL OBJECTIVES	134.2	170.2
REINFORCEMENT	135.7	152.1
RESPONDENT LEARNING	219.2	144.5
SHAPING	236.3	139.6
PHYSICIAN	132.1	78.1[b]
RELIGION	157.7	261.5

[a]The null hypothesis that the model fits the data is rejected at $\chi^2 \leq 82.22$ where $\alpha = .05$ and df = 63.
[b]Three-factor model fits.

It was concluded at this point that if these five or six factors were clear from a rational (psychological) standpoint, they could serve as the basis for the tests of

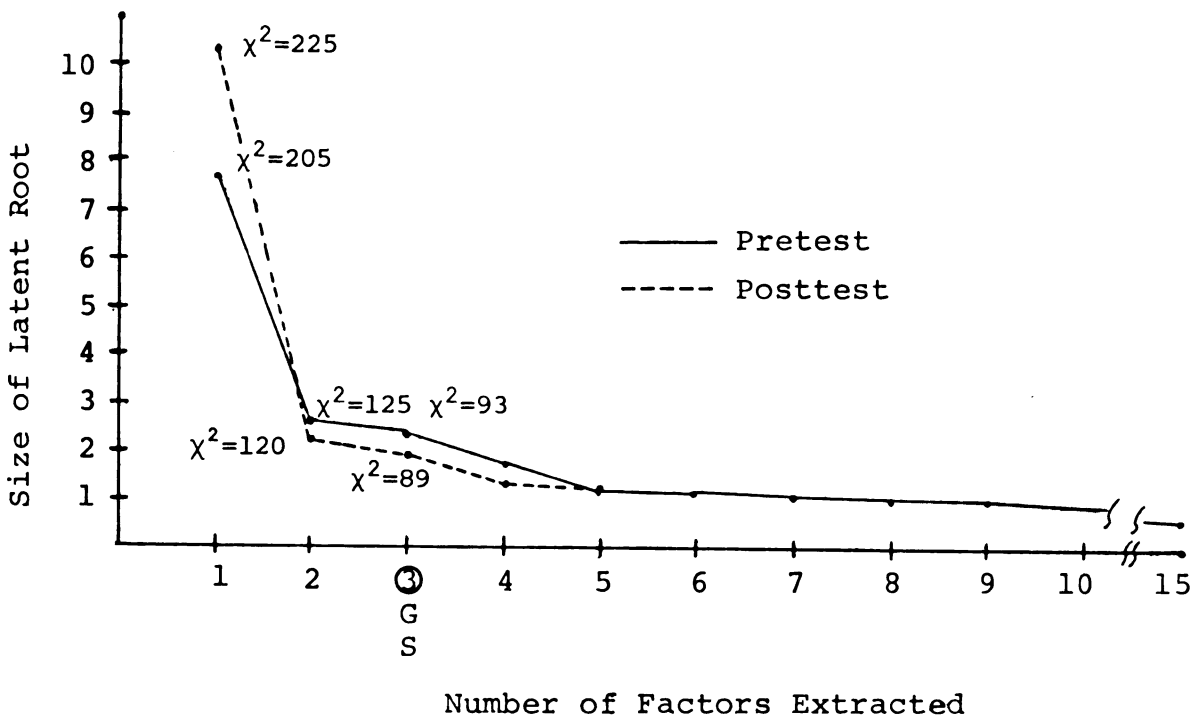

Figure 3.--Fit of the factor solutions for the concept Myself as a Teacher; size of the latent roots associated with each factor for the Scree Test-best fit designated S on the abscissa; Chi Square associated with each model for test of fit-best fit designated G on abscissa.

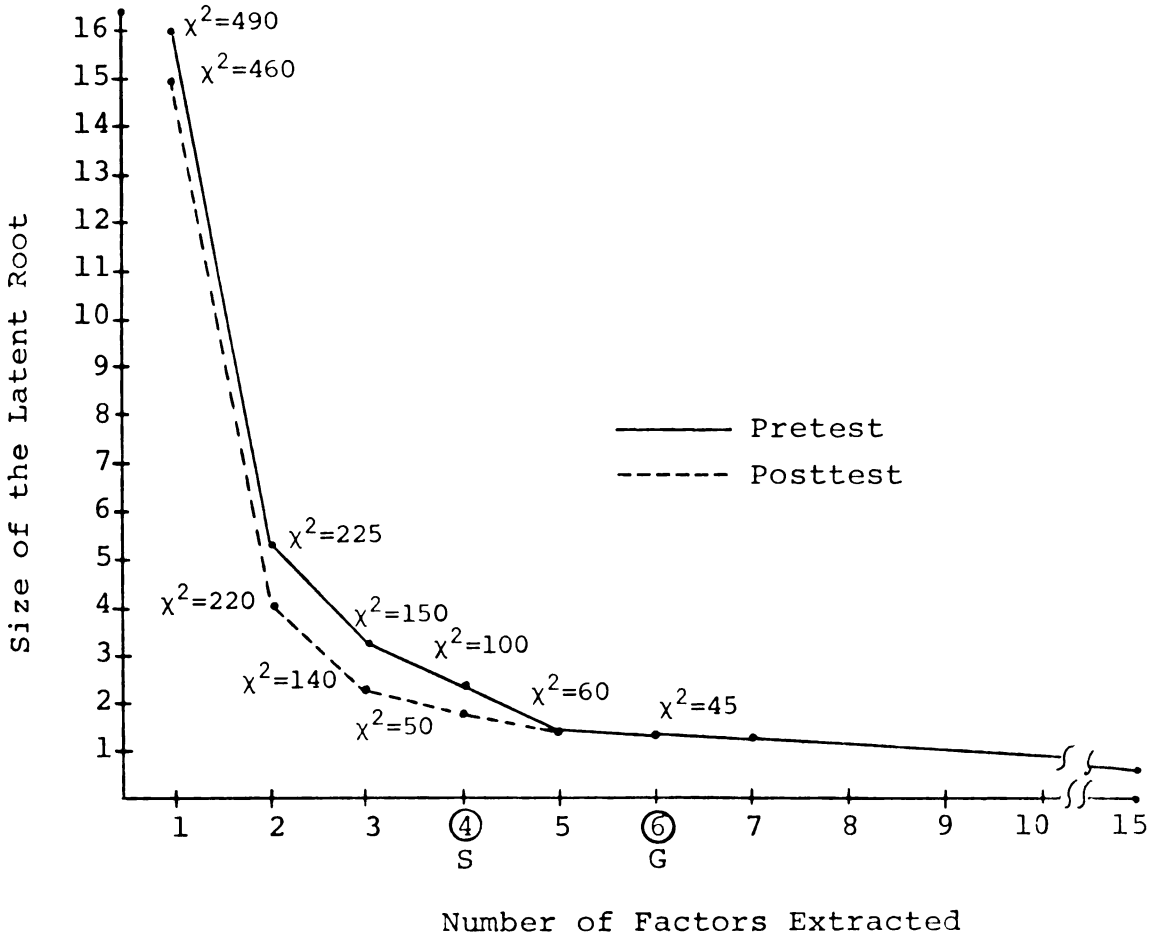

Figure 4.--Combined data assessing fit of factor models for the concept Nonverbal Behavior; size of latent roots associated with each factor for Scree Test-best fit designated S on abscissa; Chi Square associated with each model-best fit designated G on abscissa.

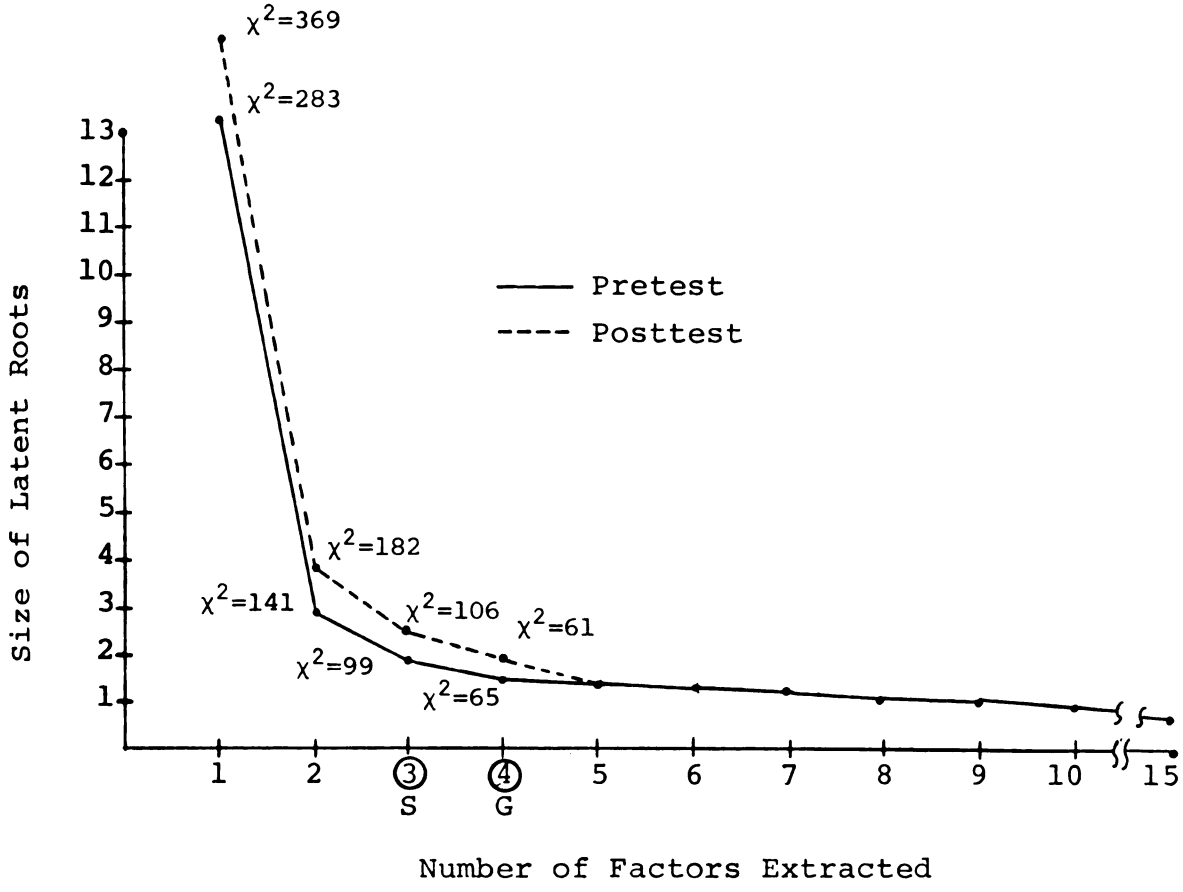

Figure 5.-- Fit of the factor solutions for the concept Myself; best fit according to Scree Test designated S; best fit by Lawley χ^2 criterion designated G on abscissa.

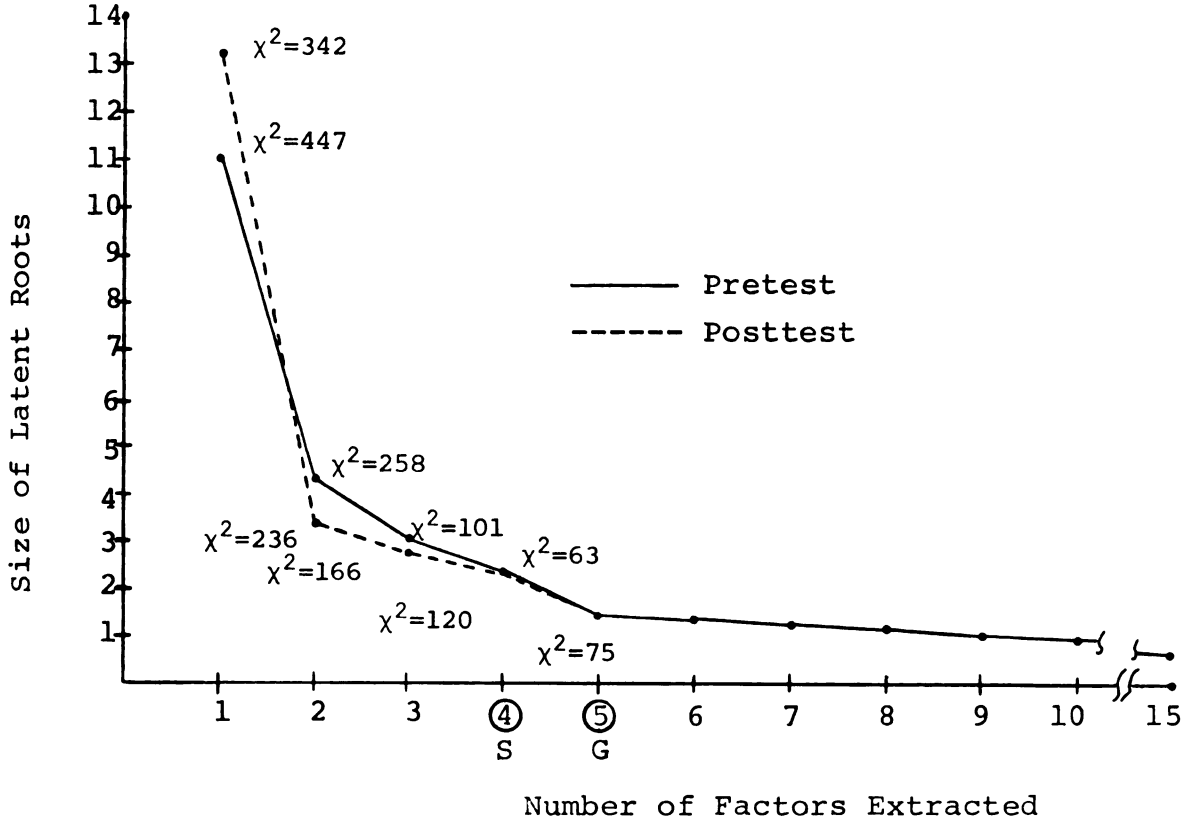

Figure 6.-- Fit of the factor solutions for the concept Questioning and Listening Skills; size of the latent roots associated with each factor for Scree Test - best fit designated S; chi square associated with each model for test of fit - best fit designated G.

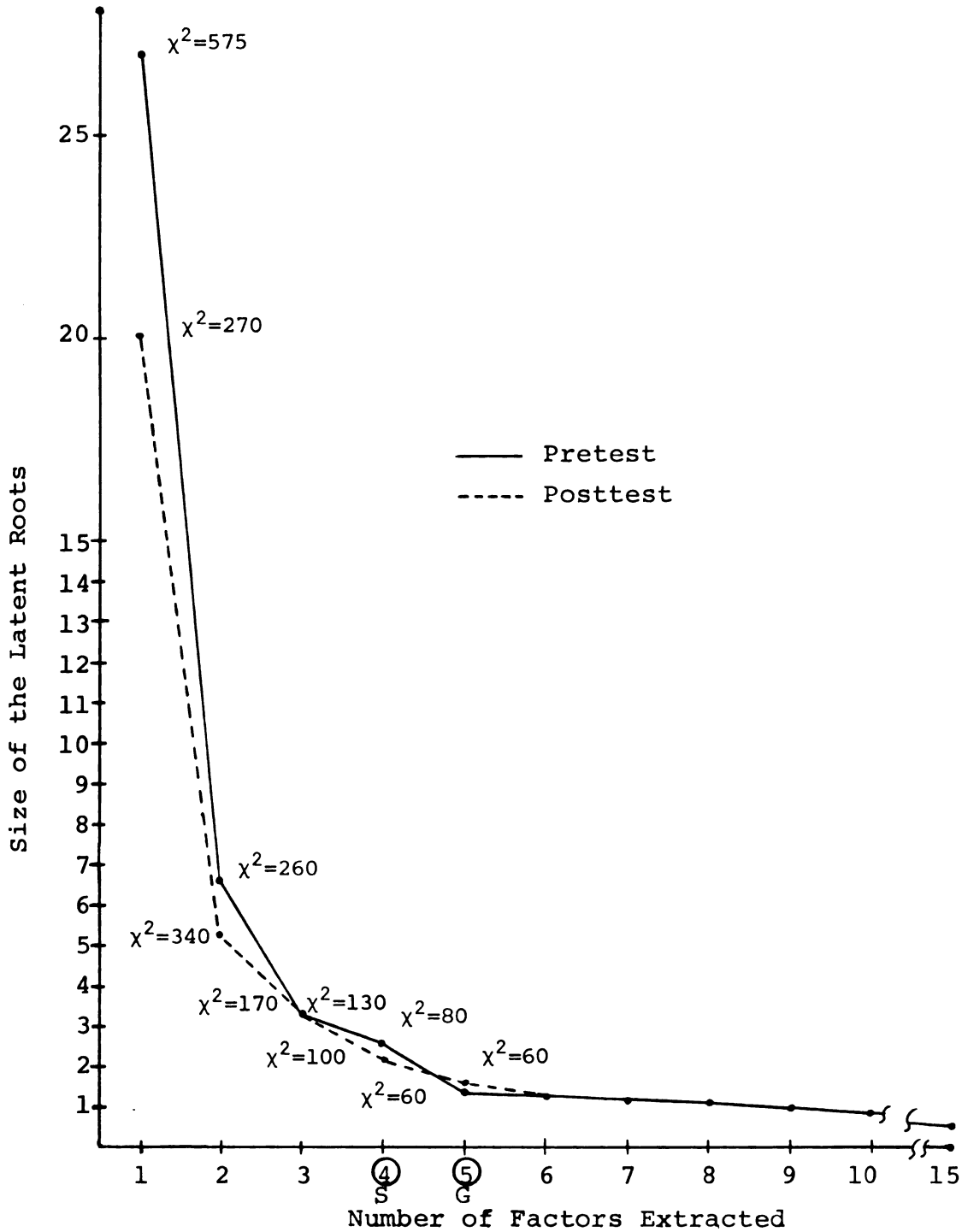

Figure 7.--Fit of factor models for the concept Behavioral Objectives; size of the latent roots associated with each factor for Scree Test-best fit designated S; chi square associated with each factor for test of fit-best fit designated G on the abscissa.

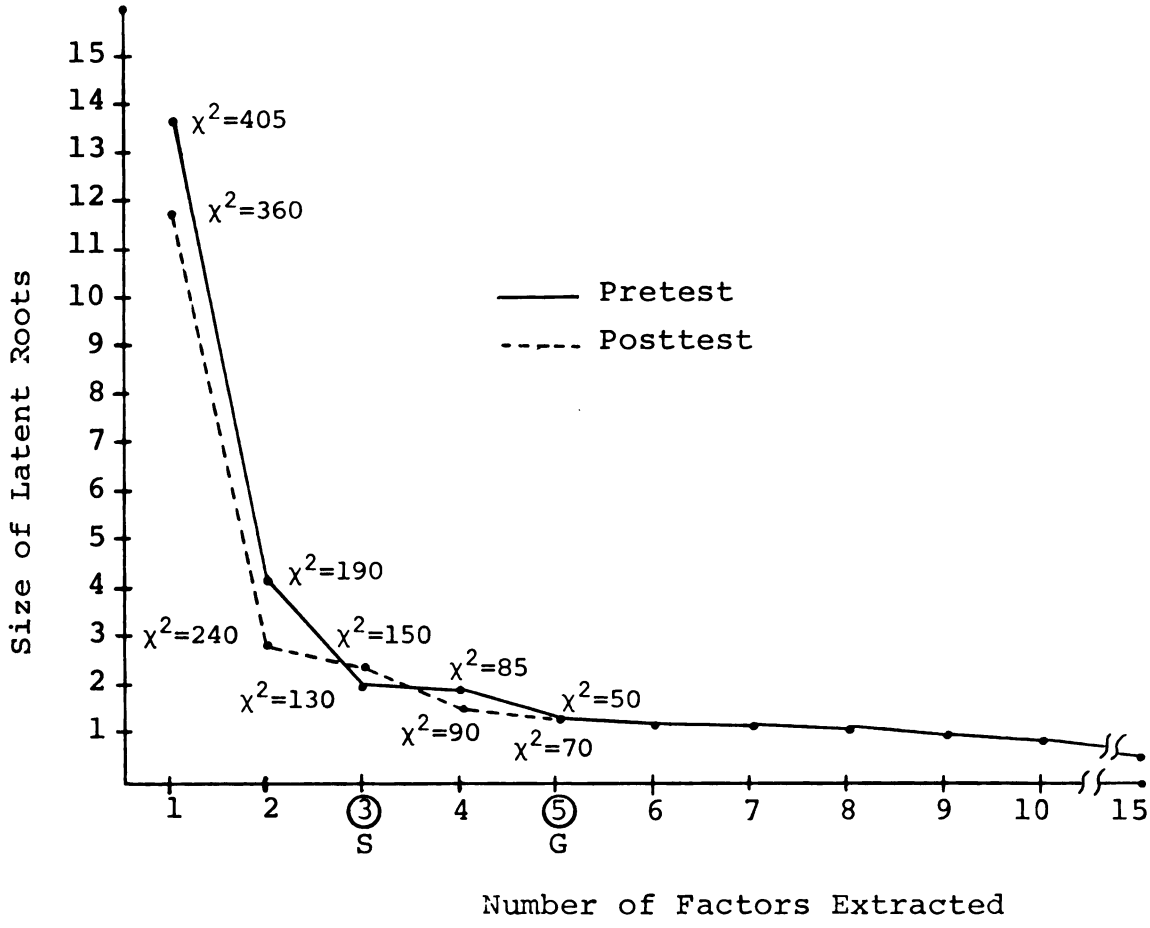

Figure 8.--Fit of factor models for the concept Reinforcement; size of the latent associated with each factor for Scree test - best fit designated S; chi square associated with each factor for test of fit - best fit designated G.

