THE EVALUATION OF A TEACHER TRAINING PROGRAM

By

Kim Kanaga

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Communication

1978

G112980

Accepted by the faculty of the Department of Communication, College of Communication Arts and Sciences, Michigan State University, in partial fulfillment of the requirements for the Doctor of Philosophy Degree.

Director of Dissertation

Guidance Committee: ______________, Chairman

E P Bettinghaus

David C Ralph

James L. Phillips

ABSTRACT

THE EVALUATION OF A TEACHER TRAINING PROGRAM

By

Kim Kanaga

The evaluation of instruction has traditionally been a controversial issue among educators. While there are several plausible explanations for this situation, one important contributing factor certainly appears to be the lack of adequate procedures for assessing instructional performance. Of the current practices in instructional evaluation, students appear to be the most promising source of information because they have much more exposure to the instructor in the classroom setting, they are recipients of the teaching process and their numbers are usually of sufficient size to minimize biases resulting from individual differences. Student rating forms are a quite widely used means of formally gathering evaluative information. However, they only indirectly tap into the teaching-learning process by examining various underlying components which are assumed to be a reflection of teaching effectiveness. A more appropriate and seldom used means of determining teaching effectiveness is to examine student learning. Learning is, after all, the goal of teaching.

A cognitive-mediation oriented approach to student learning appears to be the most fruitful for the purpose of evaluating instruction. From such a perspective, learning can be defined as the evolution of meaning. It would follow that the

function of teaching would be to develop consensually shared meanings for the basic concepts that define a specific domain of instruction. Metric multidimensional scaling provides a precise and reliable set of procedures for assessing the meanings for such a set of concepts over time. The experiment reported here was designed to provide an initial test of these procedures.

The current study evaluates a training program for teaching assistants by examining the teaching effectiveness of a group of trained assistants in contrast to a group of assistants that were not trained. Student learning and perceptions of actual classroom performance were used as the criteria for the evaluation. Data was gathered from student volunteers at four different times during the term.

Prior to discussing the results of this study, several problems should be pointed out. The particular data set used here is really not appropriate for the analyses that were performed on them. Specifically, every available student was asked to participate at each data collection point. The samples were thus not randomly drawn but the data was instead correlated to the extent that some subjects participated in more than one data collection. Moreover, the data was correlated within time intervals since some courses had more than one teaching assistant. As a result, the multiple responses by students in these courses were not independent. The analyses performed does, nevertheless for illustrative

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to a number of people whose contributions and support were instrumental in the completion of this dissertation. Above all, I would like to thank my advisor and dissertation director, Dr. Joseph Woelfel. His knowledge, guidance, and understanding have been an inspiration to me. For these things I am deeply grateful. Most of all, I am grateful to Joe Woelfel for having the flexibility to be there when I needed him and to otherwise allow me to find my own way.

Thanks are also due to the other members of my supervisory committee. Dr. Erwin Bettinghaus, Dr. James Phillips and Dr. David Ralph. Their assistance always seemed to come at the right times. In particular, Dr. Bettinghaus and Dr. Ralph provided the departmental approval for the study to be conducted and backed it through some occasionally difficult times. For this, I am indebted. It is also with a great deal of appreciation that I thank Dr. Larry Barker, my advisor for both my bachelors and masters degrees. It was his initial help and encouragement that guided me in the direction which ultimately led to this dissertation.

The success of this study was dependent on the co-operation of a large number of people. All of the Com 100 instructors cooperated in allowing me to come into their classrooms to gather data. The teaching assistants very courageously permitted their classroom effectiveness to be examined. The student respondents put several hours into filling out questionnaires. In actually collecting and coding the data; several undergraduate assistants were extremely helpful. In particular, Donna Paquette, Marianne Mnich, Cliff DeMarco and Ilene Benison put a great deal of time and energy into this study. Rick Holmes provided valuable computer assistance and Mrs. Ruth Langenbacher did her always excellent job in preparing this manuscript. Without the effort put forth by all of these people, this study would not have been possible. To each one of them, I am very grateful.

I also owe a great deal to those people who have been closest to me, my family. My parents, Kent and Phyllis Kanaga, have always had confidence in me and have been supportive of me in the things that I've done. My wife Mary and son Jeremy have put up with my absense and frustrations over the past several years. I cannot express what their love and understanding has meant to me. To them, I dedicate this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The evaluation of instruction has traditionally been a controversial issue among educators (Doyle, 1975; Fenker, 1975; Kent, 1966; Lewis, 1975; Miller, 1972). While most instructors would probably agree that there are potential advantages in carrying out such evaluations, Gray (1969) points out that they are usually less than enthusiastic about having their own classroom performance examined. This is particularly the case in higher education where the reasons for this lack of enthusiasm are quite evident. At the college level, very few instructors receive any sort of formal training (Doyle, 1975). In the absence of such training, instructors are left to either develop their own teaching skills or model their teaching after instructors from which they have had courses. The underlying assumption is that the instructor has, for certain courses, the necessary substantive knowledge and should be able to convey that information to students without much difficulty. The resistance to the evaluation of instruction may, then, be partially due to the instructor's insight into this deficiency in their own instructional training.

Another issue contributing to this controversial situation is the current status of the reward structure within institutions of higher education. Retention, promotion, tenure, and pay raises at many major universities tend to be based more on scholarly productivity than on instructional achievements (Astin and Lee, 1966; Doyle, 1975; Fenker, 1975; Hammond, Meyer, and Miller, 1969; Luthans, 1967). Even when teaching is taken into consideration, it is not unusual for the quantity of contact with students to be emphasized (Astin and Lee, 1966). Committee membership and other such responsibilities, in and of themselves, are usually not weighted heavily in the reward system. They do, however, involve working with peers and as such would be expected to elevate their importance to faculty members. Moreover, consulting and other outside activities typically provide substantial incentives. It would appear, then, that university systems are not structured to promote quality instruction. Instead, the emphasis on other responsibilities would seem to lower the priority placed on developing and maintaining superior teaching skills.

The lack of enthusiasm among instructors may, in fact, reflect their dissatisfaction with the outcomes resulting from current evaluation practices. Hence, instructors may see potential advantages to evaluating classroom performance, but have yet to see any of those advantages realized. Although administrators may now be at least looking at evaluation forms, they still appear to have little faith in them

(Astin and Lee, 1966; Lewis, 1975). Furthermore, the information provided by present evaluation forms is of limited value to instructors who wish to use them for assistance in improving their teaching skills. Typically, these forms rate the instructor on various criteria (Genova, Madoff, Chin, and Thomas, 1976). The instructors are then informed as to how they rate in comparison with other instructors. In some circumstances, there is no explicit indication of whether the individual instructor's position is more or less effective than the other instructors, but merely how much they differ from other instructors. Even if specific strengths and weaknesses were precisely pinpointed, these forms rarely contain any recommendations for how specific teaching skills may be improved. Thus, the skepticism surrounding instructional evaluation may be, in part, due to the failure of current evaluation procedures to produce any meaningful results.

This latter point suggests that at least some of the problems associated with instructional evaluations may be found in the instrumentation rather than the lack of concern among instructors and administrators. An extensive survey conducted by the American Council on Education (Astin and Lee, 1966) revealed that systematic measuring instruments are, in fact, not a widely used means of evaluating classroom performance. Instead, they found that by far the most frequent sources of information are the subjective judgments of department chairpersons (85.1%) and deans (82.3%) followed distantly by scholarly productivity (43.8%) and the informal opinions of colleagues (48.9%) and students (41.2%). As Gustand (1961) put it, ". . . the most frequently cited sources of information can be described simply as hearsay."

The extensive use of such imprecise procedures may, very well, be an indication of the quality of the measuring instruments that are currently available. Doyle (1975) and other contemporary educational researchers (Blackburn and Clark, 1975; Kauffman, Hallahan, Payne, and Ball, 1973; Lewis, 1975; and Miller, 1972) have suggested that current evaluation procedures are inadequate. This conclusion was also reached by the Commission on Academic Tenure in Higher Education (1973). In their report, they recommend that "professional associations and research centers continue and expand their efforts to develop more sophisticated conceptions of teaching effectiveness and more reliable methods of evaluating teaching . . . ." The research reported in this paper is such an undertaking. More specifically, the experiment reported here was designed to provide an initial test of a new set of procedures for evaluating instruction.

The remainder of this chapter will first discuss the purposes for evaluating instruction. The various types of instructional evaluations that are currently in use will then be critically examined. One specific type, namely student evaluations, will be singled out as the most fruitful means of evaluating instruction. The problems in this area will be thoroughly reviewed. The chapter will then conclude by laying out the organization for the remainder of this dissertation.

Purposes for Evaluating Instruction

Evaluations are generally conducted to examine some program or activity. More specifically, they attempt to determine the extent to which some objective or objectives are being accomplished. The primary concern for instructional evaluations is the assessment of the instructors success in maximizing student learning of course material. While there are a wide variety of specific purposes for which the information obtained from such assessments may be used, Doyle (1975) points out that they typically fall within the following four basic categories: administrative decisions, self-improvement, criterion for research, and advising students.

Administrative decisions can be aided by the meaningful evaluation of instruction. As in most organizations, someone or some group within the educational system has the responsibility for the distribution of internal rewards and punishments. Since instruction is a primary responsibility of most faculty members, precise information regarding their effectiveness as instructors would seem to be a highly desirable and useful commodity for those who have to hire, fire and promote. Considering trends in student enrollment and in available financial support for higher education, this information should become even more important in the future.