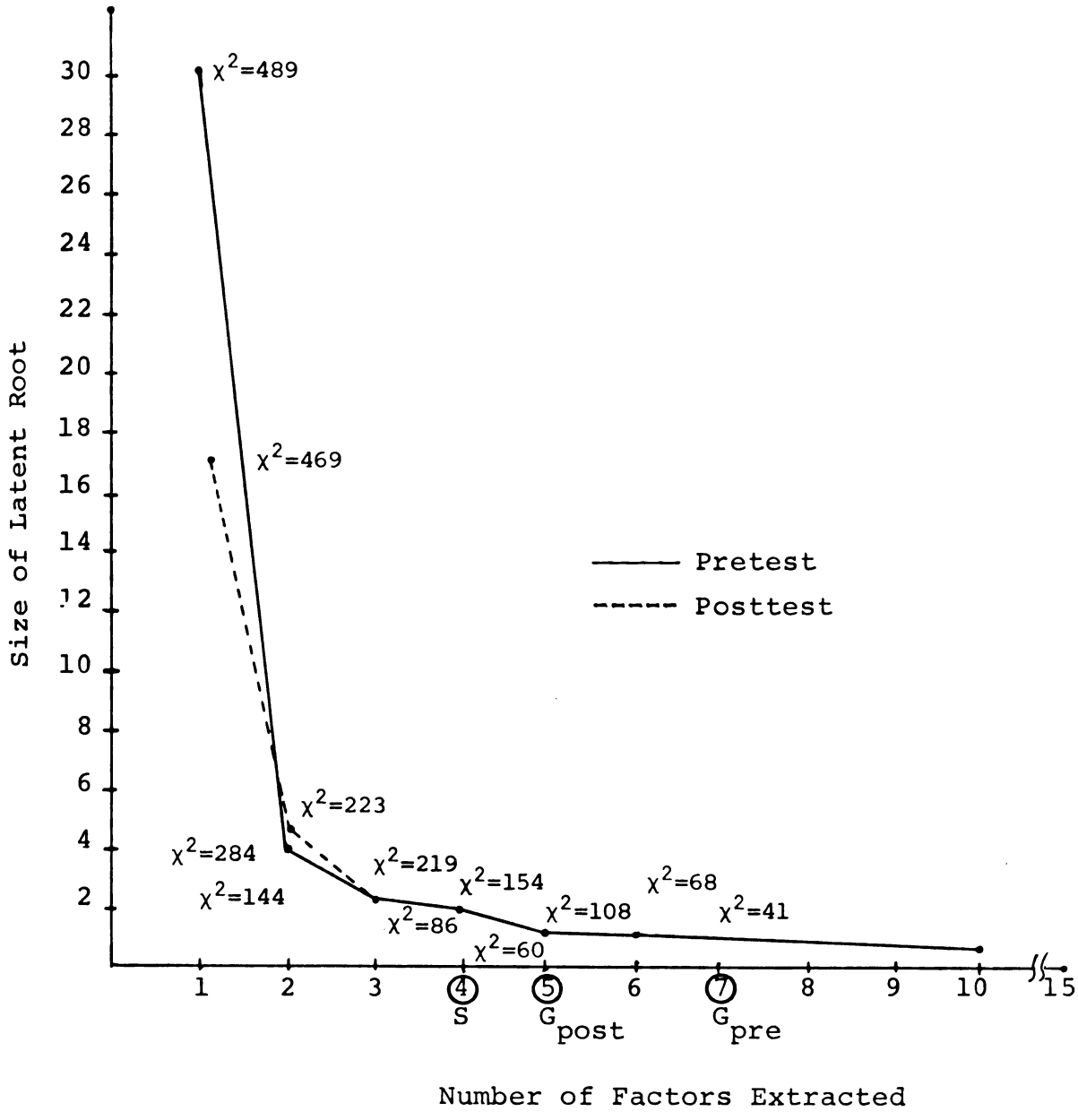

Figure 9.--Fit of factor solutions for the concept Respondent Learning; best fit via Scree Test designated S; best fit via χ^2 test designated G_{pre} for pretest results and G_{post} for posttest results.

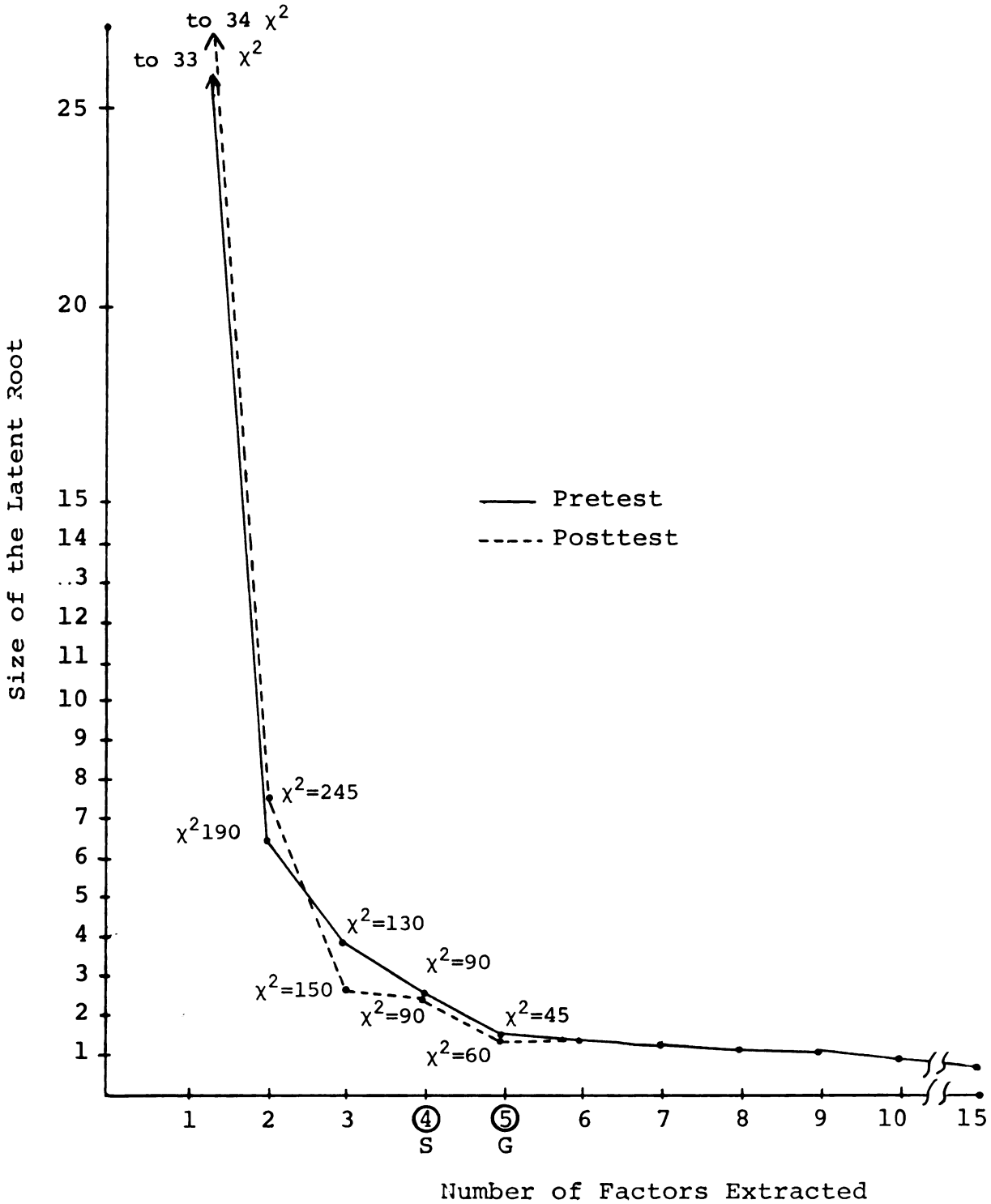

Figure 10.--Fit of factor solutions for the concept Shaping; size of the latent roots associated with each factor for the Scree test best fit designated S on the abscissa; chi square associated with each model for test of fit - best fit designated G.

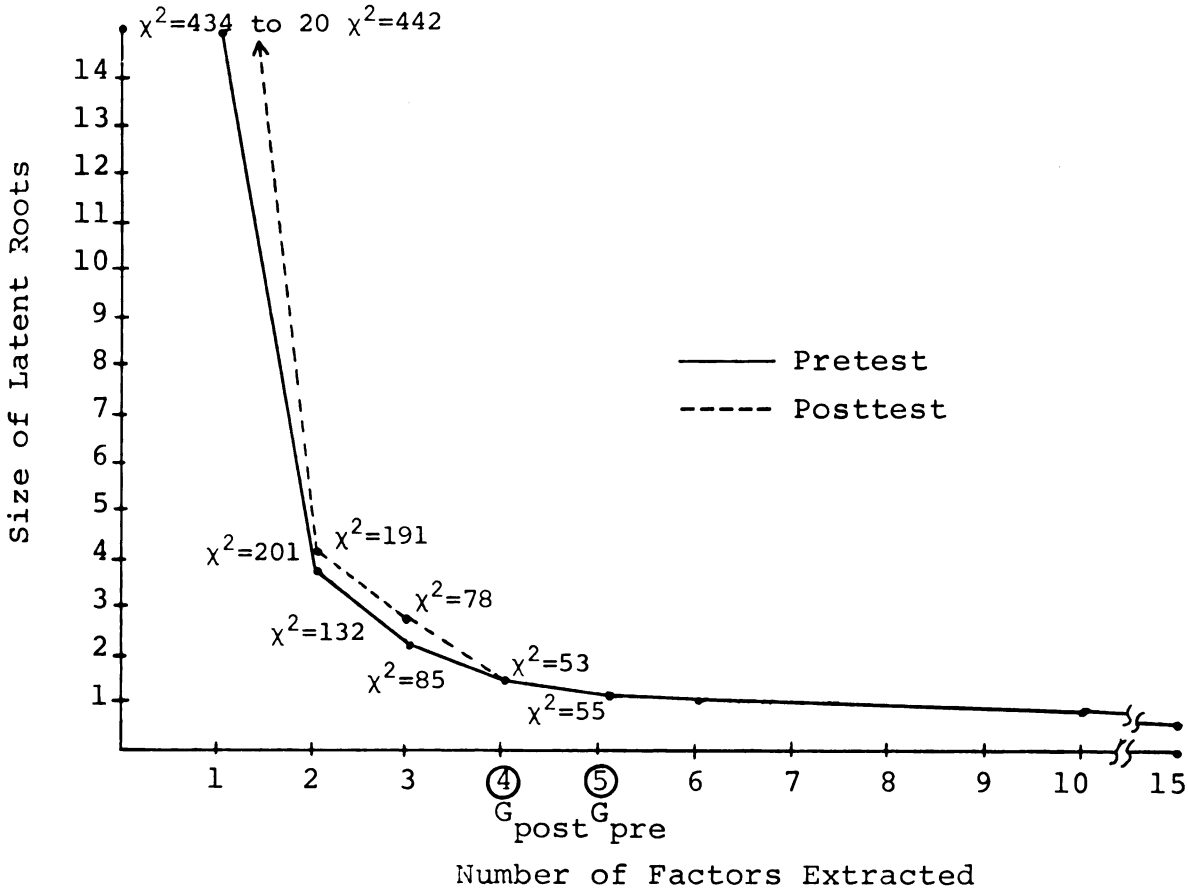

Figure 11.--Fit of factor solutions for the concept <u>Physician</u>; best fit using Scree Test and chi square test designated S, G_{pre} and G_{post} respectively on abscissa.

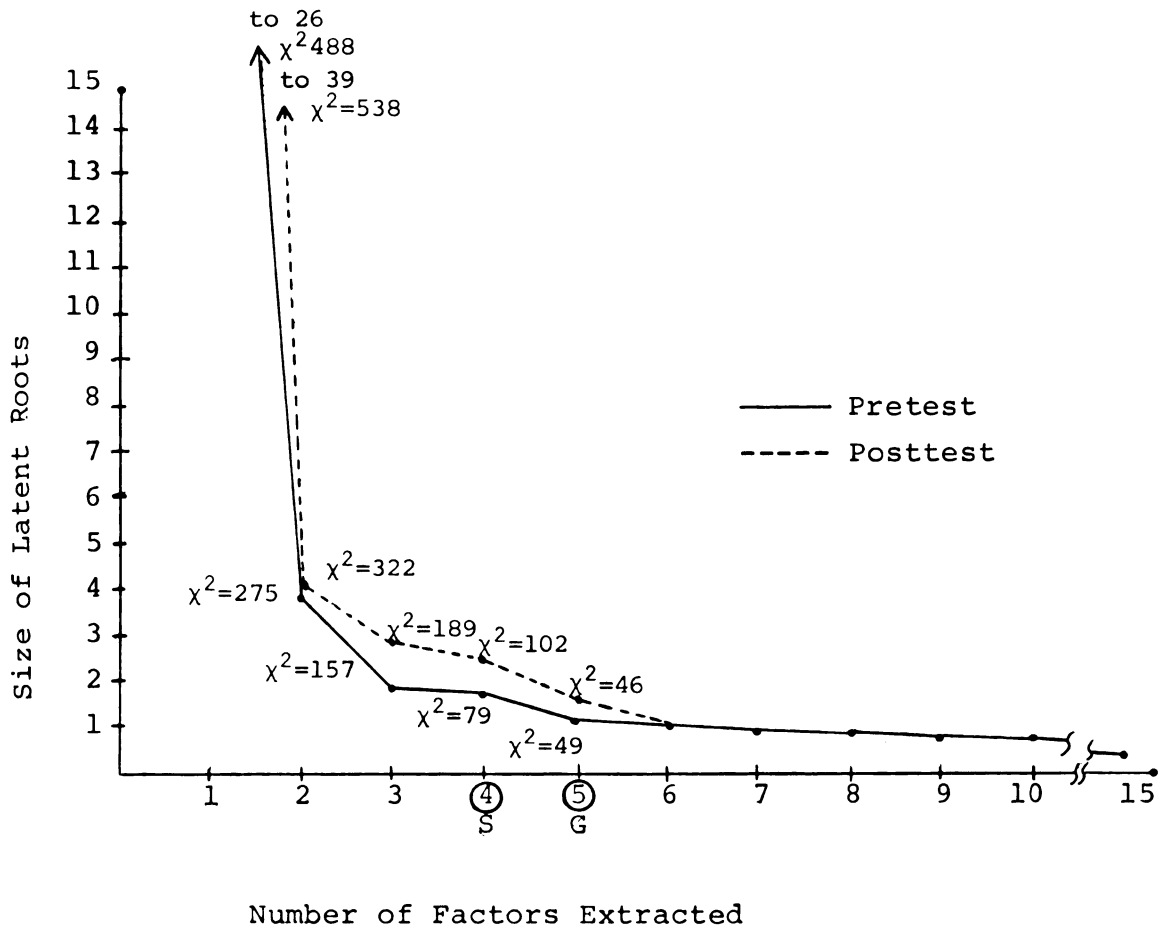

Figure 12.--Fit of factor models for the concept Religion; S and G on the abscissa indicate best fit by Scree and χ^2 test of goodness of fit respectively.

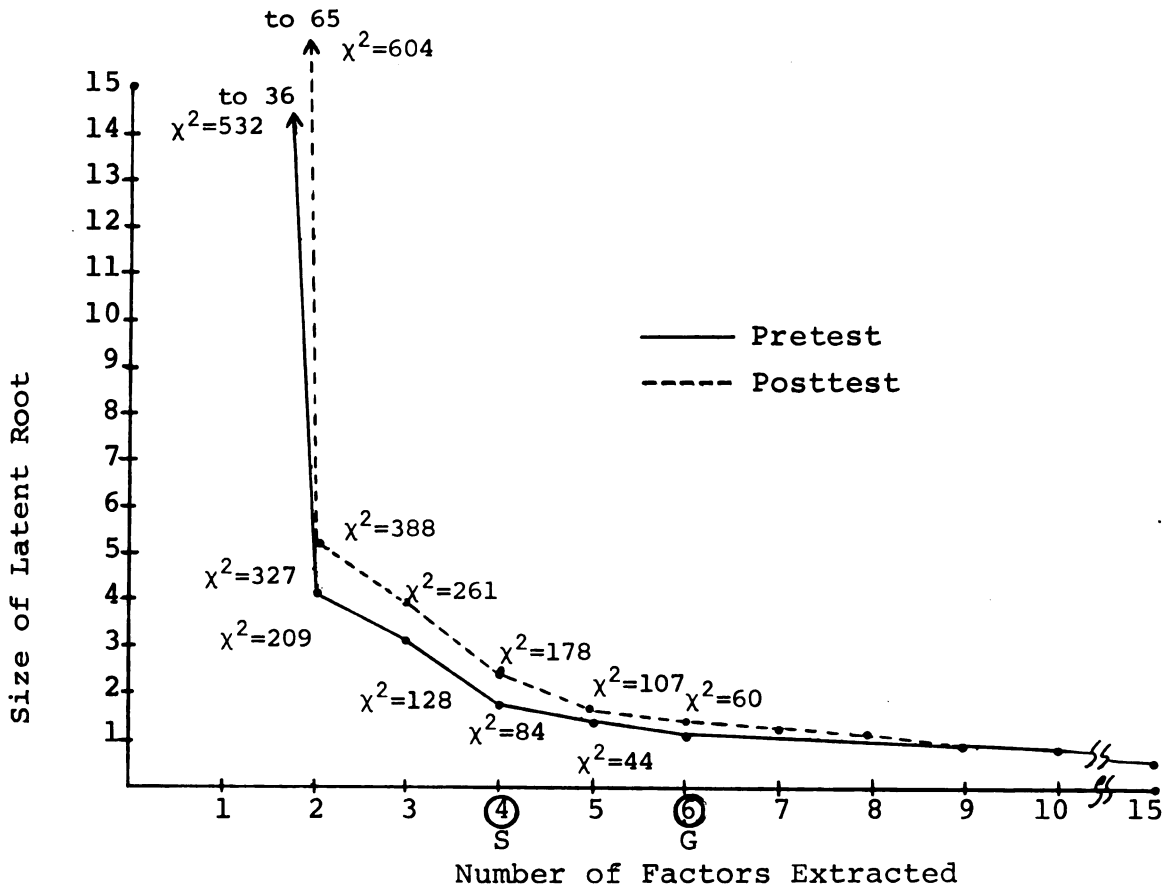

Figure 13.--Fit of factor solutions for the concept Marijuana; S and G on the abscissa indicate best fit by Scree and goodness of fit tests.

factor structure invariance. However, examination of the factor loadings of the various models for concepts revealed little that made psychological sense around the five- or six-factor levels of analysis. At about the four-factor level there were indications of what seemed to be a common thread of factor structure, but, beyond that, single scales started separating themselves out as factors. If these had been the same scales over numerous concepts, it would have been feasible to acknowledge these as inefficient yet important factors and to let them stand. However, this was not the case, since different scales seemed to spin off for different concepts and at different times. Consequently, the fifth or sixth factor and beyond were of little use in obtaining a simple and parsimonious explanation of the data.

A further exploration of the data (see Figures 3 through 13) by means of a scree test or graphing the latent roots and noting the largest of these, as suggested by Cattell (1966), demonstrated that in most cases there were four large latent roots and the remainder were unimportant. This further supported the hypothesis (initiated by looking at the factor loading) that a four-factor solution was the best explanation.

The disagreement between these two tests indicated that it would be advisable to go back and look very closely at the factor loadings to determine the exact nature of the

statistically unclear latent constructs.* On close inspection, it was apparent that, in nearly all of the four-factor solutions (16 out of 22), it was possible to identify a factor characterized by large loadings on these scales: worthless-valuable, bad-good, negative-positive, and unimportant-important. This makes immediate sense in terms of Osgood's traditional evaluative factor. It was also apparent that the scales unpleasant-pleasant and unenjoyable-enjoyable were highly intercorrelated and were being separated as a second factor of an evaluative sort but reflecting more of a personal dimension as opposed to the "other" oriented traditional Osgood evaluative dimension described above. A third group of scales which loaded highly together were severe-lenient, tense-relaxed, and violent-gentle. For clarity of discussion this was termed a leniency dimension, though it bore some similarity to a factor Wittrock (1964) and Husek and Wittrock (1963) termed a tenacity dimension of meaning. Finally, the three scales weak-powerful, passive-active, and still-moving were apparently tapping a fourth latent construct revolving around the potency of the concept being rated. The remaining three scales, as yet unaccounted for, were unfair-fair, insensitive-sensitive, and uninteresting-interesting. There were some indications that these also fit into the pattern

---

*Appendix D contains the complete factor loading matrices for each of the 11 concepts (pre and post) for the four-factor solution and the solution of best fit by Lawley's criterion.

of interrelationships described above: unfair-fair into the other oriented evaluative dimension; insensitive into the leniency dimension; and uninteresting-interesting into the personal evaluative factor. However, these were by no means conclusive indications.

The analysis to this point made two things quite apparent: (1) There was a high degree of consistency in the statistical results though they were unclear from a criterion point of view as to which factor model was most appropriate. (2) There was a relatively high degree of rational consistency and interpretability in the four-factor solution. Consequently, in search of a means of filtering out the noise that was inhibiting a clear reception of the relevant latent constructs, two additional data manipulations were carried out.

The first was to take one concept and eliminate the three scales (listed above) which had no clear place in the apparent psychological structure. The intent was not to test the possibility of totally eliminating these variables from the analysis, because this would have been an undue sacrifice of a great deal of data. Rather, the purpose was to take a closer look at the factor structure with some noise removed in the hope this would yield a clearer outline of the structure elements. This served the purpose well, as reported in Figure 14. The results of the factor analysis of the 12x12 variance-covariance matrix demonstrated that by both statistical criteria, the scree

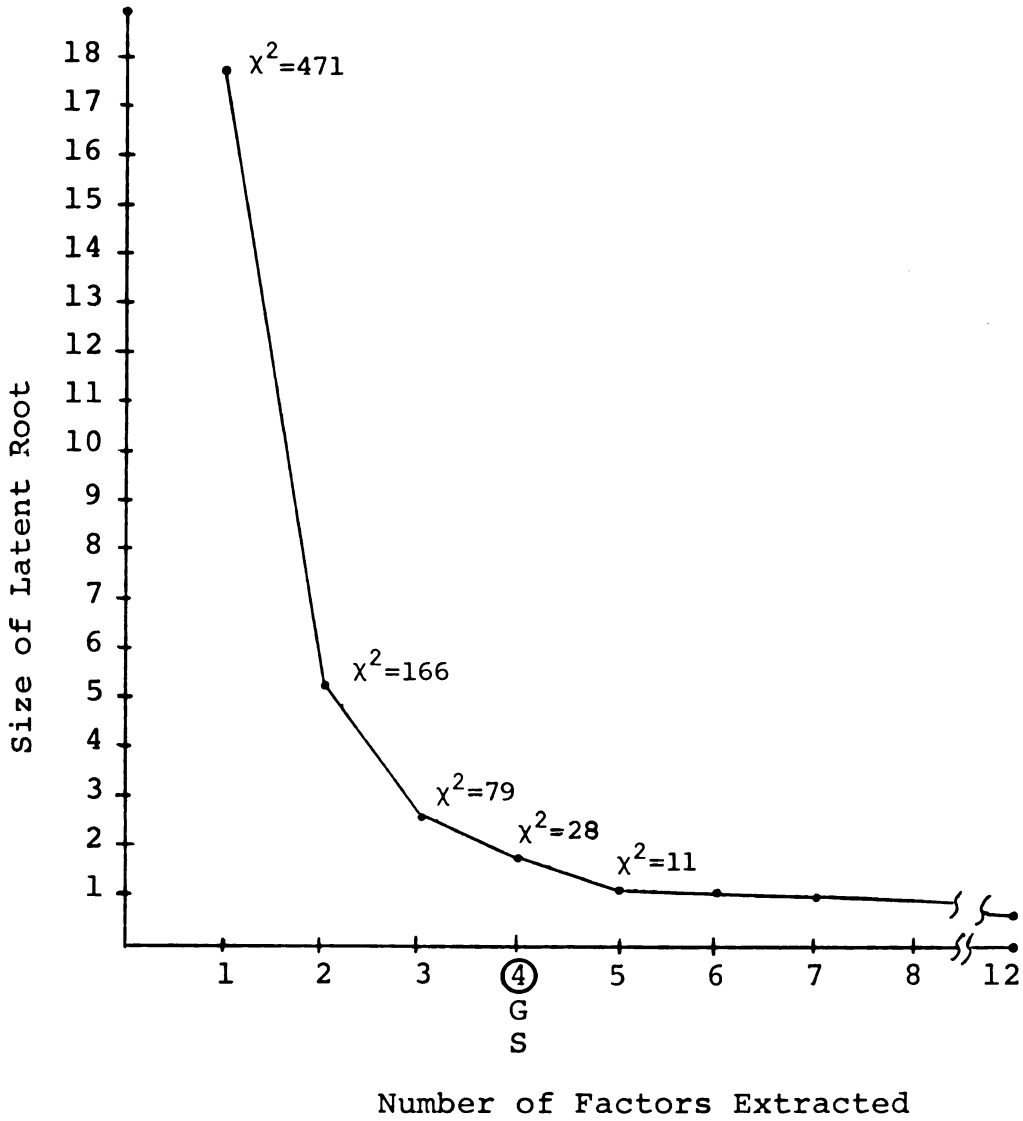

Figure 14.-- Fit of factor solutions for the concept Behavioral Objectives employing 12 of the 15 scales; size of the latent roots associated with each factor for the Scree Test - best fit designated S on the abscissa; chi square associated with each model for test of fit - best fit designated G.

test of the size of the latent roots and χ^2 test of goodness of fit, the four-factor solution was correct. Examination of the factor loadings also supported the psychological interpretation presented above.

The next manipulation performed was to pool the responses over concepts on the pretest and on the posttest and to reanalyze the pooled variance-covariance matrices. This was done for three reasons: (1) it was most difficult to deal with the idiosyncracies of 22 separate factor analyses considering 9 to 10 factor models for each; (2) in the hope that clarity would be gained through errors canceling themselves out over a larger number of responses; and (3) because of the high degree of consistency in the factor patterns across concepts. The results of these two factor analyses are reported in Figure 15, along with the factor pattern matrices in Table 4. These data provided some support for the four-factor model, some additional insights, and some disappointments. In short, they provided some good news and some bad news.

It was disappointing to find that the errors did not cancel out--they accumulated, thus increasing the discrepancy between the two statistical tests. The four-factor solution was appropriate, once again, with the scree test as the criterion. However, by summing over 2,772 rather than 252 responses, what was accumulating was not noise which would be canceled out, but true variance. Thus a larger number of factors was required to explain enough of

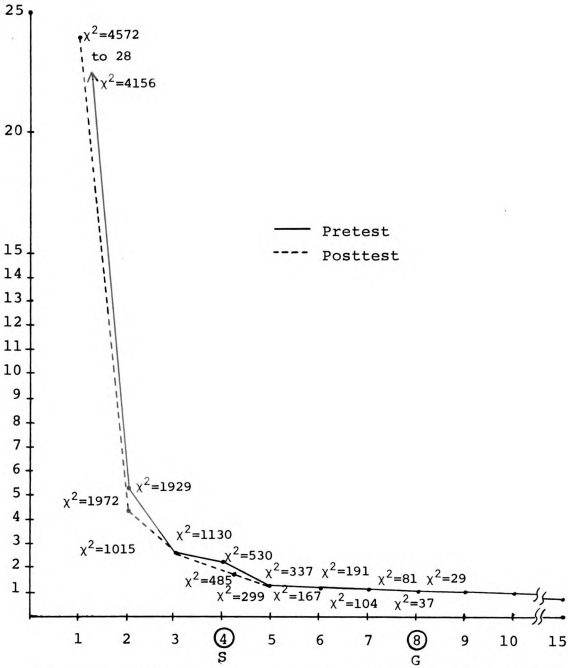

Figure 15.--Fit of factor solutions for the pretest and posttest results pooled over concepts; size of the latent roots associated with each factor for the Scree Test - best fit designated S; chi square associated with each model for test of fit - best fit designated G.

Table 4.--Rotated factor matrices for four-factor solution of pooled correlation matrices, pre and post.

Factor I.D.:	Pretest					Posttest				
Variable	Evaluative	Personal Evaluative	Potency	Leniency	h^2	Evaluative	Personal Evaluative	Potency	Leniency	h^2
Worthless-valuable	.789	-.170	.205	.268	.765	.815	-.161	.186	.231	.778
Severe-lenient	.020	-.067	-.040	.531	.289	.085	-.070	-.018	.548	.313
Unfair-fair	.416	-.178	.219	.576	.585	.506	-.199	.266	.479	.596
Good-bad	-.760	.209	-.166	-.294	.736	-.752	.236	-.169	-.310	.745
Tense-relaxed	.206	-.220	.117	.391	.357	.413	-.186	.152	.514	.342
Weak-powerful	.487	-.077	.370	.005	.380	.470	-.099	.409	-.005	.398
Negative-positive	.626	-.235	.257	.436	.703	.609	-.267	.281	.392	.675
Violent-gentle	.182	-.161	.152	.682	.548	.205	-.228	.227	.686	.652
Unimportant-important	-.733	.163	-.282	-.174	.673	-.767	.153	-.289	-.152	.718
Active-passive	-.418	.116	-.614	-.087	.573	-.370	.152	-.655	-.107	.601
Insensitive-sensitive	.300	-.152	.356	.553	.546	.347	-.241	.373	.469	.538
Uninteresting-interesting	.445	-.323	.400	.307	.556	.419	-.399	.346	.264	.523
Pleasant-unpleasant	-.320	.532	-.250	-.530	.729	-.301	.609	-.251	-.505	.779
Moving-still	.184	-.172	.657	.159	.520	.185	-.177	.714	.238	.632
Enjoyable-unenjoyable	-.291	.797	-.253	-.374	.926	-.308	.703	-.245	-.416	.823
Prop. Var.	57.7	10.9	4.5	5.4		54.9	9.8	6.0	5.0	
Cum. Var.	57.7	68.6	72.9	78.3		54.9	64.7	70.7	75.7	

the variance in responses for a statistical fit by the χ^2 criterion to be achieved. This led, of course, to a loss rather than a gain in clarity.

However (now the good news), it is interesting to note that an average of 77 per cent of the common variance is explained by the four-factor solution (see Table 4). If one proceeds beyond this point to the eight-factor solution, it becomes apparent that, in order to find an adequate fit, one must explain approximately 89 per cent* of the common variance. It can be concluded from this that the additional four factors extracted are each explaining a relatively inefficient 2 per cent of the variance, while each of the first four is explaining a minimum of twice that amount. The χ^2 test must therefore be considered a very conservative test of fit.

In addition to this, a close examination of the factor pattern matrices (Table 4) by the reader will reveal the nature of the scale interrelationships with respect to the psychological model presented earlier. Note prior to this examination, however, that the random reversal of scales which was described in the instrument development section of Chapter III has sometimes rendered scales representing

---

*This figure was arrived at from the results in Figure 15. The amount of common variance explained can be computed by summing the latent roots up to a certain factor model and dividing by the total sum of latent roots. For the four-factor model in Table 4 this proportion is .783 for the pretest and .757 for the posttest. In the eight-factor case, where a "good fit" is achieved, the proportion is .89.

the same factor opposite in direction--thus resulting in factor loadings with opposite signs. The scales are listed in this table as they appeared in the instrument.

Having carried out this extensive analysis of the interrelationships of the bipolar scales through numerous factor analyses, the decision was made to proceed on to the test of factor structure stability with the four-factor solution where the factors were interpreted to represent the following latent constructs or dimensions of meaning of the concepts to the respondent:

Meaning Dimension	Characteristic Scales	Meaning Definition
Evaluative	worthless-valuable bad-good negative-positive unimportant-important	This dimension represents the respondent's favorable or unfavorable reaction to the concept in terms of how it affects others.
Personal Evaluative	unenjoyable-enjoyable unpleasant-pleasant (uninteresting-interesting)	This dimension reflects the respondent's favorable or unfavorable reaction in terms of his own personal values or from the point of view of how it affects him.
Leniency	severe-lenient tense-relaxed violent-gentle (insensitive-sensitive) (unfair-fair)	These scales reflect the meaning of the concept from a flexibility point of view or on an open vs. closed dimension.
Potency	active-passive still-moving weak-powerful	This factor represents a reflection of the manifest or observable power and activity potential of each concept rated.

Evidence in support of this decision can be drawn from three sources. The most important of these is their consistency and intuitive appeal from a rational point of view. That is, they seem to make sense. In addition, they consistently satisfy one of the two statistical criteria for the appropriate factor model, though this was the more liberal of the two. Finally, one must consider their ability to consistently explain a large proportion of the common variance. The evidence counter to this decision comes primarily from the failure of this model to explain a sufficient amount of the variance to be considered appropriate from the point of view of the second mathematical criterion. This is a serious limitation, since this was established as the criterion of choice at the outset.

Overall, however, the weight of evidence would appear to support this decision, given that the decision is made with the full realization of the limitations of the data. The nature of the data and its limitations are not perfectly clear and could yield erroneous conclusions. The temptation to overgeneralize or to place too much store in these factors is to be avoided and cautious interpretation is called for. With this in mind, the next step in the analysis was carried out--that of assessing the stability of the factor structure over time.

Tests of Factor Structure Invariance

The results of four of the five progressively more stringent tests of factor structure stability of the dimension model are reported in Table 5. Each of these tests employed progressively more of the information contained in the factor loadings of the overall pretest and posttest factor pattern matrices reported in Table 4 to assess the degree of stability of comparable factor loadings for each factor. The nominal scale test, based only on the sign of the factor loadings and the trichotomization of the actual factor loading value, demonstrates a consistent rejection of the null hypothesis that the factor loadings are independent. The ordinal scale test, based on the ranking of factor loadings according to size or salience, also reveals total rejection of the hypothesis of no relationship between the loadings of comparable factors. Assuming that factor loadings have the quality of equal intervals among values, it is apparent that a very high degree of relationship exists between comparable factors as seen in the product-moment correlation coefficients. The ratio scale test, as defined by Burt, also shows strong relationships among factors pre and post. However, a statistical basis for this decision apparatus (based on a sampling distribution of known qualities) does not exist. But such a test would, at least in this case, be an unnecessary tour de force--given the high degree of factor stability reflected in the other tests.

Table 5.--Results of all statistical tests assessing the degree of factor invariance over time when comparing the pooled (over concepts) pretest and posttest factor solution.[a]

LEVEL OF MEASUREMENT[b] ASSUMED	NOMINAL	ORDINAL	INTERVAL	RATIO
Test[b]	Harris Test of Factor Salience	Spearman Rho	Product-Moment Correlation	Coefficient of Congruence
Sampling Distribution	χ^2	t	r	untested
Critical Value (Required to Reject Hypothesis of No Relation)	3.84 df=1, α=.05	1.716 df=13 1 tail, α=.05	.426 df=14 1 tail, α=.05	none
Factor: Evaluative	.59	.92 t=8.452	.998	.998
Personal Eval.	.000	.83 5.374	.993	1.000
Leniency	.000	.91 7.99	.887	.903
Potency	.000	.96 12.237	.891	.881

[a]Each is a comparison of loadings on the same variable from the two-factor matrices. H_o: No Relationship N = 15

[b]As summarized by Cattell (1966, p. 337); and Cattell and Baggley (1956).

The results of the fifth and final test of the invariance of factor structure, the analysis of covariance structure (ANCOVST), were not able to be completed due to the frequent failure of the computational package employed. The primary advantage of this test would have been its ability to consider the entire factor model from the pretest data and compare it with the variance-covariance matrix of the posttest data from the point of view of fit. In the ANCOVST case, however, the factor model from one set of data is compared with the criterion variance-covariance matrix from a separate set of data to assess the fit.

Even in the absence of any concrete data, however, it is possible to draw some tentative conclusions about what the results would have been. In most cases where the pretest four-factor model for any concept or for the pooled data had been compared with its posttest variance-covariance counterpart, it is safe to conclude that a lack of fit would have been demonstrated. But, just as with the factor analyses themselves, the reason for the lack of fit would have been the minor undefinable and undiscoverable additional factors which have tended to prevent a clear definition of scale interrelationships throughout this analysis. The essence of this argument is that a specified factor structure which fails to explain a sufficient amount of its own variance to be considered an adequate fit can hardly be expected to explain enough of the variance of a separate and independent set of observations to be considered an

adequate fit. In this case, the reason for the lack of fit would once again have been the high degree of precision required by the statistical test.

Therefore, based on the results of the four tests of invariance that were carried out, and tempered by the expected results of the fifth test if it had been completed, it was decided that the four-factor solution was stable or that the frame of reference within which students rated the concepts remained constant over time. This, of course, led to a comparison of comparable variables at the beginning and at the termination of instruction to determine the nature of changing meaning within this constant frame of reference.

Test of Change in Concept Meaning

The factor standard scores were computed and were transformed so as to approximate the original scale as described in the previous chapter. The results of these computations were ten scores for each subject on each concept: four orthogonal factor scores and a composite measure of meaning intensity for both pretest and posttest observations. These can then be transformed by subtracting post from pre to indicate change. All results will appear in this form, without further consideration of individual scales.

According to the first hypothesis, which assessed the change in meaning of individual concepts, the following decisions are possible for each factor: Either there is an

overall change in mean for each factor (tested separately) or there is not; if there is a significant overall change, either an individual concept contributes to this or it does not.

The results of each of these four-factor score change tests appear in Tables 6 through 10. It is apparent from these tables that there was a significant overall change in three of the four factors and that changes in individual concepts can, therefore, be considered for these three.

Table 6, representing changes on the evaluative dimension, reveals that the vector of mean differences differs significantly from a vector of zeroes, thus reflecting significant change in that dimension ($F=28.4324$, $p<.000$). The column of mean changes shows that the greatest changes took place for the most part in the carrel concepts, with Shaping, Respondent Learning, and Behavior Objectives showing the most extensive changes. Note that the changes have positive signs, indicating a change in meaning from less favorable to more favorable. The contributions of each concept to the significant overall change are represented in the step down F-ratio. Reading up from the bottom of that column, it is apparent that the first significant F is associated with the concept Shaping. All F ratios above that point must be considered significant and the three below insignificant. This indicates that the instructionally relevant concepts did, in fact, change in meaning, while the

Table 6.--Multivariate analysis of variance results of change in the evaluative dimension.

Concept	Individual Concept Changes				
	$\bar{X}$ Change	Univariate F	p<	Step Down F	p<
Myself as a Teacher	.1748	14.6749	.0002	14.6749	.0002
Nonverbal Behavior	.3288	30.3106	.0001	26.2311	.0001
Questioning and Listening Skills	-.0594	1.2661	.2616	1.8232	.1782
Myself	.2376	28.7822	.0001	16.2958	.0001
Behavioral Objectives	.5745	57.2381	.0001	46.8987	.0001
Reinforcement	.1273	4.8684	.0283	1.2254	.2694
Respondent Learning	.6662	89.0564	.0000	27.6389	.0001
Shaping	1.0439	205.0043	.0000	87.6834	.0000
Physician	.2826	14.4736	.0002	0.5649	.4531
Religion	.1295	4.3497	.0381	0.9137	.3401
Marijuana	.1762	7.3671	.0072	1.3500	.2465

Overall change: F = 28.4324 p< .0000

Table 7.--Multivariate analysis of variance results of change in the potency dimension.

Concept	Individual Concept Changes				
	$\bar{X}$ Change	Univariate F	p<	Step Down F	p<
Myself as a Teacher	.2739	30.4171	.0001	30.4171	.0001
Nonverbal Behavior	.2594	10.6794	.0013	4.7458	.0304
Questioning and Listening Skills	.2426	12.1528	.0006	3.2651	.0720
Myself	.1766	8.7039	.0035	1.6756	.1968
Behavioral Objectives	.4383	34.3504	.0001	14.7340	.0002
Reinforcement	.3536	26.3572	.0001	9.1825	.0028
Respondent Learning	.4112	36.1644	.0001	6.8243	.0096
Shaping	.4318	38.0995	.0001	7.4633	.0068
Physician	.2606	15.4934	.0002	1.6291	.2031
Religion	.1841	5.7773	.0170	0.5850	.4451
Marijuana	.0440	0.3373	.5620	1.1745	.2796

Overall change: $F = 8.2354$ $p < .0001$

Table 8.--Multivariate analysis of variance results of change in the leniency dimension.

Concept	Individual Concept Changes				
	$\bar{X}$ Change	Univariate F	p<	Step Down F	p<
Myself as a Teacher	.1032	5.6118	.0186	5.6118	.0186
Nonverbal Behavior	.0567	0.7388	.3909	0.0209	.8852
Questioning and Listening Skills	.1768	7.9529	.0052	7.6727	.0061
Myself	.0682	1.7833	.1830	0.2320	.6305
Behavioral Objectives	.0186	0.0659	.7976	0.3154	.5750
Reinforcement	-.0247	0.1224	.7268	0.5453	.4610
Respondent Learning	.2550	17.3521	.0001	11.1204	.0010
Shaping	.4318	37.3589	.0001	25.6025	.0001
Physician	.0292	0.2161	.6425	0.5677	.4519
Religion	-.0341	0.2038	.6521	0.3195	.5725
Marijuana	.0166	0.0530	.8182	0.0608	.8055

Overall change: F = 4.9861 p< .0001

Table 9.--Multivariate analysis of variance results of change in the personal-evaluative dimension.

Concept	Individual Concept Changes				
	$\bar{X}$ Change	Univariate F	p<	Step Down F	p<
Myself as a Teacher	.1215	4.6531	.0320	4.6531	.0320
Nonverbal Behavior	.0409	0.2629	.6086	0.2241	.6361
Questioning and Listening Skills	.1191	1.8549	.1745	1.7434	.1880
Myself	.1544	6.2616	.0130	4.6452	.0322
Behavioral Objectives	-.0739	0.8365	.3613	1.4277	.2333
Reinforcement	.0019	0.0006	.9809	0.1368	.7118
Respondent Learning	.0197	0.0646	.7996	0.0362	.8498
Shaping	.1528	4.0215	.0460	3.6331	.0579
Physician	-.1269	2.3823	.1240	2.5557	.1112
Religion	-.1275	0.1454	.7034	0.0019	.9649
Marijuana	.0265	0.0937	.7598	0.0044	.9472

Overall change: F = 1.7427 p< .0651

Table 10.--Multivariate analysis of variance results of change in the overall meaning intensity.

Concept	Individual Concept Changes				
	$\bar{X}$ Change	Univariate F	p<	Step Down F	p<
Myself as a Teacher	.1891	25.3483	.0001	25.3483	.0001
Nonverbal Behavior	.2317	12.7417	.0005	7.9162	.0053
Questioning and Listening Skills	.0871	2.4672	.1176	0.1974	.6572
Myself	.1784	17.3381	.0001	4.1426	.0429
Behavioral Objectives	.3967	28.5496	.0001	14.6027	.0002
Reinforcement	.1737	7.8106	.0056	0.0067	.9348
Respondent Learning	.6785	89.2117	.0000	48.6665	.0001
Shaping	.8274	127.4983	.0000	44.8386	.0001
Physician	.1907	9.8464	.0020	0.7216	.3965
Religion	-.0375	0.3157	.5748	1.7083	.1925
Marijuana	-.0355	0.2187	.6405	2.3147	.1295

Overall change: F = 16.8961 p< .0000

noninstructional concepts remained relatively unchanged in evaluative meaning. However, there were two instructional concepts which changed little in meaning as demonstrated by the small univariate and step down F's. These were Questioning and Listening Skills and Reinforcement.

Table 7 presents the results of the analysis of change in the second orthogonal dimension, potency. Once again there is a significant overall change ($F=8.2354$, $p<.0001$) and all changes have positive signs, reflecting a change from less potent to more potent. The high degree of similarity in change patterns with the evaluative dimension is also reflected in the large changes in the carrel concepts relative to the others and in the significant change in instructional concepts but not in the noninstructional concepts.

This same pattern is again reproduced in Table 8, which reports the results of the change analysis for the leniency dimension, but there are some important differences to be noted. The vector of mean differences does depart significantly from zero, but fewer individual concepts make important contributions to this difference. Once again, reading up from the bottom of the step down F's, it is apparent that the noninstructional concepts did not change in meaning. However, in this case it is also apparent that only two of the four carrel concepts (Shaping and Respondent Learning) and two of the four IPL concepts (Myself as a Teacher and Questioning and Listening Skills) contributed

significantly to the overall change. The column of mean changes in this case reflects changes to a meaning which is more "lenient" in nature, since positive signs predominate. The two negative values reveal relatively minor changes toward a more "severe" meaning.

In Table 9 the results of the analysis of change in the personal-evaluative dimension are reported. In this case there was no significant departure from zero, thus no change took place here. This means that the concepts were not seen as being any more enjoyable or pleasant after instruction than they were before instruction.

Finally, in Table 10 the results of the analysis of change in overall meaning intensity are reported. Given significant changes in three of the four orthogonal dimensions of meaning reported above, this analysis was academic, because this score is a composite of the others. The primary reason of the execution of this phase was to determine where the largest changes in overall "meaningfulness" occurred. The reader will recall that as a result of the mathematical computation of this value, a positive change would reflect movement away from the origin of meaning where the axes intersect, thus indicating a greater intensity of meaning or more meaningfulness. Examination of the column of mean changes reveals that nearly all of the instructional concepts are seen as more meaningful after instruction, while the three noninstructional concepts showed little change and two were in a less meaningful

direction. Further examination of this column reveals that two of the eight instructional concepts tended to change little in meaning intensity. These were the concepts Questioning and Listening Skills and Reinforcement. The largest change in overall meaning was in the concept Shaping, followed by Respondent Learning. The remaining concepts in order of magnitude of change from highest to lowest were Behavioral Objectives, Nonverbal Behavior, Myself as a Teacher, and Myself. It is apparent that the largest changes occurred in the carrel or more technical concepts.

The results of this analysis of change in meaning provide a fairly clear sketch of the alterations in meaning that occurred during the period when instruction was being carried out. However, the analysis as it has been described up to this point provides only part of the complete picture which can be drawn from the data. It says little about the actual meaning ascribed to the concepts by the respondents. In order to complete the picture, mean factor scores on each of the four dimensions of meaning for each concept were computed for pretest and posttest responses. These means are reported in Table 11. In order to facilitate interpretation of these values, a three-dimensional Euclidean space was constructed to represent only the three dimension in which a significant change in meaning was detected. It was then possible to place each of the concepts as points in this space. This graphic representation appears in Figures 16

Table 11.--Factor score means for each of the four major meaning dimensions for each concept.

Concept	Pretest					Posttest				
	Evaluative	Evaluative Self	Leniency	Potency	Overall Intensity	Evaluative	Evaluative Self	Leniency	Potency	Overall Intensity
Myself as a Teacher	.976	1.588	1.828	1.542	3.352	1.152	1.710	1.931	1.160	3.541
Nonverbal Behavior	1.189	.978	.989	.994	2.755	1.518	1.019	1.046	1.253	2.987
Questioning and Listening Skills	1.937	1.188	.984	1.069	3.228	1.877	1.306	1.161	1.312	3.315
Myself	.948	1.361	1.872	1.312	3.238	1.186	1.516	1.940	1.489	3.417
Behavioral Objectives	1.105	.970	.907	.770	2.538	1.679	.869	.925	1.208	2.935
Reinforcement	1.738	1.098	1.005	.995	3.011	1.865	1.101	.981	1.350	3.185
Respondent Learning	.925	1.067	.727	.736	2.232	1.591	1.087	.982	1.148	2.911
Shaping	.544	.881	.689	.788	2.114	1.588	1.034	1.121	1.220	2.942
Physician	1.332	.699	1.189	1.038	2.485	1.614	.573	1.217	1.299	3.037
Religion	.629	1.206	.915	.919	2.834	.758	1.179	.881	1.103	2.797
Marijuana	-.819	1.098	.689	.644	2.789	-.642	1.125	.705	.688	2.754

through 18. The IPL, carrell, and noninstructional concepts are reported on separate sets of axes for clarity.

It is apparent from these figures that each concept from the instructional categories is seen as favorable, lenient (gentle), and potent and these reactions tend to intensify during instruction. In addition, there is an apparent cluster of instructionally relevant concepts along a value of just less than one on the leniency scale. This cluster includes the following concepts: Nonverbal Behavior, Questioning and Listening Skills, Behavioral Objectives, Reinforcement, Respondent Learning, and Shaping. A second cluster appears further out on the leniency scale, which includes Myself and Myself as a Teacher. The noninstructional concepts have a large variance in meaning between them, as might be expected because of differences in their nature. The single departure from the lenient-favorable-potent response pattern is seen in the noninstructional concept Marijuana, which is portrayed as unfavorable (bad) in meaning.