Evaluations can also be useful to the instructors themselves. Scholars of communication have long recognized the value of feedback in modifying inappropriate or ineffective behaviors (Barker, 1971; Mortensen, 1972). Friedrich, Galvin and Book (1976) discuss several cases in which systematically manipulated feedback such as head nods, smiles, verbal praise, etc. was successfully used to change specific instructional behaviors. While such direct behavioral modifications are beyond the scope of instructional evaluations, feedback provided by evaluations can be used to help instructors identify instructional skills that need improvement.

The third purpose for evaluating instruction that Doyle mentions is in the area of educational research. Since the ultimate goal of educational research would presumably be to better understand the educational process so that improvements can eventually be made, it would seem that a means of precisely determining how effective instructors are in their classrooms would be of great value. That is, meaningful information regarding the teaching-learning process could, if used constructively, be beneficial to the field of education.

Finally, Doyle points out that students and academic advisors have recently taken an increased interest in obtaining information from instructional evaluations. In most educational institutions, the only formal information provided for students and their advisors is information regarding the content that will be covered in various courses. Any information about specific instructors is usually obtained through informal and often inaccurate or incomplete channels. The manner in which an instructor presents course material can be an important determinant of how much student learning will occur. As the consumer of the educational process, it is not surprising to see students concerned with obtaining information about prospective instructors and their styles of instruction. Seeking more formal and precise information does not seem unreasonable and a more concentrated effort to do so should not be unexpected.

Most evaluation instruments currently in use were developed to serve one or possibly two of the four purposes mentioned. While even the most precise and thorough evaluation procedures may not be able to adequately fulfill each of these purposes, an attempt in this direction should still be made.

Current Practices in the Evaluation of Instruction

There are probably as many different instructional rating instruments as there are colleges and universities which employ them. Each one purports to tap the most essential characteristics of teaching effectiveness. It would follow that if these instruments are, in fact, measuring the same phenomena, there should be some commonality among them. While some of the instruments are similar in appearance and claim to be measuring the same dimensions (Doyle, 1975), the items used to tap these dimensions are quite different (Genova et al., 1976). Since the items are not comparable from instrument to instrument, it is quite possible that even if they were examining the same dimensions, they may be focusing on a different range of those dimensions. Still other instruments are quite diversified in terms of the source and type of information they obtain. Consequently, a meaningful discussion of current evaluation instruments as a whole is questionable. Instead, these instruments can be more clearly examined within the context of the following four categories: administrator evaluation, colleague evaluation, self evaluation, and student evaluation.

Administrator Evaluation

In most cases, it is ultimately the administrators responsibility to evaluate the teaching performances of faculty members. The administrator is typically asked to make an overall assessment of teaching effectiveness as well as of other faculty duties. While more specific information is probably desirable, the more crucial issue in administrator evaluation is to determine on what basis such general assessments are made. Most administrators are unable to actually observe the classroom behavior of their faculty. The reason for this boils down to a simple matter of priorities. Although the evaluation of faculty teaching is an administrative concern, it is only a portion of the responsibilities delegated to a dean or department chairperson. Consequently, administrators do not have the time necessary to thoroughly evaluate each faculty member. They must, instead, rely on other sources of information.

Even if time permitted, the utility of an overall assessment based on the observations of any single individual is questionable. That is, direct observational ratings of this sort are based on individual perceptions which are at least partially determined by past experiences. While explicitly stated criteria for making ratings may minimize this potential source of bias, it is difficult, in the absence of other information, to determine if the criteria have been appropriately applied by any one individual. Several possible problems are noteworthy. Administrators faced with such a task may be partial to certain teaching styles which would not necessarily render other methods as less effective but would merely be an indication of personal preferences. It could also be difficult for them to determine the extent to which the instructor is facilitating or impeding student learning of course materials. What is clear and understandable to the administrator may be vague and incomprehensible to the students. What may seem as disorganized to the administrator may instead provide the students with the flexibility needed to get them more involved in class projects. Another potential source of bias may result from a difficulty in assessing teaching performance independently of other components which comprise the faculty members' overall contribution to the department and/or college (Blackburn and Clark, 1975). Moreover, these and other individual biases could result in a lack of uniformly applied standards which, in turn, may result in serious morale and/or legal problems.

This is not to say that these biases will occur or that they are by any means intentional. But even if administrators were aware of their biases, they may not be able to eliminate them and in attempting to do so may cause biases in other directions. It would appear, then, to be to the administrators' advantage to seek out the most precise, reliable, and valid information at their disposal and to utilize that information in evaluating teaching effectiveness.

Colleague Evaluation

Colleague evaluations tend to parallel administrator evaluations in both purpose and problems. These evaluations are primarily used to provide administrators with information regarding the performances of faculty members. While some peer rating forms (Genova et al., 1976) ask for summary information, others (Blackburn and Clark, 1971; Miller, 1972) allow colleagues to make more specific assessments and as such these instruments may also be used to assist instructors in improving their teaching skills.

There are several problems which accompany this type of evaluation. Faculty members are not immune to the biases that administrators are subject to when directly observing teaching behavior. In fact, they are probably more prone to bias. Faculty members have more personal contact with one another and would themselves be in a position to subsequently be evaluated by those faculty members whom they must evaluate. However, Genova et al. (1976) points out that faculty members rarely visit each others' classroom for the explicit purpose of assessing teaching performance. As a result, their evaluations must also be based on other sources of information.

A more serious issue resulting from peer evaluations is the impact that they may have on the faculty. Eble (1970), as well as others (Fenker, 1975; Miller, 1972), has found strong opposition to colleague evaluations. Fenker (1975) reported that resistance was so strong that his attempt to collect colleague evaluations had to be abandoned. This suggests that, for at least administrative purposes, the cost of such information may outweight the benefits.

Self Evaluation

As evidenced by the lack of attention in the literature, formal self evaluations are relatively rare in higher education. Moreover, the research that is available (Blackburn and Clark, 1971; Centra, 1972) tends to suggest that instructors inflate the assessments of their own teaching skills. If, on the other hand, faculty members were able to accurately assess their own effectiveness as a teacher, it is questionable whether individuals should be put in the position of providing the information that would warrant their own dismissal or would call for a promotion or pay raise. It may be desirable to think that professionals could and would make such self-judgments. To make such an assumption is, however, unrealistic and would establish a potentially detrimental reward system.

Formal self evaluation may however, be used as a means of improving teaching skills. Having instructors seriously reflect on their teaching behaviors could be quite helpful to some instructors; particularly those interested in improving their teaching ability but not knowing how to proceed. Self evaluation could provide them with some needed direction. Under these conditions, additional information from other sources would also be desirable.

Student Evaluation

Student rating forms are by far the most widely used means of formally gathering evaluative information for each of the purposes previously mentioned (Doyle, 1975). That is, student evaluations have been used to: (1) provide information for administrative purposes, (2) provide diagnostic feedback for instructors, (3) provide data for educational researchers, and (4) provide information for students and their advisers.

The most profound issue regarding student appraisals of teaching effectiveness is a debate over the students' ability or competence in making such judgments. While administrator and colleague evaluations involved assessments by individuals with at least comparable experiences and training, the opponents of this type of evaluation argue that students are not qualified to make such assessments (Bryant, 1967; Hildebrand, 1972). Supporters of student evaluations counter by arguing that as the consumer of the instructor's efforts, students are in the best position to actually know the effectiveness of the instructor (Astin and Lee, 1967; Grush and Costin, 1975). The issue is, however, contaminated by the limitations of the evaluation procedures that are presently in use.

Most educators would probably agree that student responses based on learning are desirable. But current instruments fail to actually measure student learning as a criterion for determining teaching effectiveness. Instead, they typically require students to indicate, on some sort of forced choice scale, the extent to which a given statement accurately characterizes a specific instructor. These statements are intended to reflect the various underlying components which, when taken as a whole, constitutes teaching effectiveness. Students are asked to make ratings on such characteristics as the instructors' flexibility, motivation, rapport, teaching style, and interaction patterns (Doyle, 1975). Extensive reviews of empirical research by Doyle (1975) and Costin _et al_. (1972) show rather convincingly that students are quite capable of making such assessments. Although these results are still occasionally questioned, it becomes more crucial to determine the extent to which the components are related to the actual learning that takes place in the classroom.

In the absence of direct measures, it is assumed that, collectively, the dimensions tapped by the measuring instrument are a reflection of the teaching-learning process. The issue, then, becomes as Rippey (1975) points out, not a question of whether or not students should be used for the purpose of determining teaching effectiveness, but rather a question of how to best utilize this potentially valuable source of information.

In addition to the problems created by the nature of the information that present evaluation procedures obtain from students, they also suffer from several methodological shortcomings. All of the instructional evaluation instruments that are either reproduced or described in the literature are subject to at least one of the following limitations:

1) All relevant criteria which prospective respondents use to evaluate instruction must be known in advance and/or appropriately captured in the measuring instrument. Statements characterizing specific instructional qualities must relect these criteria. The lack of agreement regarding what are the essential criteria and how they are to be tapped suggests that current instruments may not measure all of the criteria by which instruction is evaluated.

2) Respondents are typically required to assess the instructor separately for each of the criteria that are employed in the instrument. Since these criteria are seldom independent of one another, obtaining additional information regarding the interrelationships among these criteria would provide a more complete picture of the evaluation process.