Instructor Role in Meaning Change

The final research hypothesis to be tested by statistical means anticipated that there would be differential changes in meaning depending on the instructor with whom the student was associated during the course. In order to test this hypothesis, two analyses of different portions of the data were carried out. First, in order to

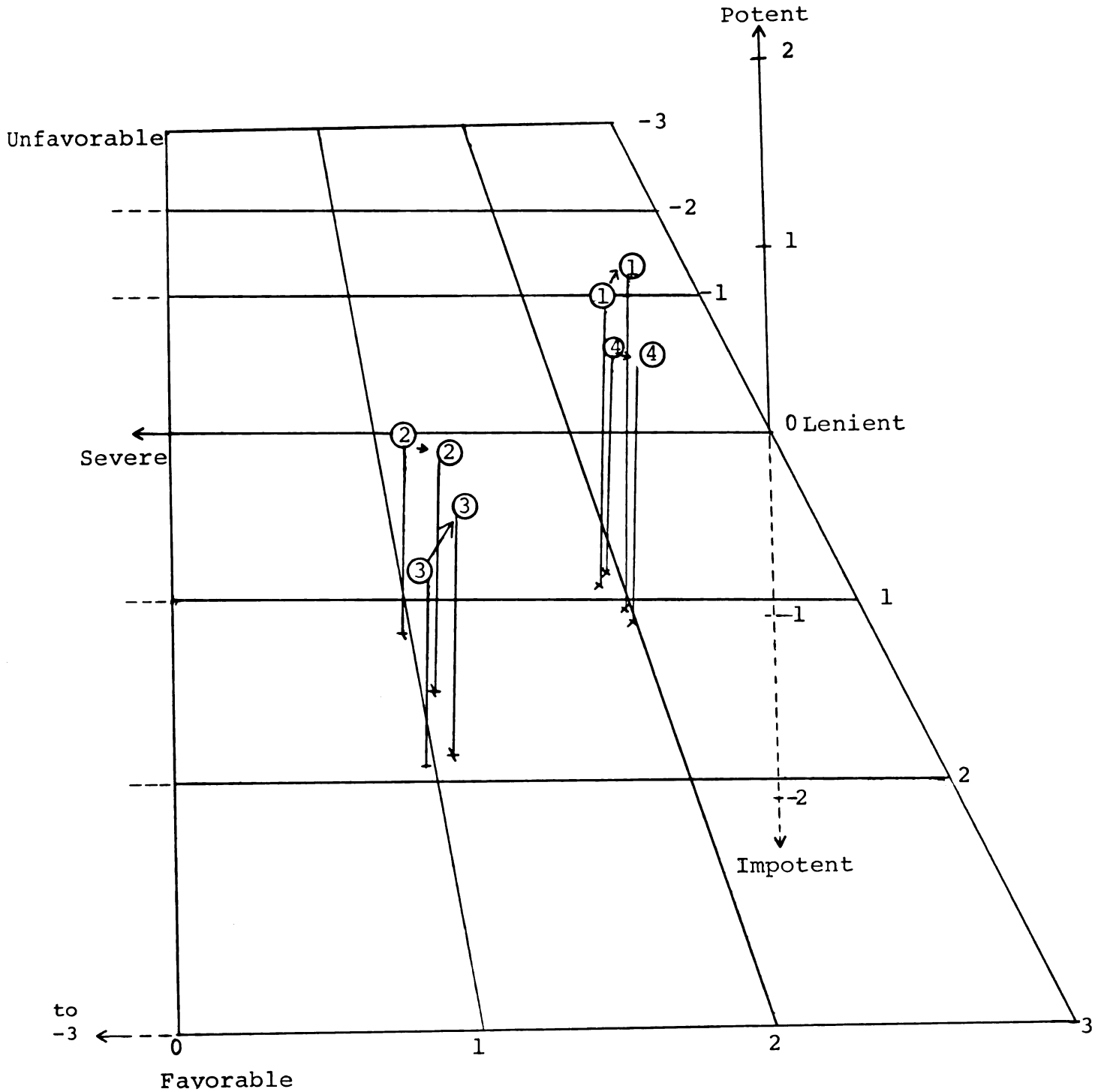

Figure 16.--Graphic representation of pretest and posttest meaning assigned to IPL concepts:

1. myself as a teacher
2. nonverbal behavior
3. questioning and listening skills
4. myself

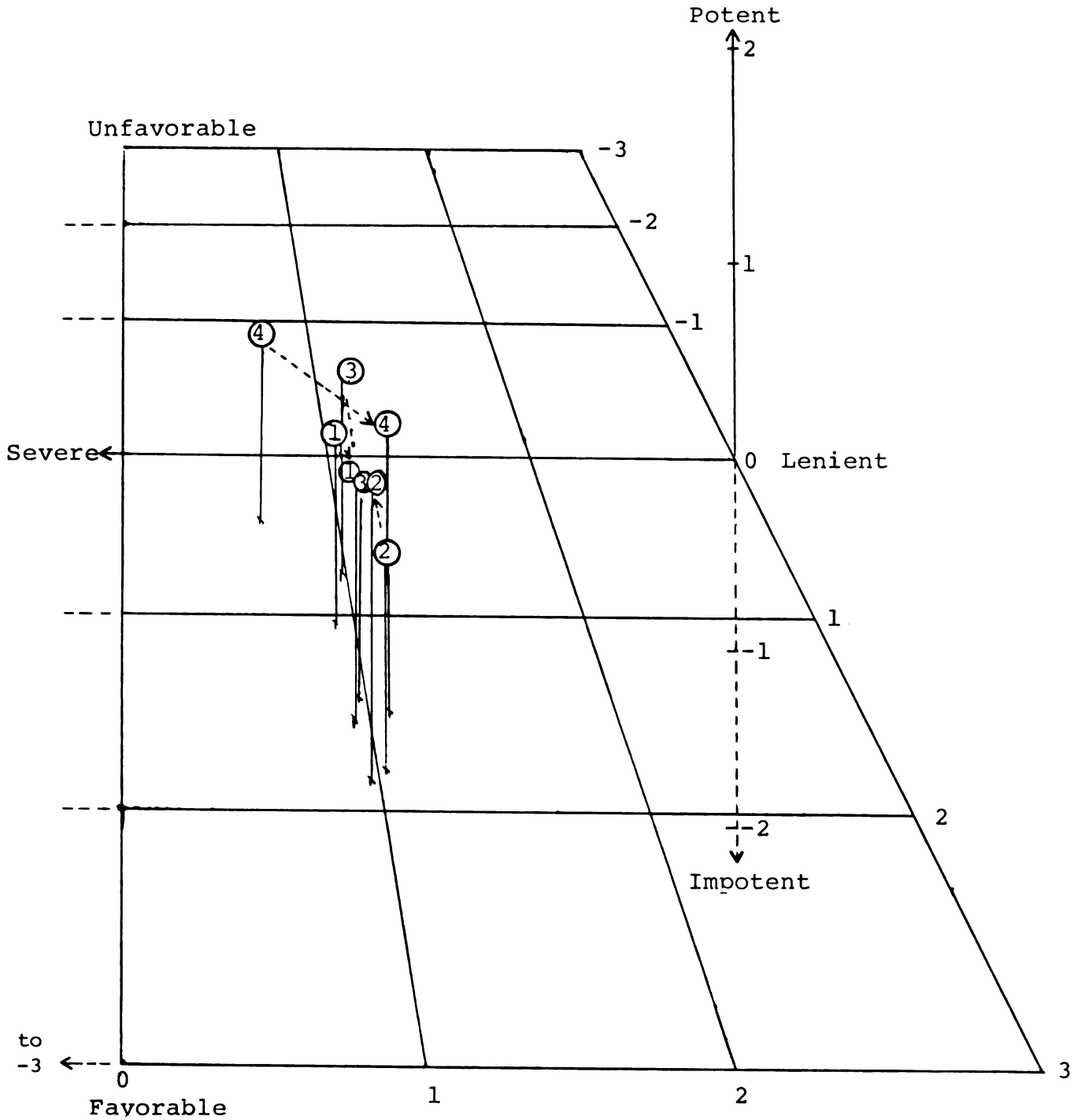

Figure 17.--Graphic representation of meaning assigned to Carrel concepts:

1. behavioral objectives
2. reinforcement
3. respondent learning
4. shaping

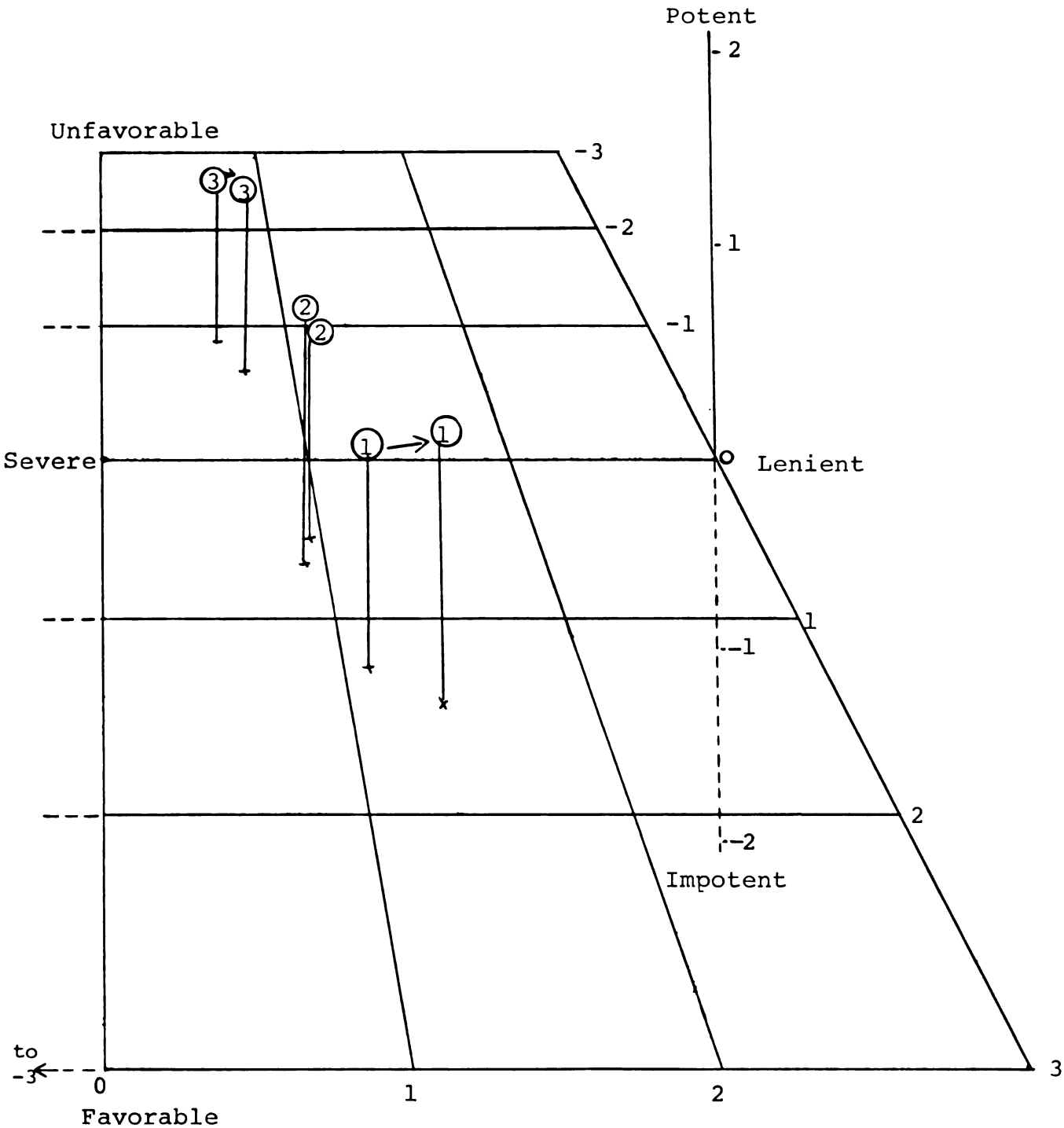

Figure 18.--Graphic representation of pretest and posttest meaning assigned to noninstructional concepts:

1. physician
2. religion
3. marijuana

establish the base line on which to plot the changes by instructor, differences between instructors on the pretest responses were examined. In this analysis, instructor with 19 levels was the independent variable and there were three dependent measures: average ratings of carrel concepts for each student, average rating of IPL concepts, and average rating of noninstructional concepts. A one-way multivariate analysis of variance was carried out for each orthogonal dimension of meaning where a significant overall change in meaning had been detected. The results of this test, reported in the left hand columns of Table 12, indicate that there were no differences in initial ratings of the concepts by students studying with different instructors.

In order to assess differential change in meaning, a similar analysis was carried out with instructor as the independent variable, but with the three dependent measures redefined as the difference between the average carrell, IPL, and noninstructional ratings for each subject on each of the three orthogonal dimensions where significant change had occurred. Once again there was a negative result, as demonstrated in Table 12. This was a firm indication that there were no differences in the amount or direction of change in meaning for students studying with different instructors. The mean pretest rating and changes for different instructors can be used to clearly represent the nature of this result.

Table 12.--Results of manova of instructor differences in pretest meaning and change in meaning on the three dimensions where a significant change was detected.[a]

	Univariate F	p<	Step Down F	p<
Evaluative Dimension				
Pretest: Multivariate F = 1.3248 p< .0643				
IPL Concepts	1.0274	.4297	1.0274	.4297
Carrel Concepts	1.9881	.0112	1.8386	.1221
Noninstructional Concepts	1.3084	.1833	1.1324	.3217
Change: Multivariate F = 1.0180 p< .4418				
IPL Concepts	.6556	.8522	.6556	.8522
Carrel Concepts	1.4406	.1139	1.4229	.1217
Noninstructional Concepts	1.0577	.3968	.9935	.4681
Leniency Dimension				
Pretest: Multivariate F = .9290 p< .6207				
IPL Concepts	1.0961	.3570	1.0961	.3570
Carrel Concepts	1.0356	.4207	.8288	.6654
Noninstructional Concepts	.9880	.4744	.8725	.6126
Change: Multivariate F = .9265 p< .6257				
IPL Concepts	.4398	.9779	.4398	.9779
Carrel Concepts	1.1185	.3349	1.1098	.2698
Noninstructional Concepts	1.0288	.4281	1.1675	.2897
Potency Dimension				
Pretest: Multivariate F = 1.0457 p< .3891				
IPL Concepts	.8950	.5852	.8950	.5852
Carrel Concepts	.7549	.7513	.8964	.5835
Noninstructional Concepts	1.2844	.1989	1.3580	.1542
Change: Multivariate F = 1.0366 p< .4062				
IPL Concepts	1.1179	.3355	1.1179	.3355
Carrel Concepts	.8861	.5960	1.0751	.3785
Noninstructional Concepts	1.0810	.3740	.9271	.5464

[a]Dependent variables were average over carrel, IPL, and noninstructional concepts for each subject.

A sample of these means has been plotted in Figure 19. Two major points can be made from this representation. First, this randomly selected group of 5 of the 19 instructors began the term with students who had very similar reactions to the concepts on the evaluative dimension as demonstrated by the narrow range of means (1.05 to 1.37 and .91 to 1.42). Second, the parallel nature of the lines indicating change indicates the high degree of similarity in the changes which took place among students associated with different instructors. Though this figure represents only a subset of the data, the results reported in Table 12 indicate that graphic representations of the remaining data would be very similar to that pictured above.

Given this negative result, it was decided that the data might be more closely scrutinized from the point of view of instructor differences by looking at these differences on individual concept ratings. The prior analysis was based on a composite (mean) of all the subject's responses to the carrel concepts, for example, not his individual responses to each concept in that category. It must be recognized at this point that the ideal follow-up to the negative findings in the prior analysis would have been to seek instructor differences in the change in meaning of each individual concept on each of the meaning dimensions. However, since this would have required 44 separate analyses (4 dimensions by 11 concepts), the decision was made to further assess the degree of variations among instructors

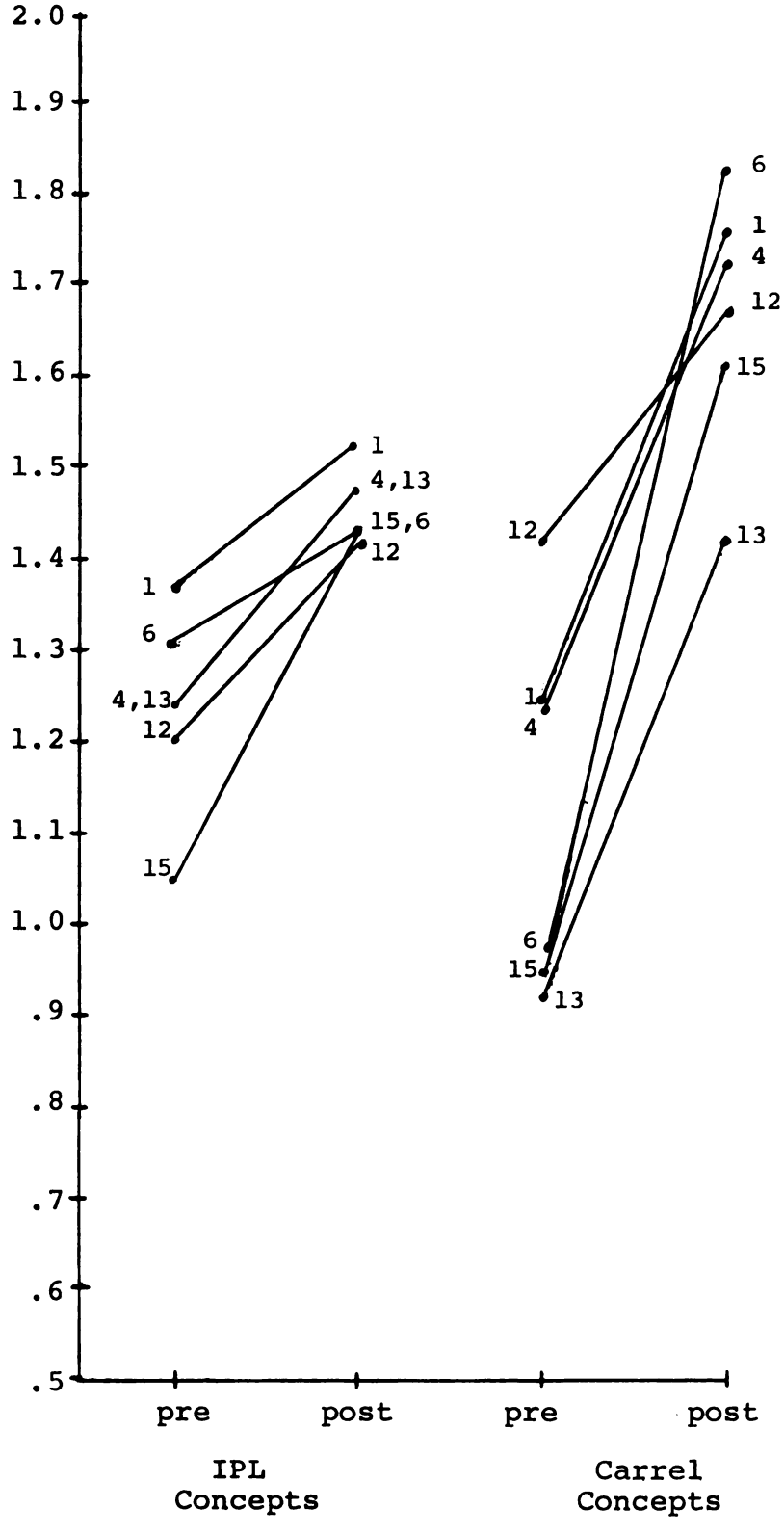

Figure 19.--Sample of data representing role of instructor in meaning change. Five of 19 instructors selected at random are graphed with respect to the average rating of their students on the IPL and Carrel concepts on the evaluative dimension.

on concepts where they would be most clear, and yet attempt to represent each concept category and meaning dimension. For this reason, nine ANOVAs of change scores were carried out with instructor as the independent variable and changes in the meanings of the concepts and dimensions listed in Table 13 as the dependent variables. Once again, however, there were no differences among instructors, as indicated by the F-ratios and their probabilities of occurrence.

Table 13.--Results of analyses of variance of change in meaning over instructors for individual concepts (not averaged within concept category).

Dimension	Concept	Univariate F	p<
Evaluative	Nonverbal Behavior	1.2993	.1891
	Shaping	1.4718	.1012
	Physician	1.6773	.0443
Leniency	Questioning and Listening Skills	16844	.8253
	Shaping	.5998	.8981
	Religion	.5143	.9501
Potency	Myself as a Teacher	1.2228	.2436
	Behavioral Objectives	.6311	.8735
	Physician	1.0715	.3822

Conclusion

The major results of this attempt to clarify the nature of the alterations in connotative meaning which take place during the period of teacher training experience defined above can be summarized in terms of the three major research hypotheses tested. A three-factor solution to the

SD ratings gathered is not the most appropriate factorial model. In fact, a fourth factor must be added to gain a parsimonious explanation of the data. The factors are not the traditional evaluative, potency, and activity dimensions of meaning which has characterized SD research, though there are indications of part of that structure. The four-factor model has demonstrated a high degree of stability over time where the interval between measures was about nine weeks. However, there are significant changes in concept meaning within this stable meaning framework. The largest of these changes are associated with more technical concepts from among the 11 rated and the smallest were those associated with the noninstructional concepts. For the most part, changes in these noninstructional concepts were chance occurrences. Finally, it was apparent that the instructor played little role in terms of differential changes in their respective changes. Given these results, there is some discussion and speculation which seems warranted. This will be presented in the fifth and final chapter and will revolve around the statistical tests which sought answers to three of the questions posed, as well as the questions which did not lead directly to research hypotheses.

CHAPTER V

DISCUSSION

Introduction

The common thread throughout this dissertation has been the set of questions of interest which were posed at the outset. Once again these can serve as the framework for discussion, but since the analysis is complete and the results are now available, it is possible to formulate direct answers to each. This will comprise much of the discussion in this chapter. Then, in order to bring all of the threads of discussion together near the starting point, some conclusions will be drawn concerning the potential of semantic differential research in the systematic and scholarly evaluation of instruction in teacher training.

Scale Interrelationships and Factor Analysis

The purpose of the series of algebraic manipulations termed factor analysis is to thoroughly examine and draw conclusions about the interrelationships among a large number of variables by maximizing the conciseness and simplicity of the exploration of the intercorrelations. Though many procedures have been developed for performing factor analysis (see Cattell, 1966), each is aimed at the major purpose stated above. More specifically, according to Thurstone

(1947) and Nunnally (1967) there are two primary uses of factor analysis: (1) to test hypotheses about the existence of certain latent constructs (factors) and/or (2) to reduce data into an interpretable number of distinguishable primary factors. The "and/or" conjunction in the last sentence is critical in discussing the results of this research, because it was necessary to use both of these in combination.

The first of the two primary uses assumes that the researcher has sufficient reason to expect that a certain latent structure should be manifested in his data. The actual structure may then be compared with the anticipated structure to test the truth or falsity of the hypothesis. This was precisely the case in the research being presented here. It was hypothesized that a three-factor structure (evaluative, potency, and activity) would explain the responses, and this prediction was based on extensive semantic differential research and careful scale selection to tap those dimensions. However, the latent structure which generally characterizes responses to concepts from the general domain of meaning did not seem to apply in the rating of educational concepts. Therefore, the answer to the first question posed: "Does EPA structure fit?" was no. There was another, more adequate explanation to be discovered.