3) There is a general lack of precision of measurement in instructional evaluation forms. These instruments extensively utilize some sort of Likert and/or semantic differential type scales which, at best, can be considered ordinal measure. The necessary precision of measurement is dependent on the phenomenon under consideration. Ambiguous conceptualizations of teaching effectiveness make this issue unclear. However, unless evidence is available to suggest that ordinal scales would be adequate, more precise measures should be sought.

4) Instructional evaluation instruments are rarely designed with over-time analysis in mind. Virtually all of the instruments that are in use are set up as ex-post-facto measures. If the teaching-learning relationship is a dynamic one, which few educators would dispute, it should be examined across time. End of the term measures only provide a static and potentially misleading description of a complex process. Moreover, they come too late for instructional modifications to be made.

5) Criteria that are explicitly stated in the form of statements concerning specific instructional qualities runs the risk of being quite easily distorted by respondents. Administrators, colleagues, and students all have vested interests in the effectiveness of the instructors they are asked to evaluate. Consequently, the ability to bias their responses should always be minimized.

6) In order to determine just how effective an instructor is, what constitutes effective instruction must first be identified. While much attention has been focused on what criteria ought to be incorporated into an instrument, there is usually no provisions made for determining some optimal level of performance for each of these criteria. Without an explicit standard for comparison, interpretation of any results would be at least somewhat questionable.

In summary, it has been pointed out that assessments of teaching effectiveness have been used to provide information for administrative purposes, for the improvement of teaching skills, for educational research, and for counseling students. As sources of information, it was suggested that administrators and colleagues are unable to invest the time necessary to thoroughly evaluate teaching effectiveness and are highly susceptible to a number of biases. Self evaluation was found to be impractical. Students appear to be the most promising source of information because they have much more exposure to the instructor in the classroom setting, they are recipients of the teaching process, and their numbers are usually of sufficient size to minimize error resulting from individual differences. However, until a precise means of measuring student learning is developed, their potential value will not be fully realized.

It seems quite clear, then, that if the formal evaluation of instruction is to play a vital role in the educational process, the teaching-learning relationship must be the focus of theoretical development and a measurement system not suffering from the limitations of current procedures must be devised.

Organization of the Report

The present chapter provides an introduction to the problem under investigation as well as a specific direction in which a solution to the problem will be pursued. The remainder of the report is separated into four additional chapters. In Chapter II, a theoretical framework and measurement system are developed. Chapter III contains a description of the experimental methods and procedures used in the study. The results of that study are then presented in Chapter IV. Chapter V concludes with a discussion of the study, summary, and recommendations for future research in this area.

CHAPTER II

THEORY AND MEASUREMENT

In pointing out that "the ultimate criterion of good teaching is student learning," McKeachie, Lin, and Mann (1971) are not making any startling revelation. They are merely stating what educators have known for a long time. Learning is, after all, the purpose of teaching (Tyler, 1958; Cohen and Brawer, 1969). It has, however, rarely been used as a determinant of teaching effectiveness because traditional measures of learning are imprecise (Kauffman, Hallahan, Payne, and Ball, 1973) and not conveniently applied in the educational environment. Before turning attention to the theoretical under-pinnings of learning measures, it would be informative to first discuss the means that have been used to measure student learning when such attempts are made. One means of getting at learning has been to tap into the grading system (Costin et al., 1972). While grades are intended to differentiate levels of student learning, they do not necessarily reflect the instructors' teaching effectiveness. The assignment of student grades may also indicate the quality of students in the course, the difficulty of the course material, as well as the instructors' standards for assessing student performance. For example, a final exam in which all students correctly respond to all items does not necessarily mean that the instructor was maximally effective. It would probably suggest that the instructor was unable to construct a test that adequately discriminated between various levels of student learning. Conversely, an instructor assigning abnormally low grades may be very effective but rigorous. Standardized achievement tests have also been used to determine student learning (Gessner, 1973). However, they often lack flexibility and are not available for all subject matters. The most rarely used means of measuring student learning has been to merely ask students what or how much they have learned. Even though researchers (Doyle and Whitely, 1974; Frey, 1973) have found that students' responses to these types of questions are positively related to final exam scores, their accuracy in making such direct judgments is somewhat questionable. Moreover, their responses are quite easily distorted. Although these latter two types of measures have shown some promise, they are still not without serious limitations. This lack of adequate measures of classroom learning can be traced to the theories from which the measurement systems were derived.

Learning Theory

The most influential theories of learning (Bolles, 1975; Hilgard and Bower, 1975; Pittenger and Gooding, 1971; Snelbecker, 1974) have been behaviorally based in the traditions of Thorndike, Pavlov, and Skinner. These theories generally define learning as a relatively stable change in a behavioral pattern resulting from repeated exposure to some external stimuli (Snelbecker, 1974). They were comprehensively designed to encompass learning in all animals and as such were specifically tested on rats, pigeons, and monkeys with the express purpose of generalizing the results to humans. That is, it was felt that if other animals adhere to these principals of learning, humans, despite their cognitive potential, must also be subject to them.

To use such a theoretical framework in formal education would require specific behavioral outcomes for all course material to be identified and measured. That is, learning could only be detected by observing whether the appropriate response followed exposure to some specified evoking stimuli. Consequently, an assessment of teaching effectiveness must take into account the specific student responses that would be expected to result from exposure to course material. Moreover, procedures for making observations of these responses would have to be established. While this may be a desirable goal, an evaluation system based on this information must either assume comparability of outcomes and measures, or make some allowances for differences. In other words, it may be difficult to compare the teaching effectiveness of a typing instructor whose students have improved their typing efficiency with an art history instructor whose students have become able to identify specific paintings and sculptures with a calculus instructor whose students have increased their ability to solve differential equations. Differences in specific student responses which reflect learning may prevent any meaningful comparison of the instructor's effectiveness. A more serious problem has to do with the incompatibility of the behavioralist learning theory and the typical way courses proceed. Course material is usually covered continuously by topics. Once taught, topics are not frequently repeated. Consequently, the repeated exposure to stimuli, which is usually a condition for measuring learning, is quite often not present. Moreover, behavioral learning theories typically call for the response to follow immediately after exposure to the stimulus. In the classroom situation, it is not uncommon to have weeks go by between the time certain course material is covered in class and the time that students are expected to utilize that information. What happens to the information during this time between stimulus and eventual response is not easily accounted for by the strictly behavioral perspective. It seems clear, then, that while these theories may be applicable in other contexts, they are of limited utility in evaluating classroom instruction.

A more appropriate theoretical orientation can begin to be developed by first focusing attention on the work of Hebb (1958, 1960), Mowrer (1960a, 1960b) and Osgood (1957, 1963). Hebb (1958) contends that there exists thought or mental processes "which themselves independent of immediate sensory input, collaborate with that input to determine which of the various possible responses will be made and when" (p. 107). In other words, these mental processes mediate responses to incoming stimuli. What is being suggested here is essentially a two-stage activity. Stage one involves the internal interpretation of external stimuli. In stage two, the interpretation serves as an internal stimulus for a given response. This internal interpretation is what Osgood (1957) refers to as "meaning" which is gained from an association made between specific stimuli and the responses which they have elicited. Put another way, Mowrer (1960a asserts that, "External events or stimuli influence and determine behavior, not in a direct, physicalistic sense, but through the intervening action of the meanings they have acquired for the individual as a result of prior experience, i.e., through learning, either immediate or vicarious" (p. 309). It is further suggested that because certain stimuli are linked to similar responses, these stimuli also become associated with one another. New stimuli can then be learned by being associated with other stimuli that already have specific responses attached to them.

In a study investigating the associations between various stimuli, Noble (1952) found the existence of chain associations. That is, one stimulus is associated to another which in turn is associated with another and so on. In other words, the internal response to an external stimulus can serve as an internal stimulus for yet another internal response and this process can go on indefinitely or until the ultimate external response is made. Theoretically, if the chain were long enough, the resultant external response need only be remotely related to the initial stimulus. Nevertheless, this indicates that rather than internally possessing an enormous number of unrelated associations, stimuli are probably incorporated into an overall structure and are at least indirectly related to one another.

This idea of an internal structure appears to be quite consistent with the position taken by Tolman (1932) and the Gestalt theorists (Koffke, 1935; Kohler, 1947; Wertheimer, 1959). From a more cognitive perspective, they viewed learning as the acquisition and organization of information about the environment. In other words, their emphasis was on man's ability to see meaningful relationships and structure (Snelbecker, 1974). Tolman (1932) did not, however, abandon behaviorism. Instead, he also theorized that cognitive processes intervened between external stimuli and overt behavior. Stimulus information is incorporated into an individual's cognitive structure until such a time as it is called into play by some activity. Thus, the relationship between external stimuli and behavioral response is indirect and not necessarily immediate.

In order for the cognitive structure to become useful in the assessment of student learning, it must first be determined how the structure is organized. What is the basis by which associations are made? What is the criteria for relating stimuli to one another in the cognitive structure? Bruner (1966) suggests that stimuli are categorized as concepts or objects of an individual's experience. Objects are associated with one another when they belong to the same category. Categories are established according to certain essential distinguishing characteristics or what Bruner refers to as criterial attributes. "To categorize is to render discriminately different things equivalent, to group the objects and events and people around us into classes, and to respond to them in terms of their class membership rather than in terms of their uniqueness" (Bruner, 1966, p. 1). Thus, learning involves knowing: (1) what attributes are available, (2) when they should be applied, and (3) how they should be applied. Successful learning occurs when stimuli or objects are appropriately categorized.