Before proceeding further in the discussion of the scale interrelationships, some comments about the hypothesis of EPA structure seem warranted. In the original research which reports this structure, Osgood and his research

associates (discussed in Chapter II) never imply that these three dimensions are the entire basis for an individual's reaction to concepts being rated on bipolar scales. The EPA dimensions were considered the primary dimensions and the researchers were willing to acknowledge that there were other, less important dimensions which their efforts had not been able to clearly define. The "primary" nature of these dimensions can be defined in statistical terms and this will serve to clarify the point being made here. As Heise (1969) points out in his review of the methodological SD research, the evaluative dimension generally accounts for around 30 per cent of the common variance in responses and the other two dimentions (potency and activity) tend to account for around 15 per cent each. The remaining 40 per cent is explained by minor undefined dimensions orthogonal to these three. Yet in the research reported here, an attempt was made to prove that these dimensions accounted for nearly all of the common variance. Though this attempt was not stated outright, it was implied in the statistical test used (UMFLA), which, for the number of subjects used, requires that approximately 90 per cent of the common variance be explained to satisfy the criterion of an adequate fit of the factor model. In all fairness to the original researchers, this was an unnecessarily strict criterion. However, the failure of the three-factor EPA model to fit did manifest itself in rational as well as statistical terms and also led to some interesting results, as the further assessment of the nature of

the interrelationship proceeded to more complex factor models.

It was at this point, then, that the second use of factor analysis came into play. The question of interest then became: "Is there a distinguishable set of constructs which adequately and parsimoniously explain the scale interrelationships?" It might be argued at this point that, once the initial hypothesis was found to be inadequate, further examination of the correlations is unnecessary. Given a negative result for the three-factor hypothesis, it would seem that one of three courses of action is warranted: (1) Stop the analysis and report the inadequacy of the hypothesized latent structure, if that is the only hypothesis of interest; (2) Leave the question of a parsimonious reduction of the data unanswered and proceed to test other relevant hypotheses on the basis of raw score data; or (3) Attempt to search beyond the hypothesized structure for an alternative, yet equally plausible, latent structure.

The first choice was unacceptable here because the most important hypotheses were to be tested after the decision as to the most appropriate latent structure (i.e., tests of concept meaning change). The second was equally unacceptable because of the great difficulty in data manipulation and interpretation involved in explaining the large numbers of raw responses. Therefore, because of a need for a simple structure for reduction of the data and in order to clearly understand the nature of the results, the third option was

chosen. This turned out to be a fruitful course because the result was a four-factor latent network which possessed the qualities of fair statistical integrity and a high degree of psychological interpretability.

From a psychological point of view there was a high degree of similarity between the four factors at the beginning and at the end of the instructional period; that is, the same scales loaded together on the pretest as on the posttest. Statistical tests of this stability will be discussed below. The first factor was most easily identified. It was clearly an evaluative factor of the type frequently found in SD research and which reflects a favorable-unfavorable reaction to the concepts rated. The second factor was equally clear in meaning. It was comprised of the scales enjoyable-unenjoyable and pleasant-unpleasant and was named a personal evaluative dimension. It is interesting to note that conceptually this factor is quite similar in definition to the evaluative scale, but statistically it was quite independent. This is a clear indication that the respondents were registering two distinct reactions to the concepts. That is, whether or not they found a concept personally pleasing had little to do with their judgments of the concept's worth. So that, for example, the typical student reaction to the use of behavioral objectives in terms of how enjoyable and pleasant they would be to them personally had little to do with the value and importance they placed on the concept.

The remaining two dimensions of meaning were quite clear conceptually, but were difficult to name. They were a combination and redivision of Osgood's potency and activity dimensions, with both potency and activity scales combined in each of the new factors. For example, the scale weak-powerful, which is usually the primary scale on Osgood's potency dimension, was highly related to active-passive, which is usually the primary scale on his activity dimension. This was an unanticipated result and is very difficult to explain. The most satisfying explanation from the author's point of view comes from a comparison of this dimension with the fourth dimension, leniency, and the scales which characterize it (lenient-severe, tense-relaxed, violent-gentle, etc.). These scales seem to reflect more latent or less observable characteristics of the concepts, while the power and activity scales tend to reflect the more outward, observable, or manifest characteristics of the concepts. These are clearly meaningful and distinct psychological reactions which the respondents have made to the concepts rated. The naming of the factor described as latent characteristics which allowed for the most meaningful reference was taken from the first scale, severe-lenient, and was the leniency scale. The naming of the power and activity scale was more difficult, but the name potency was chosen as a combination of the nuances of meaning found in the scales.

The primary limitation in the evaluation of the scale interrelationships is found in the statistical criterion for

the definition of the most adequate factor structure. This was discussed in detail in Chapter IV, but is worthy of further discussion here. There were two statistical criteria--one liberal and one very conservative. These were the scree test and the χ^2 criterion of the UMFLA procedure developed by Joureskög. The reader will recall that the lack of precision in the data gave rise to factor solutions using UMFLA which went beyond rational or psychological interpretability (in terms of the number of latent constructs) in order to find a statistical fit. The more liberal scree test, on the other hand, gave rise to clear psychological interpretability, but fell short from a statistical point of view, because it lacked clear statistical antecedents. The decision was made to proceed with the four-factor solution with the realization that future research must clarify this issue of best factor explanation. This is simply a question of acknowledging the value of these four dimensions in terms of their relevance to the task of assessing student ratings to instructional concepts in teacher education and then verifying them by attempting to compose and try new scales which more clearly tap the dimensions of interest. This would, hopefully, tend to clarify the factor structure and eliminate the problems experienced here with the statistical criterion.

This limitation caused by unexplainable variance in the data led to a second problem in the assessment of the stability of the factor structure from pretest to posttest.

Both the UMFLA procedure and the analysis of covariance structure, which was one of the tests of the hypothesis of factor invariance, require a high degree of precision in order to satisfy the requirement of the statistical tests involved. The reader will recall from the presentation of results, that each of the other four tests of factor stability indicated a very consistent factor structure over time. However, the results of the ANCOVST, incomplete though they were, would have cast doubt on stability defined by the other tests. The procedure described above for clarifying the factor structure, by employing more scales that tap the dimensions of interest and deleting those which tend to cloud the issue, would tend to assist in the use of the ANCOVST decision apparatus also. This will be carried out in ongoing research.

Little that is of additional interest can be said about the stability question beyond the fact that there was a very high degree of consistency. That is, the responses that students made to the concepts or the manner in which they used the scales to describe their reaction were very similar across concepts and across testing times.

The Question of Concept Meaning Change

The function of the scales used to assign connotative meaning to concepts did not change in their basic nature or structure as demonstrated by the stability of the factor structure, but the meaning of instructionally related

concepts did change in intensity and definition. In addition, there is some indication that the course instruction played a role in this change because of the relatively minor changes that took place in the intensity of the reactions to the noninstructional concepts. Here again, however, there is a limitation mentioned in earlier chapters which must be restated. The degree of control over the variables of interest which is gained by using control concepts rather than a control group of subjects must be considered less adequate than if a control group had been employed. For this reason, a true cause and effect relationship between instruction and concept meaning change must remain to be proven by additional experimental manipulation of the variables. However, for purposes of the following discussion and in the absence of an immediately apparent rival explanation of the cause of change, it is going to be assumed that the course of instruction played at least some part in bringing about the change in meaning intensity which has been demonstrated.

With this assumption in mind, then, it becomes possible to proceed to a discussion of the nature of the changes that occurred and the reasons for their occurrence, and beyond that to a discussion of the questions of the appropriateness of the meanings and the changes therein for teacher education.

The reader will recall that the largest changes in the overall intensity of meaning were demonstrated in the carrel concepts Behavioral Objectives, Respondent Learning,

and Shaping. These changes can be clearly interpreted in a manner which makes psychological sense when one considers the amount of prior experience or prior meaningful associations which the students had with these concepts. Concepts such as Myself, Myself as a Teacher, Nonverbal Behavior, Questioning and Listening Skills, and Reinforcement are all familiar terms with which the respondents had had at least some prior experience. The three technical concepts mentioned above, on the other hand, were probably very new to most of the respondents and were, therefore, more easily altered through instructional experience. This is not to say that the more familiar terms did not change, because for the most part they did. But what is being suggested here is that there is an inverse relationship between familiarity with a concept and the amount of change in connotative meaning which can be achieved in a person's reactions to that concept if the amount of energy expended (instruction) in bringing about that change is held constant. This may be considered rather clear evidence of a point made by Osgood in his development of the semantic differential as an index of meaning. In the case of a lack of familiarity, a response near the neutral point in terms of connotative meaning scales is elicited. As familiarity increases, the intensity of the meaning associations increases. The direction of the change depends on the nature of the experience, but as one moves further out on the scale, it becomes more difficult to bring about change. This might be interpreted as a statistical

artifact resulting from the attenuation of responses at the limits of the scale. This probably contributed to the effect being discussed, but the amount of the contribution remains an unknown. It is true that if a person has made a response of four or neutral to a scale, the extent to which he may change this response at some later time (in either direction) is maximized. But, the fact remains that he has made that response for some psychological reason and, if he alters it at the time of the later rating, he alters it because of some change in his psychological interpretation between the scale and the concept which he is rating. Regardless of the limitations which the researcher places on the respondent in terms of the range of possible responses, it must be assumed that, if there is a desire to make inferences from the scales to the psychological operations of the respondent, there is a direct connection between the scale and the psychological function being measured.

It is also interesting to note with respect to the meaning of the self-concepts and of the various instructional tools and procedures, that the two clusters seem to differ a great deal on the leniency dimension and to a lesser degree on the potency and evaluative dimensions. These are not vast differences and should probably not be pressed too far, but it is interesting to speculate on the nature of a classroom or learning environment constructed by a person who sees the tools he uses as being more valuable and less lenient than himself as a person. This would appear to be a

wholesome situation indeed, because if these were reversed and the tools were seen as less valuable and more lenient than the self, the tools would probably never be employed.

The location of the noninstructional concepts in the three-dimensional meaning space is of interest, because they tend to add credibility to the scaling procedure. First, the location of the concept Marijuana in the unfavorable domain discounts the potential argument that there is a favorable overall halo effect being demonstrated in the responses. Second, the wide differences in the location of the three concepts Physician, Religion, and Marijuana suggests very different meanings, as there surely must be.

There was suggestion made above concerning the appropriateness of the meanings assigned to the concepts with respect to teaching and teacher training. This was one of the questions posed at the outset which didn't imply any statistical test and which must ultimately by answered by those responsible for the planning of teacher training instruction. However, some discussion and conclusions seem warranted at this point. Certainly, since all of the educationally relevant tools, concepts, and procedures were seen as favorable, lenient, and potent at the outset, students could be considered to be predisposed to use them in the classroom. But above and beyond this, the changes which took place during instruction rendered the concepts even more favorable, potent, and somewhat more lenient, thus appearing to increase the predisposition for their later use. It would appear that

the associations with these few concepts were appropriate for the teaching profession and that the instructional sequences served the purpose for which they were intended.

Instructor Differences in Change in Meaning Intensity

Perhaps the most puzzling result from the point of view of formulating a sensible explanation is that of no differential effect among the 19 instructors. The changes noted in the intensity of the respondents' rating occurred regardless of the instructor with whom the student was associated. No differential effect of instructor in terms of the carrel concepts and the noninstructional concepts might have been expected because these concepts are not systematically presented in the IPL, where the instructor makes his contributions. But to reveal that there were no differences in the IPL concepts either, where it is assumed that the instructor has the greatest influence, was unexpected.

One can, of course, only speculate as to the reasons for this result. It is possible, for example, that the instructor simply does not make a contribution which is clearly and identifiably his in the change which takes place in the students during the ten weeks of IPL meetings. It is equally possible that he does have an effect but that it occurs in ways and is reflected in changes not tapped by this instrument. Third, it may be true that the instructors influence these reactions differentially, that the instrument measures the correct change, and that all instructors

are operating at the same high level of instructional ability. Each of these possible explanations has some merit.

The first implies that there is a general diffuse increase in awareness concerning the meaning of education and teaching which is reflected in the results, and in which individual instructor plays only a uniform part in this growth. This is a distinct possibility because of the short duration of the course (ten weeks) and because of the large amount of information concerning the task demands and the personal demands of teaching which is presented to the student during those ten weeks. It might be worth noting parenthetically that the hypothesis of a diffuse increase in awareness would have been testable by adding a fourth category of concepts related to education, but presented in this particular course. The nature of the change in these concepts would have been most interesting, because a change in these noninstructional but educational concepts would tend to support the general increase in awareness hypothesis.

The second explanation suggested above is based on the extremely limited scope or breadth of the instrument employed. The reader will recall that the scales were selected to represent the traditional EPA structure (see Appendix A) and that the concepts were selected simply as examples of the types of concepts presented in the course. It is very possible that, with only four IPL concepts selected from the large number of possible concepts, there is simply a sampling error being demonstrated here. From a different

point of view, it is also possible that the role of the individual instructors would be clearly demonstrated by a completely different set of dependent measures taken perhaps from sociometry or personality assessment.

Finally, there is the suggestion that the instrument is a valid assessment of an equally adequate performance by all 19 of the instructors. Though this is a difficult explanation to believe because of the large number of individual instructors and students involved, it does find some support in supplementary evaluative research carried out during the terms when these SD data were gathered. In a research report currently in preparation by the Education 200 Research Council (Stiggins and Byers, 1972), it is reported that the students perceived a very high level of performance on the part of the instructors. Students were asked to rate their instructors at midterm and at the end of the term on their performance with respect to facilitating group interaction, displaying the IPL objective, and helping the students to reach those objectives. On a 7-point scale with 7 being the best possible rating in terms of performance, all instructors scored consistently at the 6 level. Though this does not necessarily lead to any conclusions about the concept meaning changes reported here, it does reflect a level of performance in the opinions of the students. It should be made clear that this research merely lends support to the third suggestion as a possible explanation of no instructor differences in change. It is not intended to prove that an

unusually high level of performance was in fact the case.

It is also possible, of course, that some combination of the suggested reasons was operating to bring about these results. The correct reason or combination of reasons for no differential instructor effect must await further examination. Further examination is warranted, because it was suggested in the first chapter that one of the merits of this type of instructional evaluation would be its ability to supply feedback to individual instructors in such a way as to be of value in the reformulation of instruction. Such a procedure could be carried out with the results presented here, but, with the lack of surety that the results reflect something that is, in fact, part of the instructor's behavior or instructional procedures, the value of the feedback would be questionable.

The Question of Other Testable Hypotheses

It was also suggested at the outset that the merit of this type of instructional evaluation would be judged first on its performance in this task and then on its potential to contribute answers to other questions of interest. Given the demonstrated ability of the SD procedure employed here to detect changes in the intensity of concept meaning tempered by the methodological limitations suggested throughout this chapter, it would seem that the procedure was able to supply answers to most of the questions posed at the outset.

A few other hypotheses which might be examined by this procedure come to mind. Some of these are worthy of detailed discussion while others need only be suggested briefly. However, each could serve a further testing ground to the semantic differential in the role of providing feedback about instruction.

First, it was suggested at the outset that the ultimate test of the validity of this procedure would be its ability to predict teacher behavior. Given the generally positive nature of the results reported here, an interesting research design for this test would be as follows: Begin with a group of students who are undergoing some type of professional preparation to become teachers. Gather from these students achievement test data on the concepts being presented both at the beginning and at the end of the course and supplement this with SD ratings of the concepts. This procedure would be most revealing if the concepts presented were instructional tools and procedures such as those used in this study. The correlations among the achievement test items and the SD rating would be interesting enough, but even more revealing would be post-course observations of classroom behavior with special focus on the tools and procedures taught in the course. Using frequency of usage of the tools and procedures as the criterion variable, it would be most interesting to determine which of the two predictor variables was most efficient. Further, it would be revealing to know the efficiency of these two variables in

combination for predicting teacher behavior as compared with other predictors. The ultimate result of this test would determine the validity or the true value of this type of rating in teacher training. The further use of the procedure to test other hypotheses and to serve other evaluative functions would be dependent on a positive result in this test.

If such a result were achieved and a high degree of predictability were revealed, the generalizability of the procedure would be endless. Other concepts and procedures could be assessed from the various courses that compose a teacher training sequence. The role of teacher training, for example, might be closely scrutinized or various modes and methods of instruction in the teacher training sequence might be compared.

Throughout these types of evaluations further scales and dimensions of meaning might be assessed and, if a consensus of relevant scales is reached, an attempt might be launched to construct a semantic atlas of educational concepts like that developed in the general linguistic domain by Snider and Osgood (1969).

The primary interest in the research reported and discussed here has been on the result of instruction on the methodological aspects of teaching (i.e., on tools and procedures of teaching). There is an entirely different realm of interest which might serve as a valuable supplement to the concern expressed here, and that is the use of the SD procedures for a closer look at the interpersonal demands of

teaching. This implies the use of bipolar scales to rate people. This use of the semantic differential is currently being tried in the course where the data for this dissertation were gathered, and the results promise to be most revealing.

But the reader must bear in mind that most of these research possibilities are still in the distant future. The immediate research priorities are there: (1) further data collection and manipulation of the scales and dimensions of the SD instrumentation from a psychometric point of view, (2) a series of more tightly controlled experimental administrations to determine the nature of any cause and effect relationships, and (3) a test of the true validity of this procedure employing teacher behavior as the criterion.

Conclusion

It was stated at the outset that the semantic differential procedure for the scaling of psychological reactions termed the connotative meaning of concepts might have potential value for the systematic and scholarly evaluation of instructional sequences. To begin measuring this potential, an attempt was made to measure changes in the meaning of concepts presented in a course which contributed to the professional preparation of teachers. It was stated further that some indication of the potential would be gained by attempting to find the answers to a series of questions regarding the role of instruction.

It was determined that the traditional dimensionality of semantic meaning space was inadequate, but that a plausible alternative did exist. A series of statistical tests verified the fact that the basic nature of the semantic meaning space was not altered by instruction but that the intensity of reactions to the concepts rated was altered significantly. The meanings ascribed to these concepts were seen as appropriate for teaching and the changes in that meaning appropriate for teacher training. From this research, the course for future investigations seems quite clear. These were all favorable outcomes. However, there were limitations noted which remain to be accounted for in the future study of the role of the SD procedures for course evaluation.

Consequently, it must be concluded that, though there is much more work to be done, there is evidence in support of the initial statement that the use of the semantic differential procedure for the scaling of connotative meaning can, in fact, serve a useful purpose in the systematic and scholarly evaluation of instruction. The measurement technique would seem to have survived its initial test. The next research steps are clearly drawn. The validity of the measurements, using classroom behavior as the criterion, must now be determined.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ahmann, J. S. "Aspects of Curriculum Evaluation: A Synopsis." Perspectives in Curriculum Evaluation. Edited by R. Tyler, R. Gagné, and M. Scriven. New York: Rand-McNally, 1967.

Allport, G. W. "Attitudes." Readings in Attitude Theory and Measurement. Edited by M. Fishbein. New York: John Wiley and Sons, 1967.

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

Barrett, G. V., and Otis, J. L. "The Semantic Differential as a Measure of Changes in Meaning in Educational and Vocational Counseling." Psychological Reports, XX (1967), 335-338.

Bentler, P. M. "Semantic Space Is (Approximately) Bipolar." Journal of Psychology, LXXI (1969), 33-34.

Block, J. "Studies in the Phenominology of Emotion." Journal of Abnormal and Social Psychology, LIV (1957), 358-363.

Bloom, B. S.; Hastings, J. T.; and Madans, G. F. Handbook of Formative and Summative Evaluation. New York: McGraw-Hill, 1971.

Brown, F. G. Principles of Educational and Psychological Testing. Hinsdale, Illinois: Dryden, 1970.

Burt, C. L. "Factor Analysis and Canonical Correlation." British Journal of Psychology (Stat. Sec.), I (1948), 95-106.

Campbell, D. T., and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research. New York: Rand-McNally, 1963.

Cattell, R. B. Handbook of Multivariate Experimental Psychology. New York: Rand-McNally, 1966.

Cattell, R. B., and Baggley, A. R. "The Objective Measurement of Attitude Motivation: Development and Evaluation of Principles and Devices." Journal of Personality, XXIV (1956), 401-423.

Cliff, N. "Adverbs as Multipliers." Psychological Review, LXVI (1959), 27-44.

Cook, D. R. "A Study of the Relationship of the Meaning of Selected Concepts to Achievement and Ability." Unpublished Doctoral dissertation, Indiana University, 1959.

Darnell, D. "A Validity Study of the Semantic Differential." Unpublished Doctoral dissertation, Michigan State University, 1964.

Davitz, J. R. The Language of Emotion. New York: Academic Press, 1969.

Dingman, H. F.; Paulson, M. J.; Eyman, R. K.; and Miller, C. R. "The Semantic Differential as a Tool for Measuring Progress in Therapy." Psychological Reports, XXVI (1962), 271-279.

DiVesta, F. J. "A Developmental Study of the Semantic Differential." Journal of Verbal Behavior, V (1966), 249-259.

DiVesta, F. J., and Dick, W. "The Test-Retest Reliability of Children's Ratings on the Semantic Differential." Educational and Psychological Measurement, XXVI (1966), 605-616.

Ebel, R. L. "What Are Schools For?" Phi Delta Kappan, LIV (1972), 3-7.

Endler, N. S. "Changes in Meaning During Psychotherapy as Measured by the Semantic Differential." Journal of Counseling Psychology, VIII (1961), 105-111.

Ervin, S. M., and Foster, G. "The Development of Children's Terms." Journal of Abnormal and Social Psychology, LXI (1960), 271-275.

Feshbach, N. D., and Beigel, A. "A Note on the Use of the Semantic Differential in Measuring Teacher Personality and Values." Educational and Psychological Measurement, XXVIII (1968), 923-929.

Flavell, J. The Developmental Psychology of Jean Piaget. New York: Van Nostrand, 1962.

Gage, N. L., ed. Handbook of Research on Teaching. New York: Rand-McNally, 1963.

Gagné, R. M. "Curriculum Research and the Promotion of Learning." Perspectives in Curriculum Evaluation. Edited by R. Tyler, R. Gagné, and M. Scriven. New York: Rand-McNally, 1967.

Glaser, R. "Psychology and Instructional Technology." Training Research and Education. Edited by R. Glaser. Pittsburgh: University of Pittsburgh Press, 1962.

Green, R. F., and Goldfried, M. R. "On the Bipolarity of Semantic Space." Psychological Monographs, LXXIX (1965), 6 (Whole No. 599).

Greenberg, B. S.; Bowes, J.; and Kagan, N. "Dimensions of Empathetic Judgment of Clients by Counselors." Journal of Counseling Psychology, XVI (1969), 303-306.

Griggs, A. E. "A Validity Study of the Semantic Differential Technique." Journal of Clinical Psychology, XV (1959), 179-181.

Hartley, J. E. "A Semantic Differential Scale for Assessing Group Process Change." Journal of Clinical Psychology, XXIV (1968), 74.

Heise, D. R. "Some Methodological Issues in Semantic Differential Research." Psychological Bulletin, LXXII (1969), 406-422.

Henderson, J. E. "Individual and the School: Ed. 200 Course Outline and Description." East Lansing, Michigan: Michigan State University, 1972. (Mimeographed.)

Henderson, J. E.; Willard, S. M.; Barnes, H. L.; and Prawat, R. S., eds. Individual and the School. East Lansing, Michigan: Michigan State University, 1972.

Hoffman, J. E. "An Analysis of Concept Clusters in Semantic Interconcept Space." American Journal of Psychology, LXXX (1967), 345-354.

Hoover, K. H., and Schutz, R. E. "Student Attitude Change in an Introductory Education Course." Journal of Educational Research, LXI (1968), 300-303.

Howe, E. S. "Associative Structure of Quantifiers." Journal of Verbal Learning and Verbal Behavior, V (1966a), 156-162.