In the absence of more specific information, a categorical system may be sufficient to explain the organizational pattern of the cognitive structure. However, finer discriminations are usually possible. Individuals are able to identify and respond to unique entities in the environment. For example, the major means of transportation that is usually found in the garage is not just a car, but a given make and model produced in a certain year by a specific automobile manufacturer. The use of additional properties or attributes such as color, rust spots, dents, mileage and license plate numbers would eventually identify a single car. Maintenance and use of that car is likely to depend not only on the categories that it belongs to but also on its idiosyncrasies. In other words, as more information becomes available, the categories in which an object can be placed become narrower and narrower until such a time as the object becomes a category in and of itself. The object is then differentiated from everything else and can be responded to according to its unique and/or categorical characteristics. Thus, objects gain their meaning by their placement on attributes relative to other objects.

Woelfel (1974) has further suggested that the information extracted from the environment is organized into a pattern of similarities and differences among the concepts that are symbolized by words in the vernacular language. Each concept is uniquely defined by its interrelationships with all other concepts. Changes in the meaning of concepts are cognitive processes and result in modifications of the overall cognitive structure. For example, speech, as an academic discipline which serves as a categorical domain of meaning, is defined according to how it is differentiated from art, biology, business administration, history, physical education, psychology, and all the other academic disciplines. While, at one time, speech was strongly associated with the humanities, it has begun to be more strongly associated with the social sciences. As a result, the cognitive structure of those in the academic community that are aware of the changing emphasis in speech-communication departments would be expected to be experiencing a reorganization in which the location of speech within the structure is undergoing change.

Given this foundation of cognitive processes, learning can be defined as the evolution of meaning. This definition portrays learning as a continuous process where the meanings for concepts are initially created by setting observations into correspondence with specific symbols and are subsequently molded by the accumulation of concept-relevant information. Since attitudes, beliefs, and values all essentially involve relationships among concepts, changes in the meanings of concepts would appear to be of fundamental importance to these processes. In other words, learning, in addition to its function of evolving specific concept meanings, is also the basis for the development and modification of an individual's attitudes, beliefs, and values. It is in this latter sense that meanings will be examined. Rather than investigating meanings for individual concepts (Osgood, Suci, and Tannenbaum, 1957), meanings for concepts relative to one another will be emphasized.

In terms of formal education, learning must be considered in conjunction with teaching. The function of teaching is typically to develop consensually shared meanings for the concepts that define the domain of instruction. In contrast, teaching is also occasionally used to broaden student perspectives. Whether the intent is to develop individual or collective meanings, all courses are usually set up around some overriding theme or goal. The relationships between the concepts that make up this theme constitute the basic foundation on which the course rests. Within this framework, specific course objectives may be dealt with by a number of different teaching techniques. The success of students in meeting these objectives is, of course, important. Just as important, if not more so, is the extent to which the students have properly internalized the meanings of the basic concepts on which the course is based. In this respect, teaching effectiveness can be seen as the progress of the students, as an aggregate, in learning the interrelationships between the concepts which define the overall structure of the course.

A potential limitation of a cognitive perspective is its generalizability to all learning situations. In particular, the utility of an evaluation system based on a cognitive conceptualization of learning may be called into question in skill or performance oriented courses. The primary objective of a calculus course, for example, would probably be to teach students to actually calculate differentials and integrals. An evaluation based on the student's ability to make these calculations would be a more direct and probably more desirable measure of teaching effectiveness. On the other hand, a good grasp of the course material should also be reflected in the development of specific meanings for these operations. The distinction appears to be one of knowing how to perform a specific activity versus understanding when it is appropriately performed. Both need to be learned and neither is precluded when cognitive processes are considered to intervene between external stimuli and resulting behaviors.

It should also be explicitly pointed out that this theoretical framework is not strictly cognitive. That is, when students provide information regarding their cognitive structures, they are in essence behaviorally responding and it is assumed that those responses reflect the classroom stimuli that they have been exposed to. Rather than focusing on ultimate behavioral outcomes which vary from course to course, the emphasis is on the mediating cognitive processes which it is assumed that all students must use to interpret and assign meaning to stimuli. The result is that the behavioral responses required of students, regardless of their course, is completely standardized. Moreover, it is assumed that the cognitive structure serves as the basis upon which other behaviors are made.

Measurement of Teaching Effectiveness

One of the principle objectives to early cognitive oriented theories of learning was the lack of adequate measurement of cognitive processes (Snelbecker, 1974). It was assumed that since cognitions are unobservable, they could not be measured with any degree of success. Ironically, this argument can now be turned around and used in support of a cognitive perspective of learning. It is not that cognitive processes are now observable but rather that useful procedures have been devised which allow these processes to be accurately measured. Metric multidimensional scaling provides a means of assessing cognitive processes in a manner that has been found to be precise and reliable (Gillham and Woelfel, 1976; Woelfel, 1977). Moreover, a multidimensional analysis as is to be described here satisfied at least some of the limitations of measuring teaching effectiveness that are found in traditional evaluation instruments.

Multidimensional scaling is based on the fundamental concept of psychological distance (Helm, Messick, and Tucker, (1959). That is, the perception of difference between stimuli or objects is the basis of the measurement scheme. Woelfel (1972) suggests that differences "among objects (whatever those objects may be) may be represented by a continuous numbering system such that two objects considered to be completely identical are assigned a paired dissimilarity score or distance score of zero (0), and objects of increasing dissimilarity are represented by numbers of increasing value." To utilize the set of real numbers in such a manner, a rule must be established for setting the numbering system into correspondence with perceptions of difference.

Stevens (1951), Torgerson (1958) and others (Campbell, 1928; Suppes and Zines, 1963) have stated that a rule for quantifying distance or difference must stipulate an arbitrary standard difference to which all other differences are to be compared. Such a rule is provided by Einstein (1961):

> For this purpose (the measurement of distance) we require a 'distance' (Rod S) which is to be used once and for all, and which we employ as a standard measure. If, now, A and B are two points on a rigid body, we can construct the line joining them according to the rules of geometry; then, starting from A, we can mark off the distance S time after time until we reach B. The number of these operations required is the numerical measure of the distance A B. This is the basis of all measurement of length.

To use Einstein's rule for measuring the perceived differences among objects, all that needs to be done is to arbitrarily stipulate that the difference between any two objects is some designated distance and that this distance is the standard of comparison for all other pairs of objects. This can be accomplished by wording a question in the following form:

> If the difference between *a* and *b* is *u* units, how different are *x* and *y*?

Responses to this type of question would be ratio measures of the perceived differences between pairs of objects.

When this ratio rule is applied to N concepts, N(N-1)/2 non-redundant paired comparisons are possible. Completing all such pairs produces a NxN symmetric matrix D for an N concept domain. This matrix, then, represents the overall pattern of differences among the concepts in the domain. Woelfel (1972) further states that:

> . . . the definition of an object or concept is constituted by the pattern of its relationships to other objects, the definition of any object may be represented by an 1 x n vector, d_{11}, d_{12}, d_{13}, . . . , d_{1n}, where d_{11} represents the distance or dissimilarity of object 1 from itself (thus $d_{11}=0$ by definition), d_{12} represents the distance or dissimilarity between objects 1 and 2, and d_{1n} represents the distance between the 1st and nth objects. Similarly, the second object may be represented by a second vector, d_{21}, d_{22}, d_{23}, . . . , d_{2n}, and the definition of any set of concepts or objects may therefore be represented in terms of the matrix
>
> $$\begin{array}{cccc} d_{11}, & d_{12}, & \ldots, & d_{1n} \\ d_{21}, & d_{22}, & \ldots, & d_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ d_{n1}, & d_{2n}, & \ldots, & d_{nn} \end{array}$$
>
> where any entry d_{ij} represents the dissimilarity or distance between _i_ and _j_.

Since the primary concern in evaluating teaching effectiveness is with aggregates of students, potential unreliability resulting from individual differences in responses is minimized by averaging dissimilarity scores across the students in a specific course. This procedure yields a means distance matrix D which represents the average dissimilarity among concepts which constitute the domain of instruction. This matrix is transformed into a centroid scaler products matrix B (Young and Householder, 1939) which when factored

(Jacobi, 1846) results in a cartesian coordinate system of orthogonal dimensions or axes. A rectangular matrix F is then constructed with the dimensions as columns of the matrix and with the rows representing the projections of the concepts on each of the dimensions. These procedures, which are described in much more detail by Torgerson (1958), Woelfel (1972, 1974, 1977), Woelfel and Danes (1977), and Serota (1974), essentially map the structure of the domain.

Cognitive processes such as learning can be represented by comparing a series of these spatial configurations gathered at several points in time. The coordinate systems generated at each point in time can be rotated and translated into a least square best fit with one another (Woelfel, 1977; Woelfel and Saltiel, 1974). In other words, this procedure is used to establish a common frame of reference from which changes in the meanings of objects or concepts can be observed. Changes in the spatial location of the concepts over time is then interpreted as motion through the space and as such can be mathematically expressed as velocities and over multiple time periods as accelerations. This enables at least portions of the learning process to be precisely assessed over the length of instruction.