_____. "Probabilistic Adverbial Qualifications of Adjectives." Journal of Verbal Learning and Verbal Behavior, I (1962), 225-242.

_____. "Verb Tense, Negatives, and Other Determinants of the Intensity of Evaluative Meaning." Journal of Verbal Learning and Verbal Behavior, V (1966b), 147-155.

Husek, T. R., and Wittrock, M. C. "The Dimensions of Attitude Toward Teachers as Measured by the Semantic Differential." Journal of Educational Psychology, LIII (1963), 209-213.

Jöreskog, K. G. UMFLA: A Computer Program for Unrestricted Maximum Likelihood Factor Analysis. Research Memorandum RM-66-20. Princeton, New Jersey: Educational Testing Service, 1965.

Kaiser, H. F. "The Varimax Criterion for Analytic Rotation in Factor Analysis." Psychometrika, XXIII (1958), 187-200.

Kane, R. B. "Minimizing Order Effects in the Semantic Differential." Educational and Psychological Measurement, XXXI (1971), 137-144.

Kerlinger, F. N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, 1965.

Krathwohl, D. R.; Bloom, B. S.; and Masia, B. B. Taxonomy of Educational Objectives: A Classification of Educational Goals. Handbook Z: Affective Domain. New York: McKay, 1964.

Kubinec, C. M. "The Relative Efficacy of Various Dimensions of the Self-Concept in Predicting Academic Achievement." American Educational Research Journal, VII (1970), 321-336.

Kumata, H., and Schramm, W. "A Pilot Study of Cross-Cultural Methodology." Public Opinion Quarterly, XX (1956), 229-238.

Lawley, D. N., and Maxwell, A. E. Factor Analysis as a Statistical Method. London: Butterworth, 1963.

Maltz, H. E. "Ontogenic Change in the Meaning of Concepts as Measured by the Semantic Differential." Child Development, XXXIV (1963), 667-674.

McNeil, K. "Multivariate Relationships Between the Semantic Space of Various Subcultures and Various Personality Variables." Unpublished Doctoral dissertation, University of Texas (Austin), 1967.

Messick, S. J. "Metric Properties of the Semantic Differential." Educational and Psychological Measurement, XVII (1957), 200-206.

Michon, J. A. "An Application of Osgood's 'Semantic Differential' Technique." Acta Psychologica, XVII, (1960), 377-391.

Miron, M. S. "The Influence of Instruction Modification on Test-Retest Reliabilities of the Semantic Differential." Educational and Psychological Measurement, XXI (1961), 883-893.

_____. "What Is it That Is Being Differentiated by the Semantic Differential?" Journal of Abnormal and Social Psychology, XII (1969), 189-193.

Mitsos, S. B. "Personal Constructs and the Semantic Differential." Journal of Abnormal and Social Psychology, LXII (1961), 433-434.

Mordkoff, A. M. "An Empirical Test of the Functional Autonymy of Semantic Differential Scales." Journal of Verbal Learning and Verbal Behavior, II (1963), 504-508.

_____. "Functional vs. Nominal Autonymy in Semantic Differential Scales." Psychological Reports, XVI (1965), 691-692.

Morrison, D. F. Multivariate Statistical Methods. New York: McGraw-Hill, 1967.

Mosier, C. I. "The Psychometric Study of Meaning." Journal of Social Psychology, XIII (1941), 123-140.

Norman, W. T. "Stability Characteristics of the Semantic Differential." Journal of Psychology, LXXII (1959), 581-584.

Nowlis, W. T., and Nowlis, H. H. "The Description and the Analysis of Moods." Annals of the New York Academy of Science, LXV (1965), 345-355.

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1967.

Oles, H. J., and Bolvin, J. O. "The Reliability and Useability of a Semantic Differential Attitude Scale With Third Through Fifth Grade Students." A paper presented at the National Council on Measurement in Education Annual Meeting, Chicago, 1972.

Osgood, C. E.; Suci, G. T.; and Tannenbaum, P. H. The Measurement of Meaning. Urbana: University of Illinois Press, 1957.

Provus, M. "Evaluation of Ongoing Programs in the Public School System." Educational Evaluation: New Roles, New Means. Edited by R. Tyler. NSSE Yearbook, 1969, pp. 242-283.

Remmers, H. H. "Rating Methods in Research on Teaching." Handbook of Research on Teaching. Edited by N. Gage. New York: Rand-McNally, 1963.

Rentz, R. R.; Fears, E. B.; and White, W. F. "Personality Correlates of Group Structure: A Canonical Correlation Analysis." Journal of Psychology, LXX (1968), 163-167.

Rowan, T. C. "Some Developments in Multidimensional Scaling Applied to Semantic Relationships." Unpublished Doctoral dissertation, University of Illinois, 1954.

Schlosberg, H. "Three Dimensions of Emotion." Psychological Review, LXI (1954), 81-88.

Schmidt, W. H., and Scheifley, V. Computer Programs for Maximum Likelihood Factor Analysis. Center for Urban Affairs Research Report No. 2. East Lansing, Michigan: Michigan State University, October, 1971.

Scriven, M. "The Methodology of Evaluation." Perspectives in Curriculum Evaluation. Edited by R. Tyler, R. Gagné, and M. Scriven. New York: Rand-McNally, 1967.

Seigel, S. Nonparametric Methods for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Smith, R. G. "A Semantic Differential for Theater Concepts." Speech Monograph, XXVIII (1961), 1-8.

Snider, J. G., and Osgood, C. E. Semantic Differential Technique: A Sourcebook. Chicago: Aldine, 1969.

Spino, W. D. "Semantic Differential Patterns of Selected College Freshmen as a Basis for Academic Differentiation." Unpublished Doctoral dissertation, Indiana University, 1969.

Stake, R. E. "The Countenance of Educational Evaluation." Teachers' College Record, LXVIII (1967), 523-540.

Stiggins, R. J., and Byers, J. L. "Annual Report of the Evaluation of Education 200." East Lansing, Michigan: College of Education, Michigan State University, 1972. (Mimeograph in preparation.)

Stufflebeam, D. L. "Evaluation as Enlightenment for Decision Making." Columbus: Ohio State Evaluation Center, 1968. (Mimeographed.)

Suci, G. J. "A Comparison of Semantic Structures in American Southwest Culture Groups." Journal of Abnormal and Social Psychology, LXI (1960), 25-30.

Tanaka, Y.; Oyama, T.; and Osgood, C. E. "A Cross-Culture, Cross-Concept Study of the Generality of Semantic Spaces." Journal of Verbal Learning and Verbal Behavior, II (1963), 392-405.

Tannenbaum, P. H. "Attitudes Toward Source and Concept as a Factor in Attitude Change Through Communication." Unpublished Doctoral dissertation, University of Illinois, 1953.

Tucker, L. R. "An Interbattery Method of Factor Analysis." Psychometrika, XXIII (1958), 111-137.

Tyler, R. Education Evaluation: New Roles, New Means. National Society for the Study of Education Yearbook, 1969.

Walberg, H. J. "Dynamics of Self-Conception During Teacher Training." Unpublished Doctoral dissertation, University of Chicago, 1964.

Walberg, H. J.; Metzner, S.; Todd, R. M.; and Henry, P. M. "Effects of Tutoring and Practice Teaching on Self-Concept and Attitudes in Educational Studies." Journal of Teacher Education, XIX (1968), 283-291.

Wiley, D. E.; Schmidt, W. H.; and Bramble, W. J. "A Class of Covariance Structure Models." Journal of the American Statistical Association (in press).

Wittrock, M. C. "The Connotative Meaning of Concepts: Teachers and Children." California Journal of Educational Research, XV (1964), 60-67.

Wrigley, C., and Neuhans, J. O. "The Matching of Two Sets of Factors." American Psychologist, X (1955), 418-419.

Wundt, W. Grundriss der Psychologia. 7th rev. ed. Leipzig: Englemann, 1905.

APPENDICES

APPENDIX A

SUPPLEMENTAL GENERAL INFORMATION

1. Concept Definitions

2. Research Basis for Selection of Scales

DEFINITIONS OF CONCEPTS

IPL CONCEPTS

Myself as a Teacher--(Chosen on the basis of research by Walberg, 1957)
"Self-concept in teacher trainees as they imagine themselves in the role of teachers" (p. 84).

Nonverbal Behavior--(Stipulated under objective #3 for the small group experience in Ed. 200 Course Outline, p. 3)
"The student will be able to recognize and interpret through description and explanation, diverse modes of nonverbal behavior, i.e., hands, face, arms, etc."

Questioning and Listening Skills--(Ed. 200 Course Outline obj. #1 and #2, small group experience, pp. 2-3)
"The student will be able to not only restate what has been said but to relate the feelings and intended meaning of the speaker to the speaker's satisfaction."
"The student will be able to seek further information or clarification without cueing a particular response."

Myself
Self-concept independent of any specific role other than an individual as a person.

CARREL CONCEPTS (Definitions Taken From Ed. 200 Handbook, Winter, 1972)

Behavioral Objectives

"A behavioral objective is a specific statement of what the teacher expects the student to do after completing a prescribed unit of instruction. The critical attributes of a behavioral objective are: intended behavior must be observable, achievable, and relevant, . . . specifying the following: Terminal behavior, conditions for intended behavior [and] criteria. . . ." (p. 214).

Reinforcement

"Positive reinforcement is the presentation of a rewarding stimulus following a response; the presentation of the reward (stimulus) is made contingent upon the occurrence of a specific behavior (response). The result of the stimulus presentation is that it increases the probability that the response will be repeated" (p. 328).

"Negative reinforcement is the removal of a painful stimulus (aversive situation) that is contingent upon a response (behavior). The removal of pain increases the probability that the behavior which results in removal will be continued" (p. 352).

Respondent Learning

"Respondent learning is a change in behavior which results from the pairing of two stimuli. Initially the first stimulus elicits a particular behavior which the second stimulus does not elicit. Following a number of experiences in which the two stimuli occur together in time and space, the stimuli become so closely associated that the conditioned stimulus alone comes to elicit similar behavior to that originally elicited only the unconditioned stimulus" (p. 384).

Shaping

"Shaping is the selective reinforcement of successive approximations of desired terminal behavior" (p. 368).

The table below lists the scales chosen for the pilot test and an example of the researcher(s) who had previously determined a dimension placement for the scale. * indicates those 15 carried on into the pretest/posttest procedure.

Scale	Evaluative	Potency	Activity	Other (By Whom)
*Good-bad	Osgood			
*Important-unimportant	Osgood		McNeil	
*Pleasant-unpleasant	Rentz			
*Positive-negative	Osgood			
*Valuable-worthless	Hoover & Schutz			
*Fair-unfair	Rentz; Hartley			
*Enjoyable-unenjoyable				
*Weak-powerful	McNeil	Wittrock; Osgood		
*Severe-lenient		Osgood		Tenacity (Wittrock)
Rugged-delicate		Wittrock; McNeil		
Large-small		Coyne & Holtzman		
*Interesting-uninteresting	McNeil	Wittrock		Expressiveness (Wittrock)
Violent-gentle		Solomon		
Safe-dangerous				Security (McNeil)
Unique-commonplace				
*Active-passive			Osgood	Expressiveness (Wittrock)
Excitable-calm			Osgood	Stability (Wittrock)
*Tense-relaxed			Solomon	
*Sensitive-insensitive			Hartley	
*Moving-still	McNeil		Osgood	
Labored-easy			Osgood	
Fast-slow			Rentz	Tenacity (Wittrock)
Remote-intimate				

APPENDIX B

PILOT TEST SUPPLEMENT

1. Cover Letter
2. Instructions
3. Sample Response Sheet

MICHIGAN STATE UNIVERSITY EAST LANSING • MICHIGAN 48823

COLLEGE OF EDUCATION • ERICKSON HALL

TO

FROM: Dr. Judith E. Henderson
Dr. Joe L. Byers
Coordinators, Education 200

In our attempt to improve the student's experience in ED 200 we have constructed this instrument to assess the influence of the experience on the student. This instrument, like the Student Instructional Rating Report which you filled out in your last small group meeting, is designed to provide feedback so that instruction may be altered to optimize its role in assisting you in your preparation to become a professional educator.

In order to assist us in this endeavor and not consume too much of your class time, we ask you to complete this form outside of class and return it to the secretary in Room 238 Erickson Hall (ED 200 testing room) by Wednesday of final exam week (March 15). When you return the form to the secretary, be sure that she checks your name off on a list of students. Please do this at your earliest possible convenience so that your input can be sued in planning instruction for the Spring term.

INSTRUCTIONS

The purpose of this study is to measure the meanings of certain things to various people by having them judge them against a series of descriptive scales. As you respond, please make your judgements on the basis of what these things mean to you. On each page of this booklet, you will find different concepts to be judged and beneath each will be a set of scales. You are to rate the concepts on each of the scales.

Here is how you are to use the scales:

If you feel that the concept at the top of the page is very closely related to one end of the scale, you should place your response in the box at the right corresponding to the numbers 1 or 7, as follows:

fair: 1 : 2 : 3 : 4 : 5 : 6 : 7 :unfair.... [1] 2 3 4 5 6 7

or

fair: 1 : 2 : 3 : 4 : 5 : 6 : 7 :unfair.... 1 2 3 4 5 6 [7]

If you feel that the concept is quite closely related to one or the other end of the scale (but not extremely), you should respond as follows:

strong: 1 : 2 : 3 : 4 : 5 : 6 : 7 :weak..... 1 [2] 3 4 5 6 7

or

strong: 1 : 2 : 3 : 4 : 5 : 6 : 7 :weak..... 1 2 3 4 5 [6] 7

If the concept seems only slightly related to one side as opposed to the other side (but not really neutral), then you should respond as follows:

active: 1 : 2 : 3 : 4 : 5 : 6 : 7 :passive.. 1 2 [3] 4 5 6 7

or

active: 1 : 2 : 3 : 4 : 5 : 6 : 7 :passive.. 1 2 3 4 [5] 6 7

The direction toward which you respond, of course, depends upon which of the two ends of the scale seems most characteristic of the concept you're judging.

If you consider the concept to be neutral on the scale, both sides of the scale equally associated with the concept, or if the scale is completely irrelevant or unrelated to the concept, select number 4 as your response:

active: 1 : 2 : 3 : 4 : 5 : 6 : 7 :passive.. 1 2 3 [4] 5 6 7

IMPORTANT:

(1) Be sure you check every scale for every concept--do not omit any.

(2) Never put more than one response on a single scale.

Sometimes you may feel as though you've seen the same scale before in rating a given concept. This will not be the case, so do not look back and forth through the items. Do not try to remember how you checked similar items earlier. Make each item a separate and independent judgement. Work at a fairly high rate of speed. Do not worry or puzzle over individual scales. It is your first impression, the immediate "feelings" about the items, that we want. On the other hand, please do not be careless, because we want your true impressions.

MAKE ALL RESPONSES IN PENCIL, THANK YOU FOR YOUR COOPERATION

QUESTIONING SKILLS

	extremely	very	slightly	neutral	slightly	very	extremely	
violent	1	2	3	4	5	6	7	gentle.....
unpleasant	1	2	3	4	5	6	7	pleasant.....
lenient	1	2	3	4	5	6	7	severe.....
intimate	1	2	3	4	5	6	7	remote.....
excitable	1	2	3	4	5	6	7	calm.....
bad	1	2	3	4	5	6	7	good.....
unenjoyable	1	2	3	4	5	6	7	enjoyable.....
light	1	2	3	4	5	6	7	heavy.....
passive	1	2	3	4	5	6	7	active.....
still	1	2	3	4	5	6	7	moving.....
sharp	1	2	3	4	5	6	7	dull.....
interesting	1	2	3	4	5	6	7	uninteresting.....
positive	1	2	3	4	5	6	7	negative.....
sensitive	1	2	3	4	5	6	7	insensitive.....
unique	1	2	3	4	5	6	7	commonplace.....
rugged	1	2	3	4	5	6	7	delicate.....
fair	1	2	3	4	5	6	7	unfair.....
weak	1	2	3	4	5	6	7	strong.....
worthless	1	2	3	4	5	6	7	valuable.....
safe	1	2	3	4	5	6	7	dangerous.....
tense	1	2	3	4	5	6	7	relaxed.....
fast	1	2	3	4	5	6	7	slow.....
unimportant	1	2	3	4	5	6	7	important.....
large	1	2	3	4	5	6	7	small.....
labored	1	2	3	4	5	6	7	easy.....

APPENDIX C

PRE/POST ADMINISTRATION SUPPLEMENT

1. Cover Letter
2. Sample Response Sheet

Your name_____

Student Number_____

TO: All students enrolled in Education 200, Spring Term, 1972

SUBJECT: Evaluation of Instruction

As you may know, one of the goals of Education 200 is to continually evaluate and revise our instruction so that it will supply maximum service to the student as he or she prepares to become a professional educator. As a student enrolled in Education 200, you must play a critical role in this endeavor by providing feedback about the instruction from time to time during the term.

For example, as you proceed through the course, you will be exposed to numerous instructional concepts, tools and procedures which will, hopefully, be of assistance to you when you become a practicing teacher. It is important, therefore, that our instruction be effective in teaching the meaning and use of these tools and concepts. In order to assess the worth of our teaching procedures, periodic measurements will be made in the form of tests to determine how much you have learned. That is, there will be an attempt made to determine whether or not you know how and when to use the instructional concepts and tools presented in the course. This is one method of determining if our instruction is as effective as it might be.

However, there is another influence which the instruction has on you as a student which is as important as how much you learn, and that is how you react emotionally to the concepts and tools as they are presented to you. This effect of instruction is almost never measured in our educational endeavors, but we feel that it could serve the useful purpose of providing information as we plan instruction for the future.

The survey which you are asked to respond to here is an initial attempt to measure some of these non-cognitive outcomes of our instruction by asking you to describe your reactions to the concepts. It is an initial attempt because we plan to use your responses to adjust and refine these measurement techniques in order to develop a systematic means of measuring your reactions. For these reasons, we require your assistance. Please read the instructions on the next page and respond accordingly. If you have any questions or need further clarification as to the intents of this survey, please do not hesitate to seek further assistance in room 238 of Erickson Hall, or phone Rick Stiggins or Rob Brann at 353-8765. Please return the completed form to Room 238 Erickson Hall by Tuesday, April 11.

Judith E. Henderson and Joe L. Byers
Co-Coordinators of Education 200

Richard J. Stiggins
Education 200 Evaluation

Please enter and code your student number.

STUDENT NO.						
		0	0	0	0	0
		1	1	1	1	1
		2	2	2	2	2
		3	3	3	3	3
		4	4	4	4	4
		5	5	5	5	5
		6	6	6	6	6
		7	7	7	7	7
		8	8	8	8	8
		9	9	9	9	9

CODE NO.						
		0	0	0	0	0
		1	1	1	1	1
		2	2	2	2	2
		3	3	3	3	3
		4	4	4	4	4
		5	5	5	5	5
		6	6	6	6	6
		7	7	7	7	7
		8	8	8	8	8
		9	9	9	9	9

BEHAVIORAL OBJECTIVES

	EXTREMELY	VERY	SLIGHTLY	NEUTRAL	SLIGHTLY	VERY	EXTREMELY	
worthless	1	2	3	4	5	6	7	**valuable**
severe	1	2	3	4	5	6	7	**lenient**
unfair	1	2	3	4	5	6	7	**fair**
good	1	2	3	4	5	6	7	**bad**
tense	1	2	3	4	5	6	7	**relaxed**
weak	1	2	3	4	5	6	7	**powerful**
negative	1	2	3	4	5	6	7	**positive**
violent	1	2	3	4	5	6	7	**gentle**
important	1	2	3	4	5	6	7	**unimportant**
active	1	2	3	4	5	6	7	**passive**
insensitive	1	2	3	4	5	6	7	**sensitive**
uninteresting	1	2	3	4	5	6	7	**interesting**
pleasant	1	2	3	4	5	6	7	**unpleasant**
still	1	2	3	4	5	6	7	**moving**
enjoyable	1	2	3	4	5	6	7	**unenjoyable**
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	

Please enter and code your student number.

STUDENT NO.						
		0	1	2	3	4
		0	1	2	3	4
		0	1	2	3	4
		0	1	2	3	4
		0	1	2	3	4
		0	1	2	3	4
		5	6	7	8	9
		5	6	7	8	9
		5	6	7	8	9
		5	6	7	8	9
		5	6	7	8	9
		5	6	7	8	9

CODE NO						
		0	1	2	3	4
		0	1	2	3	4
		0	1	2	3	4
		0	1	2	3	4
		0	1	2	3	4
		0	1	2	3	4
		5	6	7	8	9
		5	6	7	8	9
		5	6	7	8	9
		5	6	7	8	9
		5	6	7	8	9
		5	6	7	8	9

BEHAVIORAL OBJECTIVES

	EXTREMELY	VERY	SLIGHTLY	NEUTRAL	SLIGHTLY	VERY	EXTREMELY	
worthless	1	2	3	4	5	6	7	**valuable**
severe	1	2	3	4	5	6	7	**lenient**
unfair	1	2	3	4	5	6	7	**fair**
good	1	2	3	4	5	6	7	**bad**
tense	1	2	3	4	5	6	7	**relaxed**
weak	1	2	3	4	5	6	7	**powerful**
negative	1	2	3	4	5	6	7	**positive**
violent	1	2	3	4	5	6	7	**gentle**
important	1	2	3	4	5	6	7	**unimportant**
active	1	2	3	4	5	6	7	**passive**
insensitive	1	2	3	4	5	6	7	**sensitive**
uninteresting	1	2	3	4	5	6	7	**interesting**
pleasant	1	2	3	4	5	6	7	**unpleasant**
still	1	2	3	4	5	6	7	**moving**
enjoyable	1	2	3	4	5	6	7	**unenjoyable**
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	
	1	2	3	4	5	6	7	

APPENDIX D

RESULTS SUPPLEMENT

1. Factor pattern matrices for each concept (Note: all matrices report varimax rotated solutions)

2. Raw scale means and standard deviations for each concept

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Myself (pretest)

4-Factor Solution $X^2=65$, $p<.00$, df=51

	1	2	3	4
worthless-valuable	-.171	.555	.012	.071
severe-lenient	.018	-.164	.655	.032
unfair-fair	-.247	.361	.626	-.051
good-bad	.235	-.562	-.166	-.059
tense-relaxed	-.095	.386	.090	.251
weak-powerful	.041	.544	-.206	.368
negative-positive	-.269	.576	.186	.110
violent-gentle	-.355	.177	.526	.034
important-unimportant	.287	-.460	.030	-.146
active-passive	.203	-.419	.160	-.538
insensitive-sensitive	-.430	.171	.232	.024
uninteresting-interesting	-.400	.416	.072	.277
pleasant-unpleasant	.884	-.194	-.098	-.058
still-moving	-.110	.134	.105	.778
enjoyable-unenjoyable	.718	-.272	-.096	-.241

5-Factor Solution: $X^2 = 63.44$, $p < .01$, df=41

	1	2	3	4	5
worthless-valuable	-.088	.019	-.053	.505	.282
severe-lenient	.024	.053	.654	-.147	.042
unfair-fair	-.205	-.089	.620	.316	.266
good-bad	.168	-.008	-.094	-.488	-.336
tense-relaxed	-.115	.132	.162	.501	-.074
weak-powerful	-.061	.230	-.175	.617	-.100
negative-positive	-.129	.033	.181	.599	.268
violent-gentle	-.117	.065	.426	.095	.575
important-unimportant	.167	-.100	.105	-.434	-.323
active-passive	.210	-.410	.165	-.505	-.072
insensitive-sensitive	-.216	.041	.099	.073	.575
uninteresting-interesting	-.356	.208	.014	.405	.299
pleasant-unpleasant	.592	-.031	-.079	-.206	-.475
still-moving	-.097	.884	.080	.221	.085
enjoyable-unenjoyable	.928	-.150	-.090	-.246	-.219

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Myself (posttest)

4-Factor Solution: $X^2=61$, $p<.00$, $df=51$

	1	2	3	4
worthless-valuable	-.602	-.270	.165	-.028
severe-lenient	-.038	.054	.080	.583
unfair-fair	-.627	-.190	.012	.286
good-bad	.630	.192	-.168	-.131
tense-relaxed	-.186	-.102	.756	.224
weak-powerful	-.381	-.375	.400	-.233
negative-positive	-.506	-.285	.269	.196
violent-gentle	-.311	-.126	.131	.587
important-unimportant	.721	.216	-.138	-.024
active-passive	.245	.784	-.167	-.048
insensitive-sensitive	-.446	-.164	-.155	.360
uninteresting-interesting	-.600	-.248	.157	.178
pleasant-unpleasant	.713	.104	-.106	-.264
still-moving	-.286	-.662	.014	.048
enjoyable-unenjoyable	.743	.197	-.142	-.092

6-Factor Solution: $X^2 = 43.59$, $p<.07$, $df=32$

	1	2	3	4	5	6
worthless-valuable	-.153	.219	-.047	-.431	.269	.405
severe-lenient	-.016	-.064	.509	.041	.048	.128
unfair-fair	-.144	.158	.175	-.225	.126	.678
good-bad	.200	-.171	-.136	.394	-.258	-.420
tense-relaxed	-.093	.082	.236	-.093	.655	.056
weak-powerful	-.130	.364	-.191	-.227	.503	.159
negative-positive	-.077	.237	.203	-.287	.390	.379
violent-gentle	-.117	.147	.724	-.175	.129	.174
important-unimportant	.257	-.196	-.086	.774	-.158	-.290
active-passive	.046	-.684	-.041	.195	-.255	-.170
insensitive-sensitive	-.034	.155	.273	-.139	-.081	.574
uninteresting-interesting	-.299	.223	.083	-.086	.304	.574
pleasant-unpleasant	.378	-.115	-.264	.286	-.203	-.482
still-moving	-.158	.765	.034	-.083	.051	.203
enjoyable-unenjoyable	.840	-.222	-.120	.329	-.175	-.304

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: **Myself as a teacher (pretest)**

4-Factor Solution and Solution of Adequate Statistical Fit Are Identical

	1	2	3	4
worthless-valuable	-.414	.153	.155	-.448
severe-lenient	-.028	.468	.038	.085
unfair-fair	-.100	.344	.141	-.228
good-bad	.456	-.069	-.184	.302
tense-relaxed	-.129	.107	.746	-.181
weak-powerful	-.453	-.079	.391	-.017
negative-positive	-.459	.312	.179	-.203
violent-gentle	.005	.626	-.016	-.204
important-unimportant	.291	.058	-.070	.410
active-passive	.618	.079	-.045	.210
insensitive-sensitive	-.095	.362	-.082	-.349
uninteresting-interesting	-.406	.147	.101	-.437
pleasant-unpleasant	.157	-.192	-.180	.736
still-moving	-.486	.134	.012	-.139
enjoyable-unenjoyable	.380	-.142	-.065	.500

X^2=66.16, $p < .08$, df=51

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Myself as a teacher (posttest)

4-Factor Solution : X^2=50, p<.00, df=51

	1	2	3	4
worthless-valuable	.070	-.436	.129	.174
severe-lenient	.125	-.040	.340	-.105
unfair-fair	.056	-.190	.535	.144
good-bad	-.028	.670	-.192	-.031
tense-relaxed	.086	-.171	.207	.460
weak-powerful	.079	-.164	-.095	.606
negative-positive	.019	-.430	.392	.384
violent-gentle	.141	-.301	.454	.008
important-unimportant	-.108	.712	-.117	-.214
active-passive	-.234	.600	-.266	-.291
insensitive-sensitive	.026	-.147	.638	.129
uninteresting-interesting	.073	-.423	.438	.354
pleasant-unpleasant	-.053	.557	-.443	-.232
still-moving	.920	-.191	.274	.205
enjoyable-unenjoyable	-.223	.386	-.513	-.252

3-Factor Solution: X^2 = 87.76 p < .06 , df=63

	1	2	3
worthless-valuable	-.433	.135	.190
severe-lenient	-.034	.358	-.061
unfair-fair	-.201	.526	.125
good-bad	.680	-.182	-.020
tense-relaxed	-.184	.221	.429
weak-powerful	-.146	-.096	.679
negative-positive	-.440	.381	.335
violent-gentle	-.305	.463	.032
important-unimportant	.710	-.130	-.233
active-passive	.597	-.302	-.338
insensitive-sensitive	-.143	.622	.133
uninteresting-interesting	-.429	.444	.328
pleasant-unpleasant	.557	-.445	-.214
still-moving	-.261	.372	.304
enjoyable-unenjoyable	.387	-.555	-.275

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Nonverbal behavior (pretest)

4-Factor Solution: X^2=60, p<.00, df=51

	1	2	3	4
worthless-valuable	.144	-.770	-.206	.144
severe-lenient	.116	-.083	-.467	-.179
unfair-fair	.220	-.142	-.532	.177
good-bad	-.059	.587	.337	-.171
tense-relaxed	.017	-.246	-.497	.162
weak-powerful	.116	-.471	-.073	.425
negative-positive	.141	-.401	-.596	.199
violent-gentle	.186	-.051	-.610	-.025
important-unimportant	-.045	.776	.173	-.222
active-passive	-.088	.398	.023	-.615
insensitive-sensitive	.847	-.292	-.381	.228
uninteresting-interesting	.251	-.586	-.262	.353
pleasant-unpleasant	-.025	.185	.740	-.327
still-moving	.107	-.215	-.229	.626
enjoyable-unenjoyable	.034	.263	.758	-.359

6-Factor Solution: X^2 = 49.37 p < .03 df= 32

	1	2	3	4	5	6
worthless-valuable	.934	.164	.199	.217	-.095	-.072
severe-lenient	.071	.069	.514	-.131	-.079	-.025
unfair-fair	.119	.032	.617	.268	-.112	-.027
good-bad	-.432	-.101	-.289	-.259	.093	.431
tense-relaxed	.112	.042	.374	.159	-.237	-.393
weak-powerful	.321	.084	.005	.497	-.070	-.367
negative-positive	.229	.167	.541	.230	-.213	-.431
violent-gentle	.005	.018	.597	.004	-.188	-.139
important-unimportant	-.544	-.302	-.086	-.288	.148	[illegible]
active-passive	-.216	-.188	-.001	-.646	[illegible]	[illegible]
insensitive-sensitive	.236	.277	.563	.368	[illegible]	[illegible]
uninteresting-interesting	.349	.219	.219	.354	[illegible]	[illegible]
pleasant-unpleasant	-.174	[illegible]	-.537	-.264	[illegible]	[illegible]
still-moving	.106	.109	.173	.363	[illegible]	[illegible]
enjoyable-unenjoyable	[illegible]	[illegible]	-.474	[illegible]	[illegible]	[illegible]

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Nonverbal behavior (posttest)

4-Factor Solution and Solution of Adequate Statistical Fit Are Identical

	1	2	3	4
worthless-valuable	.163	-.168	.715	-.128
severe-lenient	.451	.017	.100	-.181
unfair-fair	.352	-.191	.249	-.250
good-bad	-.158	.085	-.599	.305
tense-relaxed	.564	-.122	.136	-.085
weak-powerful	.073	-.234	.581	-.003
negative-positive	.489	-.333	.315	-.266
violent-gentle	.832	-.106	.065	-.098
important-unimportant	-.154	.190	-.709	-.044
active-passive	-.101	.833	-.320	.074
insensitive-sensitive	.353	-.204	.448	-.229
uninteresting-interesting	.114	-.141	.641	-.315
pleasant-unpleasant	-.509	.175	-.168	.590
still-moving	.186	-.627	.259	-.208
enjoyable-unenjoyable	-.476	.222	-.208	.699

$X^2 = 79.65$ $p < .01$ df=51

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Questioning and listening skills (pretest)

4-Factor Solution and Solution of Adequate Statistical Fit Are Identical

	1	2	3	4
worthless-valuable	-.114	-.520	.164	.304
severe-lenient	-.051	.021	.430	-.131
unfair-fair	-.123	-.282	.425	.338
good-bad	.140	.710	-.073	-.299
tense-relaxed	-.312	-.141	.412	.239
weak-powerful	-.070	-.453	-.012	.612
negative-positive	-.137	-.475	.237	.603
violent-gentle	-.113	-.109	.641	.156
important-unimportant	.129	.786	-.107	-.214
active-passive	.137	.249	.003	-.545
insensitive-sensitive	-.037	-.392	.433	.402
uninteresting-interesting	-.226	-.459	.311	.474
pleasant-unpleasant	.351	.268	-.453	-.378
still-moving	-.173	-.159	.104	.507
enjoyable-unenjoyable	.901	.213	-.289	-.242

X^2=64.92 $p < .11$ df=51

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Questioning and listening skills (posttest)

4-Factor Solution and Solution of Adequate Statistical Fit Are Identical

	1	2	3	4
worthless-valuable	-.217	-.745	.116	.043
severe-lenient	-.066	.004	.016	.525
unfair-fair	-.335	-.401	.305	.308
good-bad	.174	.769	-.087	-.148
tense-relaxed	-.150	-.178	.160	.479
weak-powerful	-.183	-.599	.321	.052
negative-positive	-.279	-.500	.249	.186
violent-gentle	-.269	-.068	.239	.688
important-unimportant	.154	.641	-.180	-.036
active-passive	.119	.271	-.583	-.114
insensitive-sensitive	-.2[illegible]	-.400	.354	.346
uninteresting-interesting	-.54[illegible]	-.330	.250	.105
pleasant-unpleasant	.7[illegible]	.268	-.146	-.339
still-moving	-.23[illegible]	-.174	.672	.231
enjoyable-unenjoyable	.76[illegible]	.285	-.186	-.275

X^2=61.35 p < .15 df=51

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Behavioral Objectives (pretest)

4-Factor Solution: X^2=60, p<.00, df=51

	1	2	3	4
worthless-valuable	-.204	-.804	.118	-.079
severe-lenient	-.060	-.047	.701	-.030
unfair-fair	-.243	-.535	.452	-.156
good-bad	.153	.763	-.143	.117
tense-relaxed	-.219	-.093	.559	-.180
weak-powerful	-.132	-.559	-.017	-.357
negative-positive	-.230	-.688	.278	-.163
violent-gentle	-.186	-.188	.647	.125
important-unimportant	.142	.829	-.124	.181
active-passive	.166	.268	-.023	.712
insensitive-sensitive	-.230	-.464	.443	-.061
uninteresting-interesting	-.436	-.602	.237	-.178
pleasant-unpleasant	.687	.314	-.432	.214
still-moving	-.452	-.213	.150	-.377
enjoyable-unenjoyable	.871	.295	-.263	.097

5-Factor Solution: X^2 = 60.43 p<.03 df=41

	1	2	3	4	5
worthless-valuable	.139	-.206	-.704	-.065	.087
severe-lenient	.075	-.072	-.057	.001	.746
unfair-fair	.107	-.265	-.544	-.116	.436
good-bad	-.047	.156	.777	.092	-.142
tense-relaxed	.039	-.242	-.113	-.142	.574
weak-powerful	.066	-.142	-.545	-.322	-.030
negative-positive	.185	-.245	-.681	-.137	.226
violent-gentle	.328	-.200	-.154	.112	.545
important-unimportant	-.148	.146	.821	.162	-.100
active-passive	-.042	.176	.266	.782	-.028
insensitive-sensitive	.855	-.193	-.362	-.098	.301
uninteresting-interesting	.187	-.449	-.589	-.148	.191
pleasant-unpleasant	-.157	.713	.312	.173	-.385
still-moving	.025	-.467	-.224	-.344	.147
enjoyable-unenjoyable	-.138	.564	.290	.074	-.219

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Behavioral Objectives (posttest)

4-Factor Solution: X^2=50, p<.00, df=51

	1	2	3	4
worthless-valuable	-.674	.049	.217	.329
severe-lenient	-.087	-.026	.541	-.016
unfair-fair	-.594	-.173	.436	.303
good-bad	.726	.342	-.099	-.056
tense-relaxed	-.181	-.203	.660	.194
weak-powerful	-.540	-.019	.092	.544
negative-positive	-.656	-.217	.349	.310
violent-gentle	-.104	-.316	.638	.365
important-unimportant	.715	.232	-.044	-.146
active-passive	.319	.354	-.011	-.454
insensitive-sensitive	-.266	-.268	.469	.445
uninteresting-interesting	-.498	-.311	.202	.485
pleasant-unpleasant	.146	.831	-.280	-.178
still-moving	-.212	-.239	.310	.644
enjoyable-unenjoyable	.293	.783	-.197	-.202

5-Factor Solution: X^2 = 65.79 p<.01 df=41

	1	2	3	4	5
worthless-valuable	-.514	.034	.236	.179	.503
severe-lenient	-.034	-.049	.561	-.058	.077
unfair-fair	-.457	-.183	.447	.202	.421
good-bad	.645	.332	-.118	-.033	-.304
tense-relaxed	-.168	-.182	.661	.242	.081
weak-powerful	-.253	-.076	.084	.257	.765
negative-positive	-.448	-.258	.377	.095	.562
violent-gentle	-.038	-.319	.620	.337	.197
important-unimportant	.797	.193	-.053	-.216	-.195
active-passive	.284	.324	.013	-.486	-.256
insensitive-sensitive	-.233	-.238	.471	.468	.214
uninteresting-interesting	-.316	-.338	.218	.304	.519
pleasant-unpleasant	.146	.827	-.271	-.197	-.067
still-moving	-.109	-.228	.336	.585	.355
enjoyable-unenjoyable	.265	.781	-.190	-.194	-.149

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Reinforcement (pretest)

4-Factor Solution: X^2=85, p<.00, df=51

	1	2	3	4
worthless-valuable	-.667	.019	.074	.261
severe-lenient	.106	-.388	.063	-.039
unfair-fair	-.390	-.283	.513	.357
good-bad	.745	.141	-.350	-.150
tense-relaxed	-.205	-.544	.048	.121
weak-powerful	-.501	-.063	.030	.319
negative-positive	-.622	-.317	.236	.163
violent-gentle	-.119	-.597	.238	.151
important-unimportant	.725	.154	.043	-.217
active-passive	.412	.048	-.009	-.449
insensitive-sensitive	-.169	-.315	.311	.521
uninteresting-interesting	-.211	-.222	.114	.633
pleasant-unpleasant	.230	.672	.002	-.385
still-moving	-.267	-.104	.054	.464
enjoyable-unenjoyable	.260	.692	.143	-.409

5-Factor Solution: X^2 = 47.94 p<.21 df=41

	1	2	3	4	5
worthless-valuable	-.082	-.430	-.068	.526	.183
severe-lenient	-.022	.054	.484	-.052	.086
unfair-fair	-.069	-.498	.233	.170	.479
good-bad	.086	.822	-.012	-.315	-.202
tense-relaxed	-.209	-.118	.609	.249	.087
weak-powerful	-.010	-.178	.133	.661	.188
negative-positive	-.165	-.563	.213	.338	.212
violent-gentle	-.184	-.232	.574	.004	.209
important-unimportant	.205	.444	-.057	-.546	-.120
active-passive	.172	.193	.039	-.429	-.345
insensitive-sensitive	-.145	-.239	.215	.107	.002
uninteresting-interesting	-.177	-.095	.117	.297	.009
pleasant-unpleasant	.475	.210	-.394	-.149	-.016
still-moving	-.146	-.130	.039	.025	.005
enjoyable-unenjoyable	.371	.140	-.322	-.179	[illegible]

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Reinforcement (posttest)

4-Factor Solution: X^2=89, p<.00, df= 51

	1	2	3	4
worthless-valuable	-.075	-.689	.382	.165
severe-lenient	-.014	-.333	.431	-.182
unfair-fair	-.140	-.516	.373	.293
good-bad	.251	.683	-.151	-.140
tense-relaxed	-.225	-.236	.595	.111
weak-powerful	-.151	-.352	-.067	.436
negative-positive	-.235	-.462	.281	.361
violent-gentle	-.212	-.052	.652	.305
important-unimportant	.167	.732	-.380	-.152
active-passive	.133	.272	.055	-.631
insensitive-sensitive	-.215	-.352	.329	.450
uninteresting-interesting	-.421	-.298	.208	.331
pleasant-unpleasant	.490	.216	-.329	-.365
still-moving	-.184	-.085	.109	.649
enjoyable-unenjoyable	.899	.292	-.225	-.237

5-Factor Solution: X^2 = 67.45 p<.01 df=41

	1	2	3	4	5
worthless-valuable	-.051	-.661	.077	-.156	.182
severe-lenient	-.009	-.037	.428	.186	-.009
unfair-fair	-.119	-.510	.437	-.326	.002
good-bad	.242	.721	-.181	.178	.082
tense-relaxed	-.198	-.226	.592	-.095	.136
weak-powerful	-.137	-.341	-.056	-.444	.073
negative-positive	-.209	-.450	.299	-.368	.105
violent-gentle	-.182	-.040	.652	-.282	.187
important-unimportant	.135	.730	-.069	.124	-.236
active-passive	.110	.257	.042	.652	-.085
insensitive-sensitive	-.153	-.318	.316	-.397	.423
uninteresting-interesting	-.359	-.261	.180	-.260	.517
pleasant-unpleasant	.464	.198	-.347	.371	-.142
still-moving	-.158	-.080	.219	-.628	.130
enjoyable-unenjoyable	.874	.288	-.244	.246	-.182

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Respondant Learning (pretest)

4-Factor Solution: X^2=154, p<.00, df=51

	1	2	3	4
worthless-valuable	.068	.814	-.242	-.215
severe-lenient	.981	.068	-.169	-.071
unfair-fair	.298	.501	-.462	-.287
good-bad	-.122	-.755	.402	.297
tense-relaxed	.219	.127	-.593	-.222
weak-powerful	-.087	.469	-.297	-.478
negative-positive	.091	.588	-.497	-.351
violent-gentle	.046	.149	-.416	-.140
important-unimportant	-.040	-.776	.377	.311
active-passive	-.037	-.386	.236	.640
insensitive-sensitive	.075	.288	-.507	-.195
uninteresting-interesting	.039	.448	-.576	-.401
pleasant-unpleasant	-.095	-.320	.797	.174
still-moving	.174	.217	-.301	-.632
enjoyable-unenjoyable	-.042	-.408	.772	.215

7-Factor Solution: X^2 = 41.16 p< .01 df=13

	1	2	3	4	5	6	7
worthless-valuable	.065	.339	-.164	.130	-.219	.084	-.093
severe-lenient	.983	.079	-.092	.058	-.084	.094	.005
unfair-fair	.301	.520	-.280	.219	-.241	.305	.150
good-bad	-.122	-.768	.278	-.159	.296	-.211	-.024
tense-relaxed	.221	.147	-.446	-.006	-.216	.547	.059
weak-powerful	-.087	.492	-.171	.043	-.426	.237	.236
negative-positive	.100	.612	-.325	.218	-.290	.237	.379
violent-gentle	.032	.153	-.121	.221	-.076	.654	.026
important-unimportant	-.045	-.782	.284	-.170	.303	-.109	-.125
active-passive	-.027	-.381	.136	-.097	.733	-.124	.046
insensitive-sensitive	.073	.245	-.213	.795	-.162	.239	.037
uninteresting-interesting	.051	.463	-.420	.297	-.389	.194	.198
pleasant-unpleasant	-.116	-.336	.674	-.315	.216	-.244	.019
still-moving	.176	.241	-.291	.139	-.612	.089	.094
enjoyable-unenjoyable	-.064	-.430	.745	-.153	.237	-.216	-.100

PRETEST AND POSTTEST ROTATED FACTOR PATTERN

MATRICES FOR EACH CONCEPT, 4 FACTOR

SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Respondant Learning (posttest)

4-Factor Solution: X^2=86, p<.00, df=51

	1	2	3	4
worthless-valuable	-.699	.123	-.179	-.175
severe-lenient	-.094	.441	-.156	.010
unfair-fair	-.489	.400	-.243	-.154
good-bad	.728	-.254	.168	.153
tense-relaxed	-.184	.641	-.188	-.564
weak-powerful	-.506	.276	-.064	-.425
negative-positive	-.624	.276	-.283	-.230
violent-gentle	-.093	.809	-.185	-.204
important-unimportant	.723	-.368	.125	.412
active-passive	.335	-.310	.147	.697
insensitive-sensitive	-.339	.368	-.259	-.347
uninteresting-interesting	-.439	.214	-.419	-.226
pleasant-unpleasant	.223	-.427	.683	.248
still-moving	-.233	.222	-.274	-.580
enjoyable-unenjoyable	.276	-.356	.750	.207

5-Factor Solution: X^2 = 60.95 p<.02 df=41

	1	2	3	4	5
worthless-valuable	-.618	.114	-.197	-.230	-.231
severe-lenient	-.111	.425	-.119	-.006	.019
unfair-fair	-.548	.351	-.253	.065	-.205
good-bad	.739	-.237	.156	.100	.208
tense-relaxed	-.216	.578	-.230	.012	-.078
weak-powerful	-.463	.059	-.373	-.056	-.477
negative-positive	-.596	.243	-.286	-.055	-.289
violent-gentle	-.072	.894	-.165	-.078	-.208
important-unimportant	.605	-.351	.146	.674	.395
active-passive	.259	-.310	.158	.165	.681
insensitive-sensitive	-.283	.349	-.291	-.136	-.355
uninteresting-interesting	-.390	.190	-.443	-.162	-.240
pleasant-unpleasant	.198	-.400	.730	.076	.241
still-moving	-.201	.202	-.271	-.015	-.631
enjoyable-unenjoyable	.274	-.355	.746	.013	.235

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Shaping (pretest)

4-Factor Solution: X^2=90, p<.00, df=51

	1	2	3	4
worthless-valuable	-.255	-.363	.036	-.726
severe-lenient	-.075	.081	.663	-.010
unfair-fair	-.269	-.085	.472	-.580
good-bad	.227	.240	-.160	.794
tense-relaxed	-.236	-.020	.743	-.297
weak-powerful	.028	-.691	-.015	-.197
negative-positive	-.370	-.239	.253	-.694
violent-gentle	-.461	-.239	.515	-.231
important-unimportant	.272	.421	-.114	.749
active-passive	.202	.762	-.004	.279
insensitive-sensitive	-.385	-.210	.355	-.418
uninteresting-interesting	-.485	-.440	.125	-.438
pleasant-unpleasant	.545	.160	-.257	.332
still-moving	-.237	-.640	-.049	-.161
enjoyable-unenjoyable	.741	.157	-.291	.389

6-Factor Solution: X^2 = 44.18 p<.07 df=32

	1	2	3	4	5	6
worthless-valuable	-.206	.169	-.748	.105	-.353	.343
severe-lenient	-.077	.047	-.012	.817	.080	.112
unfair-fair	-.294	.198	-.585	.384	-.089	-.184
good-bad	.245	-.115	.788	-.095	.249	.089
tense-relaxed	-.276	.146	-.328	.636	-.018	-.227
weak-powerful	.040	-.026	-.209	-.005	-.696	.006
negative-positive	-.380	.140	-.699	.175	-.296	-.113
violent-gentle	-.463	.247	-.237	.433	-.224	-.068
important-unimportant	.266	-.176	.725	-.073	.420	-.067
active-passive	.223	-.049	.261	-.067	.789	.062
insensitive-sensitive	-.309	.843	-.332	.237	-.166	.003
uninteresting-interesting	-.443	.225	-.423	.087	-.432	.082
pleasant-unpleasant	.901	-.136	.314	-.202	.154	-.077
still-moving	-.208	.208	-.126	-.083	-.634	.048
enjoyable-unenjoyable	.718	-.197	.398	-.198	.160	.096