The implementation of this multidimensional scaling technique for the purpose of evaluating teaching effectiveness is straight forward. What first needs to be done is to identify the relevant concepts that will be scaled into the space. Since these concepts define a particular domain of meaning, they must be topic or course specific. Moreover, these concepts must describe the basic nature of the course, not evaluate it. In other words, they must constitute the overall framework on which the entire course is based. Such concepts typically make up the general course objectives or outline. For example, one of the primary goals of an introductory communication course is to provide students with some sort of definition of communication and a notion of how the study of human communication can provide them with some useful information. From this objective, the following key concepts can be extracted: communication, process, information, valuable, meaning, social science, humanities, and me (representing the self-concept, see Woelfel and Danes, 1977). A second goal is typically to differentiate several types of communication systems. Depending of course, on the systems that are explicated, the following concepts might emerge: communication, information processing, interpersonal relationships, group interaction, organizations, mass media, social change, persuasion, meaning, leadership, self-concept development, efficiency, and socialization. A course in research methods, which covers more technical information could be described by concepts such as: science, theory, measurement, precision, mathematics, function, causality, experimental design, reliability, and validity.

A predictable pattern of interrelationships among the selected concepts should develop over the duration of the course. Assuming the course materials were being appropriately taught, the meanings given these concepts by the students would be expected to begin to converge on the projected meanings provided by the course objectives. Going back to the example of the introductory communication course, several changes would be expected to occur. Leadership, for instance, may not be intuitively thought of as a communication construct. If it is effectively taught as such, a stronger association between communication and leadership would be expected to emerge. Assuming a strong relationship between cognitive processes (i.e. attitudes and beliefs) and behaviors (see Liska, 1975; Cushman, 1977) the subsequent communicative behaviors of the students should become more salient to them when they are placed in a position of leadership. If the course objectives are attained, the students would also have begun to more fully realize the value of communication. Since concepts associated together have been found to converge on each other in the multidimensional spatial configuration (Woelfel, Cody, Gillham, and Holmes, 1976), the perceived dissimilarity between communication and valuable should decrease.

Doyle (1974) has suggested that a useful criteria for evaluating instruction is the stimulation of student interest in the subject matter of the courses that they are taking. This can be assessed for all courses by examining the relationships between the concepts which constitute the instructional domain and the concept "me" which as previously mentioned represents the students' self-concept. Increased student interest would be reflected by the student's identifying more closely with the concept which best characterizes the course's subject matter. In other words an inverse relationship would be expected between student interest in the course and the differentiation of the concept "me" with the concept representing the domain of instruction. Woelfel and Danes (1977) assert that this type of relationship "is predictive of approach behavior" (p. 28). In an educational context, then, such a relationship may be representative of student effort to learn course material. It may even further indicate the probability of students to pursue relevant subject matter beyond the courses in which they are enrolled.

To determine the extent of teaching effectiveness, more specific information is required. It is not enough to merely examine general trends in student learning patterns. A standard for comparison of student progress towards learning appropriate interrelationships among the selected concepts must also be provided. Two primary sources are available for the construction of such a standard. Already knowing the content of the course and its goals, instructors would be in a good position to provide this information. The instructors own meaning for the specific domain of instruction may not be the same as that which would be expected of students. Instructors could instead provide a multidimensional configuration based on their expectations of the responses that could reasonably be expected of students who have mastered the course material. Unless, however, a sizeable group of qualified instructors could be used for this purpose, the procedure could result in substantial errors due to inaccurate projections.

The second source of information is students. That is, the meanings developed by students who have previously been successful in learning specific course material could provide a standard for comparing other students. There may, however, be problems in identifying the appropriate students and in comparing levels of student achievement across time. It would seem, then, that the most desirable standard of comparison would be a weighted average between instructors expectations and previous successful students. This would minimize the extent of errors in instructor estimates and would contain the flexibility for future student performances to exceed or differ from current levels. The weights are dependent upon the number of instructors and students contributing to the standard, changes in the course, and differences in students enrolling in the course.

A standard representing desired course outcomes enables student learning to be more thoroughly assessed. Periodic comparisons of student responses with the standard reveals not only what has been learned but also what needs to be learned. The extent to which students progress toward this standard is, then, a direct measure of teaching effectiveness.

While the procedures described thus far, are more than adequate to precisely assess teaching effectiveness, a thorough evaluation of instruction should also contain more detailed information regarding the specific instructors' classroom behavior. In addition to knowing the extent of student learning that took place in a given course, it is also informative to know what the instructor did or did not do that resulted in the learning that occurred. Multidimenseional scaling has been found to be a useful tool in this type of situation as well (Cody, 1976; Wakshlag and Edison, 1975).

For this purpose, salient teaching characteristics need to be identified and scaled into the multidimensional space along with the concept "instructor" which, of course represents the specific instructor whose teaching performance is being examined. A list of appropriate concepts or characteristics (i.e. clarity, stimulating, effective, rapport) should be obtained from students, instructors, administrators, and from previous research (Deshpande, Webb, and Marks, 1970; Sharon, 1970). These concepts must then be submitted to a Q-sort technique such as the one described by Wotruba and Wright (1975). Essentially what needs to be done here is to locate those concepts which are perceived to be the most important and which students are able to accurately make judgments of. Only concepts that meet both of these criteria as rated by students, instructors, and administrators should be included in the instrument. The lack of consensus would likely result in the obtained information not being fully utilized by all those who could benefit from it.

Along with the obtained concepts and "instructor," a concept entitled "the ideal instructor" could also be used for comparative purposes. The ideal instructor would be described as an individual from which the students perceive that they would maximize their learning potential. In other words, the descriptive teaching characteristics would be used to define an optimally effective instructor. The actual instructor would be compared to the ideal one according to more or less dissimilarity with these concepts. Since any direct comparison between the ideal and actual instructor by the students would be highly susceptible to bias, evaluation instruments should contain one or the other. The comparison can, then, be made more impartially. The information obtained from this procedure could be used to assist instructors in identifying particular teaching skills that need improvement and to assist students in locating instructors with specific desirable teaching styles.

Advantages

The overall advantage of the measurement system presented here is that it possesses at least the potential for overcoming the problems confronting traditional evaluation procedures. Specifically, this system of evaluating instruction (1) is based on a theory of the teaching-learning relationship and as such enables the direct assessment of teaching effectiveness to be made for certain courses. These procedures would appear to be most useful for evaluating instruction in courses that emphasize acquiring knowledge of course material in contrast to courses which stress improving skills. It (2) does not require advance knowledge of all relevant criteria on which student responses are based. This measurement technique assumes that the overall perception of difference precedes the perception of attributes. Consequently, specific attributes used for making distance judgments need not be known nor directly utilized in the measuring instrument. They can, however, be subsequently interpreted from the results. Making complete paired comparisons on all concepts that are provided (3) allows multiple criteria to be simultaneously examined. Moreover, the paired comparison technique (4) is not readily susceptible to the influences of social desirability factors. Doyle (1975) reports that several studies, using traditional Likert type measures, have found that students become more lenient in their evaluations when they are told that the results are to be used for administrative purposes. This is less likely to occur in this situation because actual comparisons to the standard of evaluation are not directly made by students and the scaling procedure provides no inherent indication of how one would go about making an instructor appear more favorably. This evaluation system also (5) has the capacity to incorporate standards for direct evaluation; (6) provides relevant information for administrators, instructors, and students; (7) is precisely ratio scaled; and (8) permits powerful time-series analysis of the learning process. This last advantage is particularly important because it enables instructors to monitor their effectiveness throughout the term. As a result, appropriate modifications in lesson plans and teaching styles can be made.

Limitations

This evaluation system is, however, faced with two potential problems that should be mentioned. One objection to the system has to do with the burden that it places on the students. Multiple measures for several courses could begin to require students to spend a good deal of their time evaluating instruction. Several steps could be taken to minimize this problem. First, samples of students could be taken at each time interval so that all students would not always be asked to participate. Secondly, the instrument could be broken up such that students would only have to fill out a portion of the total number of responses. Finally, specific relational patterns may begin to reappear with a great deal of regularity after the instrument has been in use for awhile.

It would only be necessary to periodically check to make sure that these patterns have not been broken. Consequently the number of required responses would be reduced. Although these remedies help alleviate this problem, they do not eliminate it. Active student participation is necessary to make this system successful. It is expected, however, that their efforts will be rewarded with generally improved teaching and with more detailed information regarding the teaching styles of perspective instructors.

This evaluation system may be difficult to adapt to skill and performance oriented courses. However, the comparison of the actual instructor to the ideal instructor would still provide useful information. Moreover, it would also be of interest to determine certain attitudes such as favorability toward the course's subject matter. After all, training students to be excellent typists would not reflect teaching excellence if as a result, all of the students hated to type and would avoid doing so in the future. It should also be pointed out that the skill and performance oriented courses are in much less need of new evaluation procedures. That is, the extent to which students master the skills taught has always provided a directly observable means of measuring teaching effectiveness.

CHAPTER III

EXPERIMENTAL METHODS AND PROCEDURES

The assessment of a training program for undergraduate teaching assistants was selected for an initial application of the evaluation system developed in the last chapter. It was previously suggested, in Chapter I, that the resistance of instructors to be evaluated may be at least partially due to insecurity resulting from their lack of formal classroom training. This, of course, would not pose a problem unless it was assumed that training produces more effective instruction. In other words, the results of such training are generally expected to be manifested in the actual performance of the instructors in the classroom. In the situation examined in this report, upper division undergraduate communication majors are used to facilitate small group activities in the freshman level introductory communication course. Rather than the traditional teaching methods training, the program established for these teaching assistants emphasized the improvement of leadership skills. Successful training would not only be expected to lead to more student learning but perceptions of teaching or leadership competence should also increase. However, one potentially serious problem should be mentioned. That is, the individuals whose teaching effectiveness is being examined are only teaching assistants. Consequently, any differences in their effectiveness may very well be overshadowed by the influences of the instructors with whom they are working.