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Shaping (posttest)

4-Factor Solution and Solution of Adequate Statistical Fit Are Identical

	1	2	3	4
worthless-valuable	-.148	-.799	.106	-.238
severe-lenient	-.051	-.093	.532	.010
unfair-fair	-.139	-.614	.434	-.251
good-bad	.152	.768	-.271	.216
tense-relaxed	-.175	-.177	.625	-.130
weak-powerful	-.133	-.370	.016	-.477
negative-positive	-.202	-.602	.260	-.381
violent-gentle	-.195	-.168	.654	-.336
important-unimportant	.186	.806	-.151	.279
active-passive	.165	.315	-.165	.669
insensitive-sensitive	-.201	-.398	.335	-.376
uninteresting-interesting	-.448	-.327	.047	-.461
pleasant-unpleasant	.565	.293	-.463	.382
still-moving	-.200	-.198	.218	-.629
enjoyable-unenjoyable	.813	.238	-.388	.288

$X^2=78.49$ $p < .01$ DF=51

PRETEST AND POSTTEST ROTATED FACTOR PATTERN MATRICES FOR EACH CONCEPT: 4 FACTOR SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Physician (pretest)

4-Factor Solution: X^2=55, p<.00, df=51

	1	2	3	4
worthless-valuable	-.082	.831	.165	.205
severe-lenient	.147	-.115	.236	-.273
unfair-fair	-.211	.279	.584	.001
good-bad	.140	-.662	-.357	-.146
tense-relaxed	-.211	.256	.517	.006
weak-powerful	-.052	.527	.300	.369
negative-positive	-.140	.392	.547	.251
violent-gentle	-.079	.237	.745	.135
important-unimportant	.080	-.736	-.258	-.230
active-passive	.046	-.474	-.265	-.449
insensitive-sensitive	-.124	.197	.683	.229
uninteresting-interesting	-.356	.312	.332	.428
pleasant-unpleasant	.347	-.220	-.670	-.116
still-moving	-.112	.202	.174	.678
enjoyable-unenjoyable	.906	-.074	-.371	-.189

5-Factor Solution: X^2 =55.73, p<.08, df=41

	1	2	3	4	5
worthless-valuable	-.071	-.212	.215	.753	.229
severe-lenient	.060	-.107	.215	-.190	-.230
unfair-fair	-.194	-.175	.593	.201	.032
good-bad	.127	.776	-.335	-.500	-.141
tense-relaxed	-.196	-.085	.530	.197	.047
weak-powerful	-.051	-.090	.314	.509	.383
negative-positive	-.126	-.265	.543	.286	.292
violent-gentle	-.056	-.009	.792	.276	.142
important-unimportant	.080	.126	-.305	-.752	-.214
active-passive	.046	.198	-.259	-.422	-.466
insensitive-sensitive	-.111	-.148	.669	.115	.274
uninteresting-interesting	-.355	-.130	.326	.272	.443
pleasant-unpleasant	.334	.116	-.676	-.158	-.146
still-moving	-.120	-.056	.154	.188	.681
enjoyable-unenjoyable	.901	.062	-.383	-.043	-.191

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: **Physician (posttest)**

4-Factor Solution: X^2=53, p<.00, df=51

	1	2	3	4
worthless-valuable	-.075	-.759	.024	.285
severe-lenient	-.037	.067	.320	-.053
unfair-fair	-.080	-.363	.511	.290
good-bad	.067	.715	-.269	-.226
tense-relaxed	-.206	-.140	.512	.199
weak-powerful	-.086	-.434	.122	.631
negative-positive	-.075	-.558	.364	.347
violent-gentle	-.089	-.375	.651	.401
important-unimportant	.056	.825	-.113	-.308
active-passive	.072	.367	-.123	-.682
insensitive-sensitive	-.069	-.333	.569	.257
uninteresting-interesting	-.185	-.354	.316	.583
pleasant-unpleasant	.300	.316	-.633	-.372
still-moving	-.105	-.189	.179	.754
enjoyable-unenjoyable	.874	.098	-.431	-.201

3-Factor Solution: X^2 = 78.13 p< .09 df=63

	1	2	3
worthless-valuable	-.752	-.042	.290
severe-lenient	.057	-.328	-.069
unfair-fair	-.385	-.493	.269
good-bad	.726	.266	-.213
tense-relaxed	-.158	-.555	.186
weak-powerful	-.438	-.147	.631
negative-positive	-.574	-.355	.332
violent-gentle	-.402	-.622	.373
important-unimportant	.828	.112	-.304
active-passive	.372	.155	-.674
insensitive-sensitive	-.356	-.546	.233
uninteresting-interesting	-.358	-.375	.578
pleasant-unpleasant	.324	.726	-.355
still-moving	-.199	-.212	.757
enjoyable-unenjoyable	.102	.636	-.252

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Religion (pretest)

4-Factor Solution and Solution of Adequate Statistical Fit Are Identical

	1	2	3	4
worthless-valuable	-.343	.779	.277	-.283
severe-lenient	-.368	-.511	.679	.112
unfair-fair	-.169	.371	.623	-.289
good-bad	.284	-.652	-.351	.371
tense-relaxed	-.134	.171	.686	-.128
weak-powerful	-.091	.446	-.316	-.484
negative-positive	-.294	.489	.471	-.469
violent-gentle	-.190	.205	.600	-.224
important-unimportant	.296	-.720	-.259	.343
active-passive	.221	-.335	-.149	.741
insensitive-sensitive	-.211	.386	.575	-.305
uninteresting-interesting	-.416	.469	.209	-.346
pleasant-unpleasant	.696	-.364	-.380	.325
still-moving	-.239	.203	.136	-.628
enjoyable-unenjoyable	.793	-.416	-.276	.349

$X^2 = 79.05$ $p < .01$ DF= 51

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Religion (posttest)

4-Factor Solution : X^2=79, p<.00, df=51

	1	2	3	4
worthless-valuable	-.745	.262	.358	-.280
severe-lenient	-.076	.718	-.070	-.031
unfair-fair	-.478	.566	.258	-.242
good-bad	.767	-.295	-.305	.260
tense-relaxed	-.210	.728	.228	-.222
weak-powerful	-.382	-.096	.549	-.122
negative-positive	-.571	.396	.457	-.293
violent-gentle	-.315	.568	.338	-.307
important-unimportant	.742	-.187	-.417	.206
active-passive	.267	-.076	-.674	.271
insensitive-sensitive	-.340	.430	.591	-.295
uninteresting-interesting	-.449	.203	.401	-.428
pleasant-unpleasant	.452	-.348	-.372	.628
still-moving	-.224	.275	.638	-.151
enjoyable-unenjoyable	.463	-.309	-.375	.636

5 Factor Solution: X^2 =46.53 p<.22 df=41

	1	2	3	4	5
worthless-valuable	-.685	.257	-.317	-.321	.267
severe-lenient	-.080	.714	.053	-.036	-.025
unfair-fair	-.490	.546	-.278	-.250	.046
good-bad	.826	-.260	.306	.256	-.077
tense-relaxed	-.207	.732	-.215	-.222	.094
weak-powerful	-.295	-.082	-.532	-.152	.494
negative-positive	-.559	.384	-.435	-.307	.181
violent-gentle	-.294	.573	-.298	-.318	.187
important-unimportant	.675	-.194	.305	.247	-.444
active-passive	.274	-.033	.771	.244	-.048
insensitive-sensitive	-.316	.427	-.554	-.311	.212
uninteresting-interesting	-.356	.224	-.272	-.490	.428
pleasant-unpleasant	.437	-.338	.371	.630	-.121
still-moving	-.224	.268	-.644	-.155	[illegible]
enjoyable-unenjoyable	.431	-.302	.361	.655	[illegible]

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: **Marijuana (pretest)**

4-Factor Solution : $X^2=84$, $p<.00$, df=51

	1	2	3	4
worthless-valuable	-.185	-.746	-.205	.364
severe-lenient	-.141	-.363	-.010	.560
unfair-fair	-.082	-.305	-.077	.615
good-bad	.171	.857	.109	-.401
tense-relaxed	-.224	-.419	.000	.591
weak-powerful	-.074	-.128	-.252	.212
negative-positive	-.210	-.718	-.137	.465
violent-gentle	-.173	-.376	.051	.745
important-unimportant	.119	.251	.405	-.156
active-passive	-.096	-.035	.757	.074
insensitive-sensitive	-.115	-.096	-.320	.491
uninteresting-interesting	-.202	-.347	-.487	.433
pleasant-unpleasant	.464	.548	.089	-.559
still-moving	-.079	-.011	-.693	-.014
enjoyable-unenjoyable	.771	.460	.181	-.401

6-Factor Solution: $X^2 = 44.87$ $p < .06$ df=32

	1	2	3	4	5	6
worthless-valuable	-.789	.132	-.086	.080	-.298	.197
severe-lenient	-.468	.448	-.134	-.037	-.002	.250
unfair-fair	-.319	.921	-.062	.001	-.139	.162
good-bad	.924	-.209	.019	-.029	.157	-.148
tense-relaxed	-.545	.175	-.083	-.112	-.063	.565
weak-powerful	-.152	.007	-.046	.132	-.217	.303
negative-positive	-.803	.233	-.038	.041	-.147	.268
violent-gentle	-.542	.323	-.200	-.153	-.021	.471
important-unimportant	.186	-.085	.026	-.153	.738	-.136
active-passive	-.047	.046	-.026	-.621	.249	-.023
insensitive-sensitive	-.198	.215	-.066	.220	-.145	.486
uninteresting-interesting	-.423	.200	-.150	.286	-.390	.330
pleasant-unpleasant	.721	-.246	.580	-.010	.084	-.277
still-moving	-.055	.028	-.043	.911	.014	.110
enjoyable-unenjoyable	.636	-.173	.425	-.098	.175	-.313

PRETEST AND POSTTEST ROTATED FACTOR PATTERN
MATRICES FOR EACH CONCEPT: 4 FACTOR
SOLUTION AND SOLUTION OF BEST FIT

CONCEPT: Marijuana (posttest)

4-Factor Solution : X^2=178, p<.00, df= 51

	1	2	3	4
worthless-valuable	-.857	-.076	-.126	.173
severe-lenient	-.639	.027	.028	.254
unfair-fair	-.633	.010	-.105	.363
good-bad	.941	.111	.108	-.163
tense-relaxed	-.528	-.155	.042	.461
weak-powerful	-.139	-.032	-.425	.367
negative-positive	-.796	-.265	-.116	.215
violent-gentle	-.721	-.141	.115	.351
important-unimportant	.364	.133	.334	-.285
active-passive	-.081	.042	.752	.048
insensitive-sensitive	-.404	-.091	-.175	.642
uninteresting-interesting	-.436	-.206	-.324	.472
pleasant-unpleasant	.754	.448	.128	-.295
still-moving	-.057	-.005	-.746	.129
enjoyable-unenjoyable	.729	.575	.158	-.289

7-Factor Solution: X^2 = 25.18 p < .34 df=21

	1	2	3	4	5	6	7
worthless-valuable	-.512	.137	-.257	.028	.420	.106	-.391
severe-lenient	-.354	.055	-.001	-.042	.790	.077	-.072
unfair-fair	-.397	.034	-.046	.084	.638	.224	-.154
good-bad	.778	-.104	.397	-.054	-.397	-.149	.212
tense-relaxed	-.535	-.173	-.078	-.107	.239	.457	-.195
weak-powerful	-.108	.103	-.051	.371	.061	.490	-.167
negative-positive	-.694	.560	-.059	.029	.304	.217	-.247
violent-gentle	-.661	.004	-.113	-.142	.334	.321	-.098
important-unimportant	.252	-.031	.048	-.196	-.077	-.189	.656
active-passive	-.026	-.012	-.001	-.696	.090	-.006	.116
insensitive-sensitive	-.344	.074	.041	.108	.285	.572	-.231
uninteresting-interesting	-.388	-.049	.055	.215	.232	.340	-.518
pleasant-unpleasant	.889	-.022	-.154	-.109	-.288	-.161	.207
still-moving	-.043	-.011	.008	.827	.127	.141	-.061
enjoyable-unenjoyable	.814	-.101	-.065	-.104	-.241	-.219	.296

RAW MEANS AND STANDARD DEVIATIONS
(RANGE = 1-7, N=252)

MYSELF AS A TEACHER

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	5.968	.719	6.115	.772
severe-lenient	4.500	1.008	4.611	1.060
unfair-fair	6.048	.660	6.286	.534
good-bad	2.472	1.158	2.063	.811
tense-relaxed	4.567	1.507	5.044	1.291
weak-powerful	4.690	.961	5.044	.995
negative-positive	5.905	.856	6.091	.791
violent-gentle	4.448	.954	5.627	.925
important-unimportant	2.286	1.074	2.020	.830
active-passive	2.167	1.113	1.853	.901
insensitive-sensitive	6.071	.872	6.202	.834
uninteresting-interesting	5.679	.895	5.976	.691
pleasant-unpleasant	2.123	.776	2.000	.783
still-moving	5.444	1.053	5.797	.894
enjoyable-unenjoyable	2.242	.879	1.944	.641

NONVERBAL BEHAVIOR

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	5.829	1.093	6.143	.886
severe-lenient	4.306	.846	4.345	1.039
unfair-fair	4.849	1.101	5.230	1.065
good-bad	2.662	1.192	2.262	1.046
tense-relaxed	4.488	1.334	4.631	1.301
weak-powerful	4.885	1.177	5.520	.992
negative-positive	5.111	1.209	5.429	1.143
violent-gentle	4.690	1.115	4.857	1.065
inportant-unimportant	2.302	1.209	2.000	.949
active-passive	2.988	1.367	2.702	1.325
insensitive-sensitive	5.456	1.231	5.694	1.010
uninteresting-interesting	5.425	1.200	5.746	1.078
pleasant-unpleasant	3.250	1.189	3.091	1.155
still-moving	4.817	1.314	5.083	1.274
enjoyable-unenjoyable	3.183	1.214	2.952	1.153

RAW MEANS AND STANDARD DEVIATIONS
(RANGE = 1-7, N=252)

QUESTIONING AND LISTENING SKILLS

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	6.5.36	.862	6.552	.645
severe-lenient	4.147	1.070	4.302	1.124
unfair-fair	5.492	1.141	5.742	1.086
good-bad	1.841	.977	1.790	.888
tense-relaxed	4.639	1.348	4.909	1.245
weak-powerful	5.397	1.224	5.774	1.026
negative-positive	5.849	1.119	5.917	1.012
violent-gentle	4.698	1.095	4.901	1.053
important-unimportant	1.798	1.058	1.754	1.003
active-passive	2.452	1.178	2.302	1.245
insensitive-sensitive	5.544	1.282	5.730	1.114
uninteresting-interesting	5.726	1.115	5.821	1.110
pleasant-unpleasant	2.813	1.198	2.635	1.101
still-moving	4.865	1.277	5.151	1.198
enjoyable-unenjoyable	2.774	1.318	2.532	1.127

MYSELF

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	6.016	.798	6.267	.655
severe-lenient	4.988	1.176	5.008	1.156
unfair-fair	5.766	.930	5.964	.771
good-bad	2.512	.963	2.127	.779
tense-relaxed	4.198	1.494	4.653	1.415
weak-powerful	4.675	1.145	4.980	1.129
negative-positive	5.464	1.144	5.721	1.067
violent-gentle	5.528	.979	5.606	1.039
inportant-unimportant	2.404	1.116	2.135	.962
active-passive	2.452	1.171	2.239	1.054
insensitive-sensitive	6.091	.929	6.183	.979
uninteresting-interesting	5.567	.798	5.801	.800
pleasant-unpleasant	2.345	.895	2.092	.822
still-moving	5.262	1.141	5.558	1.124
enjoyable-unenjoyable	2.476	.881	2.211	.824

RAW MEANS AND STANDARD DEVIATIONS
(RANGE = 1-7, N=252)

BEHAVIORAL OBJECTIVES

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	5.583	1.135	6.143	1.039
severe-lenient	4.194	.989	4.107	1.006
unfair-fair	5.091	1.225	5.667	1.118
good-bad	2.782	1.193	2.163	1.049
tense-relaxed	4.266	1.092	4.421	1.110
weak-powerful	4.698	1.043	5.397	1.090
negative-positive	5.226	1.188	5.643	1.129
violent-gentle	4.437	1.006	5.548	.987
important-unimportant	2.556	1.211	2.008	1.029
active-passive	3.091	1.179	2.489	1.166
insensitive-sensitive	4.980	1.283	5.246	1.257
uninteresting-interesting	5.056	1.176	5.444	1.278
pleasant-unpleasant	3.337	1.154	3.064	1.152
still-moving	4.537	1.087	4.980	1.182
enjoyable-unenjoyable	3.289	1.177	3.091	1.196

REINFORCEMENT

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	6.246	.938	6.520	.628
severe-lenient	4.198	1.079	4.020	1.127
unfair-fair	5.472	1.098	5.833	1.091
good-bad	2.107	1.034	2.012	0.996
tense-relaxed	4.627	1.179	4.623	1.219
weak-powerful	5.552	1.042	5.869	1.072
negative-positive	5.809	1.077	5.726	1.174
violent-gentle	4.746	1.183	4.821	1.130
inportant-unimportant	1.925	.869	1.758	.870
active-passive	2.647	1.170	2.298	1.137
insensitive-sensitive	5.289	1.153	5.627	1.209
uninteresting-interesting	5.289	1.139	5.492	1.088
pleasant-unpleasant	2.964	1.158	2.091	1.172
still-moving	4.782	1.124	5.103	1.117
enjoyable-unenjoyable	2.877	1.219	2.738	1.144

RAW MEANS AND STANDARD DEVIATIONS
(RANGE = 1-7, N=252)

RESPONDANT LEARNING

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	5.389	1.143	6.218	.811
severe-lenient	4.163	.902	4.329	1.063
unfair-fair	4.820	1.058	5.496	1.091
good-bad	2.866	1.110	2.278	.970
tense-relaxed	4.266	1.055	4.579	1.125
weak-powerful	4.722	1.046	5.603	1.011
negative-positive	4.948	1.133	5.556	1.034
violent-gentle	4.321	.839	4.702	1.084
important-unimportant	2.825	1.137	2.075	.883
active-passive	3.123	1.047	2.595	1.222
insensitive-sensitive	4.543	1.046	5.254	1.177
uninteresting-interesting	4.885	1.153	5.548	.987
pleasant-unpleasant	3.421	1.017	2.944	1.092
still-moving	4.516	.988	4.964	1.084
enjoyable-unenjoyable	3.325	1.096	2.921	1.083

SHAPING

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	5.040	1.160	6.222	.792
severe-lenient	4.123	1.008	4.385	1.025
unfair-fair	4.401	1.169	5.647	1.107
good-bad	3.357	1.160	2.186	.962
tense-relaxed	4.214	.954	4.639	1.129
weak-powerful	4.690	1.056	5.552	.998
negative-positive	4.579	1.203	5.655	1.069
violent-gentle	4.357	.923	4.821	1.095
inportant-unimportant	3.123	1.156	2.075	.878
active-passive	3.302	1.051	2.519	1.080
insensitive-sensitive	4.575	1.200	5.424	1.139
uninteresting-interesting	4.758	1.115	5.488	1.054
pleasant-unpleasant	3.643	1.048	2.964	1.130
still-moving	4.448	1.003	5.048	1.081
enjoyable-unenjoyable	3.627	1.080	2.885	1.111

RAW MEANS AND STANDARD DEVIATIONS
(RANGE = 1-7, N=252)

PHYSICIAN

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	5.905	1.268	6.119	1.076
severe-lenient	3.885	.952	4.016	1.071
unfair-fair	4.988	1.159	5.234	1.193
good-bad	2.448	1.272	2.099	1.112
tense-relaxed	4.662	1.375	4.571	1.353
weak-powerful	5.194	1.156	5.492	1.047
negative-positive	5.278	1.276	5.595	1.189
violent-gentle	5.301	1.145	5.313	1.157
important-unimportant	2.171	1.249	1.889	1.020
active-passive	2.730	1.290	2.444	1.191
insensitive-sensitive	5.289	1.233	5.385	1.333
uninteresting-interesting	5.019	1.155	5.341	1.137
pleasant-unpleasant	2.984	1.214	2.782	1.216
still-moving	4.726	1.218	5.052	1.101
enjoyable-unenjoyable	3.381	1.236	3.413	1.258

RELIGION

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	5.115	1.606	5.262	1.500
severe-lenient	4.143	1.358	3.972	1.396
unfair-fair	4.603	1.489	4.567	1.494
good-bad	2.948	1.544	2.845	1.424
tense-relaxed	4.278	1.398	4.341	1.454
weak-powerful	4.964	1.497	5.278	1.369
negative-positive	4.798	1.615	5.028	1.500
violent-gentle	4.702	1.248	4.817	1.368
inportant-unimportant	2.913	1.539	2.835	1.520
active-passive	3.206	1.441	3.016	1.351
insensitive-sensitive	4.980	1.446	4.988	1.457
uninteresting-interesting	5.000	1.486	5.282	1.465
pleasant-unpleasant	3.266	1.424	3.167	1.452
still-moving	4.671	1.336	4.794	1.433
enjoyable-unenjoyable	3.167	1.441	3.183	1.377

RAW MEANS AND STANDARD DEVIATIONS
(RANGE = 1-7, N=252)

MARIJUANA

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	3.468	1.578	3.687	1.520
severe-lenient	3.833	1.258	3.845	1.258
unfair-fair	3.940	1.122	3.988	1.242
good-bad	4.365	1.492	4.187	1.523
tense-relaxed	4.313	1.392	4.515	1.404
weak-powerful	4.270	1.246	4.425	1.408
negative-positive	3.512	1.545	3.742	1.554
violent-gentle	4.013	1.344	4.071	1.395
important-unimportant	4.380	1.566	4.309	1.530
active-passive	4.155	1.276	4.056	1.361
insensitive-sensitive	4.337	1.354	4.381	1.447
uninteresting-interesting	4.409	1.676	4.559	1.682
pleasant-unpleasant	3.972	1.632	3.810	1.652
still-moving	4.171	1.107	4.250	1.255
enjoyable-unenjoyable	3.869	1.637	3.821	1.721

OVERALL CONCEPTS

	PRETEST		POSTTEST	
	Mean	S.D.	Mean	S.D.
worthless-valuable	5.551	1.399	5.932	1.255
severe-lenient	4.223	1.109	4.268	1.168
unfair-fair	5.040	1.269	5.423	1.261
good-bad	2.760	1.362	2.364	1.242
tense-relaxed	4.409	1.308	4.630	1.294
weak-powerful	4.883	1.208	5.357	1.178
negative-positive	5.133	1.399	5.464	1.313
violent-gentle	4.756	1.177	4.917	1.205
inportant-unimportant	2.606	1.396	2.264	1.281
active-passive	2.936	1.318	2.592	1.307
insensitive-sensitive	5.194	1.321	5.466	1.296
uninteresting-interesting	5.163	1.260	5.500	1.209
pleasant-unpleasant	3.101	1.283	2.858	1.271
still-moving	4.747	1.207	5.070	1.224
enjoyable-unenjoyable	3.109	1.300	2.880	1.279

MICHIGAN STATE UNIV. LIBRARIES
31293100432131