Undergraduate Teaching Assistants

Over the past several years, a great deal of time and money has been put into the development of more advanced methods of instruction. Modern technology has enabled educators to increase the quality and quantity of materials that they are able to provide for the large numbers of students in today's schools (Association for Educational Communications and Technology, 1977; Cantwell and Doyle, 1974). The methods used to present these materials are, however, merely extensions of the only two basic techniques that are at the instructor's disposal: directly supplying educational information for student consumption and providing experiences from which the student is expected to master the desired material. This latter technique has been referred to in the literature as the "discovery process" (Ausubel, 1968; Morine and Morine, 1973) and also has informally come to be known as "experiential learning." It has typically been associated with creative and applied areas where such experiences have long been recognized as an appropriate and desirable means of instruction.

Recently, the experiential learning technique has begun to be utilized in the social disciplines (Barbous and Goldberg, 1974; Johnson, 1972; Pfeiffer and Jones, 1971). The field of communication is in an especially advantageous position to adopt this teaching strategy in certain situations. It seems to be a particularly appropriate means to introduce students to the field. That is, many students enter their first communication class with the notion that since they learned as a child to read, write, speak, and listen, and as a result are able to adequately function within their environment, communication is not a complex phenomenon and it can therefore be taken for granted (Williams, 1974). It would seem, then, that a major consideration of any introductory communication course would be to eliminate this misconception. Rather than merely trying to explain that to the students, it would appear to be helpful to also demonstrate this point by allowing students to participate in structured exercises where they will actually be confronted with certain designated communication problems. This is, in fact, the approach taken by many departments around the country.

Introductory courses in communication have recently been faced with a dilemma. That is, exercises in communication typically require close supervision. However, these courses have experienced increased popularity among students and curriculum committees. As a result, the enrollment per class has increased such that the instructors are no longer able to provide the personal attention necessary to successfully carry out these exercises. To help resolve this problem, many instructors have begun to utilize both undergraduate and graduate teaching assistants. The following guidelines were established by the Department of Communication at Michigan State University for the regulation of undergraduate teaching assistants in the introductory communication course:

GUIDELINES FOR THE FUNCTION AND EVALUATION OF THE UNDERGRADUATE TEACHING ASSISTANTS IN COMMUNICATION 100

The undergraduate teaching assistant (UTA) is a student who has taken at a minimum the Communication 100 course as an undergraduate and who has expressed interest in participating as a member of a team of instructors in the teaching of Communication 100. While priority is given to communication-education majors for placement into the UTA program, since it is a requirement for them, communication majors and majors from other areas are welcomed. (Those students whose majors are outside the communication department should be certain that a Communication independent study is acceptable with their department.)

The UTA registers for three credits of independent study (Com 299 for freshmen and sophomores, and Com 499 for juniors and seniors). The graduate student who is the senior instructor of the Communication 100 section in which the UTA works becomes the Independent Study Director and assumes responsibility for the grade given the UTA. To determine the responsibilities and means for evaluation of the UTA, the instructor and UTA will write a contract, which must be signed by the instructor, the UTA, the course chairman and the director of undergraduate study. The exact nature of the contract is determined through conference between the instructor and the UTA. However, it is recommended that <u>specific</u> criteria for successful completion of each grade (A, B, C, etc.) be determined in the contract in order to allow the

UTA to know exactly what he/she is required to do to attain the grade desired, and to provide the instructor with specific guides on which to base the final grade assigned to UTA. It is recommended that the contract identify a variety of teaching experiences for the UTA, such as functioning as a discussion leader, directing an exercise or game, preparing the class to see a movie or conducting a follow-up discussion, etc.

UTA's are required to attend all class sessions (unless a verified reason is provided for not attending a session) and it is expected that the UTA will participate as a member of the instructional team. THE UTA IS NOT TO BE USED AS A READER OR GRADER. While the UTA may grade some papers or examinations (the specific amount to be determined in the contract), a UTA must <u>NOT BE SOLELY RESPONSIBLE</u> for a student's grade. It is advised that both instructor and UTA grade a set of papers and compare their evaluations so that grading is a learning experience for the UTA and that the ultimate responsibility for the grade rests with the instructors.

The UTA is expected to interact fully with the undergraduates and often has been found to act as liaison between the instructor and the students. (In addition, the UTA, like the instructors, will be evaluated by the students in the section through the use of SIRS or alternative forms.)

In total, the UTA opportunity should provide undergraduate students with a teaching-learning-team functioning experience, and should be an asset to the instruction of Communication 100.

Teaching assistants are typically selected according to their: (1) desire, (2) availability, (3) compatibility with the prospective instructor, and (4) quality as a student. None of these criteria are necessarily related to the assistants performance in the classroom. Moreover, this is actually their first experience at formally teaching a group of students. Recognizing this as a potential problem, a training program was developed to provide teaching assistants with the leadership skills that they would need in their classroom role.

The Teaching Assistants Training Program

Taking into consideration the teaching assistants primary responsibility of guiding exercises and discussions, the focus of the workshop is on small group processes. In particular, the function of leadership in these classroom situations is emphasized. The program is essentially structured in the same manner as the classes that they will be assisting. That is, exercises are used to help demonstrate key points as well as to give the participants practice at applying their leadership skills in a situation where constructive feedback is provided. The teaching assistants also have the opportunity to actually participate in many of the same exercises that they will ultimately be using in their classrooms.

The program is offered each term and teaching assistants usually take it in conjunction with the class that they are working with. In order to maximize the benefits of their participation, the workshop is condensed into the beginning of the term. Four three-hour sessions are held one evening a week for the first four weeks of classes.

Session One

The initial meeting is primarily used to get the participants familiar with one another and with the workshop. Several warm-up or get acquainted type exercises are used for the purpose of getting the participants to feel more comfortable and to begin to develop an atmosphere of cohesiveness and trust. This is considered to be an important first step because the teaching assistants can become valuable resources to each other as the term progresses. Moreover, they will be asked to discuss one another's leadership performances in later sessions.

An explanation of what is to transpire at the next three sessions is then provided. In particular, they are told that each of them will be given the opportunity to select and lead a group exercise. This session ends with a discussion of the participants concerns as teaching assistants and suggestion of topics that they would like to have covered in the workshop that weren't originally scheduled.

Session Two

In the second session, the participants are divided into two groups for the purpose of working on a problem solving situation. The designated leader in one group is instructed to be a socio-emotional type leader while the other group has a leader that has been told to be task-oriented. This is done to demonstrate the general differences regarding efficiency and satisfaction that is typically found with groups having these two types of leaders. In addition, confederates acting as deviants are planted into each group to stimulate a discussion of techniques that may be used to deal with disruptive students in the classroom.

The exercise is followed by a discussion of leadership and group dynamics.

Session Three

The third meeting consists of as many of the participants as possible leading the group in an exercise and a follow-up discussion used to draw out the major points covered by the exercise. The group, then, examines each exercise in terms of how well it demonstrated the major points and additional points that could be brought out of the exercise. Suggestions are also made regarding other means of covering the same points. The specific leadership behaviors and strategies of each participant-leader are then discussed. The success of those strategies is assessed and possible alternative strategies are brought up.

Session Four

The remainder of the participants who have not previously had the opportunity to lead a group exercise are able to do so in this final session. The procedures are the same as in the third session. In addition, lecturing tips and other such topics of concern to teaching assistants are attempted to be squeezed in. The participants finish off the workshop by exchanging the experiences that they have already had in the classes that they are assisting.

In summary, the training program briefly described here is designed to prepare undergraduate teaching assistants for the classroom by increasing confidence in and improving their leadership skills as well as providing them with a general understanding of small group processes. In addition, the assistants become thoroughly familiar with the use of exercises as a teaching devise.

The Introductory Communication Course

The particular introductory communication course that the teaching assistants in this study worked with is essentially a survey course. That is, the course attempts to represent most of the various interest areas which make up the discipline. The following general areas are covered: basic concepts and models, the nature of meaning, code systems, message construction, public speaking, interpersonal relations, small group interaction, organizational systems, and mass media. In the course, students are provided with a brief and rather narrow exposure to each of these areas. In addition, the course focuses attention on communication skills. Students are required to give several oral presentations and to write several papers.

It would appear that for such a course to be successful, one of its overriding concerns must be in stimulating student interest not only in the course itself but also in what the department has to offer. A helpful first step in this direction would be to demonstrate that there is, in fact, something important to be studied. That is, students must be instilled with the notion that communication is a much more complex process than they had previously thought. Building on this, the students must also gain an understanding of how knowledge of this complexity can be beneficial to them. Exercises are typically used as a teaching technique to help make these points. Students frequently find this type of experience enjoyable as well as informative. Regardless of the method used, if the points are successfully made, they would be expected to stimulate student interest in learning about the communication scholars, then, provides direction for students interested in pursuing courses in the department beyond the introductory level.

Measuring Instrument

Given the previous accounts of the functions of undergraduate teaching assistants, the training that they receive, and the nature of the course in which they work, attention can now be focused on the development of the specific measuring instrument (see Appendix A) to be used for the purpose of evaluating their effectiveness in the classroom. Teaching effectiveness was defined in Chapter II as the progress of students in learning the appropriate interrelationships among the concepts which constitute the domain of meaning for the course. The task at hand is, then, to select the relevant concepts for this particular course and workshop. As an introductory course, there are many more concepts than could be considered in this instrument. The general meaning must be considered to be important. As a result, the following concepts were included in the instrument: communication, humanities, physical science, and social science. Since the academic approach of this department is a social scientific one, it is expected that if instruction was successful, communication and social science would become more strongly associated with one another as the term progresses. To examine more specific content, leadership and democratic were also included. Neither of these concepts are intuitively thought of as communication concepts. They are, however, taught as such in the group interaction section of the course. Effective teaching should move them both closer to communication in the cognitive structure. Moreover, they should move closer to each other since democratic is taught as a specific leadership style.

It was suggested in the previous chapter, in addition to student learning, the stimulation of student interest in the general topic area covered by the course should also be considered as a criterion of teaching effectiveness. This is, in fact, one of the overriding goals for the particular course examined in this study. The stimulation of student interest would be reflected by a reduction of the perceived differentiation between the concepts communication and me (representing the students' self concept). In other words, this change is assumed to reflect increased student interest in communication. To obtain an overall value judgment of communication, the concept good was added. A more favorable student attitude toward communication would be demonstrated by a convergence of communication and good.

In order to more directly evaluate the training program itself, it would be useful to more closely examine the actual classroom performances of the teaching assistants. For this purpose, the teaching assistants and the teachers for whom they work were included in the instrument. Since the teacher would presumably serve as a role model for the assistants, the differentiation between the two would reflect how well the assistants were carrying out their role relative to some standard. The teacher probably does not represent the ideal standard, but the goal of perceived equivalence with the teacher would in most cases be considered a substantial step in that direction. While leadership and democratic were previously used in the assessment of student learning, they can also be used here to provide more specific information regarding the teaching assistants. That is, effective assistants would be expected to be more strongly associated with leadership. Since the democratic style was presented in the workshop as the most appropriate for facilitating group exercises and discussions, effective assistants should also be more closely identified with democratic in the cognitive structure. The final two concepts that were included to reflect the assistants classroom performance are "confident" and "expertise." Assuming that confidence is behaviorally manifested, teaching assistants participating in the training program would be expected to display more confidence. The workshop discussions on group processes and leadership should make the participants more knowledgeable in the area that they will be working. Moreover, the importance placed on the follow-up discussions may stimulate them to better understand the material covered by the exercises they use. Participants in the workshop should, then, be more closely associated with expertise.

To put the questionnaire in more of a general education frame of reference, the concepts studying, thinking and learning were also included in the instrument. While no specific expectations are made regarding these concepts, they may be useful in interpreting the results. In all, the following sixteen concepts were incorporated into the complete paired-comparisons format on the questionnaire: communication, humanities, physical science, social science, good, me, instructor, undergraduate teaching assistant (2), leadership, democratic, confident, expert, thinking, studying and learning. The ordering of the pages on which these concepts appeared was alternated to equally distribute the possible influence of fatigue on the part of the respondents. The questionnaire also asked the respondents to provide basic demographic information as well as information regarding their educational backgrounds.

Research Design

This evaluation is primarily based on the idea that the success of the program should be determined by the actual classroom performances of the teaching assistants. Twenty sections of the introductory course were offered the term that this study was conducted. The teaching assistants from ten of these sections were randomly selected to participate in the training program while the assistants in the other ten sections served as the control group. The division was made according to section rather than individual assistants because six sections had two assistants. It was felt that to have one assistant participating in the workshop and the other not participating might cause some unnecessary problems. That is, the untrained assistant may indirectly benefit from the workshop by picking up on what the trained assistant had learned. Moreover, repeated exposure to both assistants could make it difficult for students to make a clear distinction when assessing the assistants separately. The repeated exposure to both assistants would also make it difficult to attribute the extent of student learning to the teaching effectiveness of one teaching assistant or the other.

The control group consisted of fourteen teaching assistants, nine male and five female, with a mean age of twenty-two. Their grade point averages ranged from 2.4 to 3.9 with 3.1 as the mean. Ninety percent of the group were juniors or seniors. Sixty-nine percent were majoring in communication. In contrast, there were twelve assistants in the experimental group. Four of them were males and eight were females having an average age of twenty. Their grades went from a low of 2.5 to a high of 3.8. The mean grade point average for this group was 3.2. Seventy-three percent were communication majors. Eighty-nine percent were upper-division students.

The ten sections whose assistants were in the control group were found to be quite similar to those sections whose assistants participated in the training programs (see Appendix C for specific comparisons). In the sections from both groups, class time was just about equally distributed between lectures (32%), class discussions (30%) and exercises (36%). The number of students ranged from 38 to 71 in the sections represented in the control group and from 34 to 76 in the experimental sections. The mean for both groups was 63 students. The mean age of the students was 19.1 and 19.0 for the experimental and control group sections respectively. Their respective grade point averages were 3.05 and 3.02. In both groups, there was approximately a 2 to 1 ratio of female to male students. Seven percent of the students in the control group sections were majoring in communication. The experimental group sections had ten percent communication majors. The only other sizable group of students were those who had not yet made an academic preference;

26% in the control sections and 19% in the experimental sections. This is not particularly surprising since the course is at the introductory level and high percentages of freshman (58% in control sections and 49% in experimental sections) were found to be enrolled in it. Only 28% of the students in sections of both the experimental and control groups were in the upper division.

The descriptive data discussed thus far suggests that the experimental and control groups are comparable in regard to the teaching assistants, the structure of their classes and the students in those classes. The differences that exist are either minor or would not be expected to influence the results of this study. There are, however, three areas in which the two groups differ that is a cause for concern. First, 85% of the teachers that the assistants are working for in the control group have had previous teaching experience while only 54% of the teachers in the experimental sections are experienced teachers. Moreover, teaching assistants had been previously utilized by 62% of the teachers in the control sections. Only 29% of the experimental group teachers had used teaching assistants in the past. Finally of the teaching assistants themselves, 38% in the control group and 21% in the experimental group have had some sort of past experience as a teaching assistant or small group leader. In each of these areas, the control group would appear to benefit from more experienced teaching backgrounds. This essentially provides for a more rigorous test of the training program's effectiveness. That is, for the training program to appear successful, the trained assistants have to be more effective in the classroom than the untrained assistants who have more experience and work with teachers who also have more experience.

The problem of unequal past teaching experience is somewhat neutralized by the longitudinal nature of the study. At approximately three week intervals, data was collected at four points in time during the term. Thus the assistants teaching effectiveness can be evaluated in terms of growth and improvement over time.

In summary, the current study evaluates a training program for teaching assistants by examining the teaching effectiveness of a group of trained assistants and a group of assistants that were not trained. The evaluation concentrates on student learning and actual classroom performance as the criteria for assessing the program's success. Data was gathered four times during the term from student volunteers.

CHAPTER IV

RESULTS

The major findings of this study will be discussed in two parts. The first and primary analyses will examine the data in a manner that will attempt to illuminate the procedures used to evaluate instruction. As was stated in Chapter I, the study reported here served as an initial application of a new set of procedures developed for instructional evaluation. Of specific interest is, then, an exploration into the success of these procedures in assessing the teaching effectiveness of this particular group of teaching assistants. This is the principle focus of the study.

It should be explicitly pointed out, however, that this study is a demonstration of how instruction might be evaluated under ideal circumstances. The particular data set used here is really not appropriate for the analyses that will be performed. Specifically, every available student was asked to volunteer at each data collection. Many students participated more than once but not at all four points in time. Consequently, there were insufficient sample sizes for a normal panel design. Moreover, the samples were not randomly drawn. Thus, the data was correlated at least to the extent that some subjects participated in more than one data collection. The analysis performed here does, for illustrative purposes, treat the data as if independent random samples, without replacement, were drawn at each time interval.

Sample sizes per section for the experimental and control groups for each data collection are presented in Table 1. In the primary analyses, individual student responses were aggregated across all sections for the experimental and control groups. In the secondary or adjunct analyses, the data will be re-examined to enable more general statements to be made regarding the teaching-learning process. For these analyses, as well as for the examination of the manipulations and the precision of measurement, each of the sections that had students participating in all four data collections will serve as a unit of analysis. The individual student responses within a section will provide an estimate for that particular section. The scores for each section should as a result be fairly stable. Each of the appropriate sectional scores will then be averaged for the experimental and control groups, and comparisons made on that basis.

In the primary examination of data, attention will first be given to an exploration of teaching effectiveness.

Table 1. Sample Sizes per Section for the Experimental and Control Groups at Four Time Intervals.

	T_1	T_2	T_3	T_4
<u>Experimental Group</u>				
Section 1	18	14	7	8
Section 3	7	24	16	15
Section 6	24	0	1	4
Section 8	7	4	2	1
Section 9	19	20	10	13
Section 12	20	16	2	9
Section 15	4	11	0	9
Section 16	10	0	0	0
Section 19	20	8	3	0
Section 21	20	4	7	11
	149	101	48	70
<u>Control Group</u>				
Section 2	15	6	7	5
Section 4	18	10	6	4
Section 5	23	6	6	3
Section 7	13	18	2	4
Section 10	10	8	4	13
Section 11	4	2	1	14
Section 13	11	42	12	16
Section 14	9	13	6	0
Section 18	21	18	8	6
Section 20	3	2	1	3
	127	125	53	68

More specifically, this set of analyses will examine differences in the extent of student learning that occurred in courses aided by trained teaching assistants in contrast to courses aided by teaching assistants who received no training. The integration of course material as reflected in the following six paired comparisons will be analyzed:

1. Communication and Social Science
2. Communication and Leadership
3. Communication and Democratic
4. Leadership and Democratic
5. Communication and Good
6. Communication and Me

The first four of these pairs focuses specifically on student consumption of course content while the latter two are primarily concerned with assessing more general student attitudes regarding communication. For the purpose of providing a standard for determining student progress, a criterion score for each pair was established. Because the teaching assistants had already successfully completed the course and the instructors presumably understand the course material, their judgments of the above items were pooled to create this criterion. The small sample size, however, makes the utility of the criterion scores for this particular study somewhat questionable. While specific comparisons with the criterion scores will be made, these scores will primarily serve in more of a directional capacity rather than in any absolute sense. The thrust of these analyses will be on examining differences in student learning between the experimental and control group classes.

The second set of analyses focuses more directly on the teaching assistant's actual classroom behavior. The emphasis here is on the relationship between the teaching assistants and several key attributes or characteristics which are fundamental to their role in the classroom. The following paired comparisons will be investigated for this purpose:

1. Teaching Assistant and Expert
2. Teaching Assistant and Leadership
3. Teaching Assistant and Democratic
4. Teaching Assistant and Confident
5. Teaching Assistant and Instructor

This set of analyses also contains a standard for comparison. Student responses regarding dissimilarities between the instructors and these same characteristics were averaged across experimental and control groups at each point in time to create specific criterion scores for each pair. Since there would be no dissimilarity between the instructors and themselves, the teaching assistant-instructor pair was set, by definition, at 0.00. The instructors were themselves relatively inexperienced and as such do not provide the ideal standard suggested in Chapter II. However, it was felt that perceived equivalence with the teacher, while not the ideal standard, would in most cases be considered a substantial step in the appropriate direction. Consequently, the criterion scores for these analyses will serve to provide direction rather than any definitive comparison. In other words, the emphasis will again be on differences between the experimental and control groups. The criterion scores will then be an indication of whether those differences reflect more favorably on one group or the other. Prior to discussing these two sets of analyses, it is important to first examine the measurement system and the success of the experimental manipulations.

Precision of Measure

Theoretically, the concepts not directly manipulated in the experiment would be expected to remain constant across all experimental conditions. That is, the means for each of the paired comparisons among these not intentionally affected concepts should be the same in both the experimental and control groups as well as at each point in time. Differences are attributable to error of measurement. Averaging first across all groups at each time interval enabled standard errors of measure to be established by the standard deviation for each of the unmanipulated pairs. The coefficients of variability were then calculated, according to the following equation: $V = 100\ (S_{xi}/\bar{X})$, and then averaged over the four time intervals. The resultant coefficient represents the percentage of measurement error for each pair (Woelfel, Cody, Gillham and Holmes, 1977). These coefficients for all paired comparisons (manipulated and unmanipulated) are presented in Table 2.

The coefficients for all pairs ranged from 27.00% to 85.76% with a mean of 51.31%. The mean for the unmanipulated concepts was 41.86% with the largest coefficient (V = 61.26%)

Table 2. Coefficients of Variability ($V = 100\ (S/\bar{X})$) for Mean Dissimilarities of Experimental and Control Groups Averaged Over Four Points in Time

	Expert	Instructor	Studying	Thinking	Me	TA	Physical Science	Leadership	Learning	Democratic	Humanities	Good	Confident	Social Science	Communication
Expert															
Instructor	52.15														
Studying	49.98	50.53													
Thinking	54.86	44.80	51.00*												
Me	54.73	52.66	51.58	50.68											
TA	44.46	34.83	56.30	83.28	63.05										
Physical Science	54.01	45.13	39.40*	41.02*	53.52	53.50									
Leadership	56.82	55.39	50.99	56.67	48.12	39.90	48.19								
Learning	51.89	52.90	44.09*	61.26*	53.55	75.35	42.02*	56.08							
Democratic	48.75	54.78	45.43	55.74	48.91	41.02	57.62	39.12	51.92						
Humanities	49.39	49.09	45.45*	31.56*	49.17	60.96	27.00*	51.72	40.78*	50.70					
Good	60.58	56.92	61.58	52.82	51.99	62.41	52.84	61.80	42.25	60.23	43.81				
Confident	58.55	56.99	45.75	57.60	44.40	41.98	52.96	51.77	64.10	49.24	42.61	54.62			
Social Science	48.43	52.30	51.00	53.85	47.15	60.60	48.88	51.42	50.14	54.99	49.41	44.17	54.11		
Communication	48.53	55.98	54.00	55.60	34.94	85.76	56.69	42.83	68.71	36.82	51.09	43.52	52.20	37.58	

* = unmanipulated pairs

being associated with the learning-thinking paired comparison. This result along with the relatively large coefficients found for other similar pairs such as thinking-studying (V = 51.00) and studying-learning (V = 44.09) are not surprising in light of the sample which primarily consisted of freshmen entering a new social and educational environment. This suggests that the coefficients of variability for the unmanipulated concepts in this study reflect more than merely error of measure. Previous research (Woelfel et al., 1977) has found coefficients averaging below 10% for unmanipulated concepts as vague and illusive as the ones used here. The magnitude of the coefficients in this study may very well be a result of the effects of the overall educational system. In other words, this study deals with only a portion of the educational environment to which the student-respondents were exposed. Educational influences outside the realm of this study should, then, be entirely expected to increase the variability of responses. Thus, the error of measurement in this study is likely to be at least somewhat less than what is reflected by the coefficients of variability.

Manipulations

The coefficients of variability can also be used to examine the success of the experimental manipulations. Coefficients for the manipulated pairs are expected to reflect differences across conditions and time resulting from experimental manipulations in addition to error of measure.

Assuming that measurement error is reasonably equivalent for manipulated and unmanipulated pairs, differences in the coefficients for these two groups are directly attributable to manipulation effects. The mean coefficients of variability for the 11 manipulated pairs ($\overline{V}_m$) and the 10 paired comparisons ($\overline{V}_u$) which contained no experimentally manipulated concept were computed. All paired comparisons involving the concepts instructor or me were excluded from this analysis because these two concepts, while not directly manipulated, would be expected to change over the three month period of this study. To examine differences between the manipulated and unmanipulated concepts, the Behrens-Fisher statistic for two means with unequal population variances was calculated. It should at this point be mentioned that the use of this statistic in this context is illustrative of what would be appropriate for the ideal case. Since the section was used as the unit of analysis, for the purpose of providing more stable means for the experimental and control groups, the samples were not randomly drawn. Moreover, as was pointed out at the beginning of this chapter, the participation of some of the same subjects who contributed to the calculation of their sectional scores, in more than one of the four data collections makes the data correlated rather than independent. The Behrens-Fisher statistic yielded no significant difference between the mean coefficients of variability for the manipulated and unmanipulated pairs. This, however, does not necessarily indicate that the manipulations were unsuccessful. As was previously stated, the influence of the educational environment was beyond the control of this study and consequently expected to contaminate the data. Some support for this contention is provided in Table 3 which reports the mean variability coefficient for each concept paired with all other concepts. The mean variability coefficient was obtained by averaging the individual variability coefficients, found in Table 1, for each concept paired with the remaining fourteen concepts. In other words, the unit of analysis here is the concept rather than the student respondents. Learning and thinking, two concepts which could be anticipated to undergo change in meaning as a result of exposure to a new educational system, had two of the highest coefficients. Teaching assistant, which was the most substantially manipulated concept, had by far the largest average variability coefficient. This, however, is not enough to warrant any claims about the effectiveness of the teaching assistant manipulation. Instead, these findings seem to suggest that outside educational influences were sufficient enough to prevent any conclusive statements regarding the success of the experimental manipulations.

Table 3. Means and Standard Deviations for the Variability Coefficients of Each Concept

Concept	Mean	Standard Deviation
Teaching Assistant	57.40	16.14
Learning	53.94	10.42
Thinking	53.64	11.48
Good	53.54	7.50
Expert	52.38	4.48
Confident	51.92	6.50
Communication	51.73	13.52
Instructor	51.04	6.07
Leadership	50.77	6.66
Me	50.33	6.23
Social Science	50.29	5.38
Democratic	49.65	7.03
Studying	49.44	6.00
Physical Science	48.05	8.42
Humanities	45.56	8.72

Student Learning

Although the Behrens-Fisher test revealed no significant differences in the mean scores for the course content related paired comparisons between students in the control and experimental group classes, several informative trends can be found in the data. In first focusing attention on general tendencies over time (Tables 4-7), it appears that what substantial changes did occur took place toward the latter part of the term. There were only minor differences between the scores in the time one and time two data sets for both the experimental and control groups. Beyond this similarity, however, the two groups tended to follow quite different patterns. For the experimental group, not only are the first two data sets similar to one another, but, in addition, the scores from the fourth point in time resemble the first two. The only distinction being that in all cases the final mean score was smaller than the initial one. Changes in this group appear, then, to have been centralized around the third data collection. More specifically, the dissimilarities between the selected pairs of content concepts generally show an increase from time two to time three which is followed by a decrease at the fourth point in time. The communication-good pair, for example, jumps from a mean score of 42.35 at time two to 71.52 at time three and then comes back down to a final score of 36.97. This pattern is clearly illustrated in Figures 1 and 2.

While a similar tendency was found in the control group for the communication-leadership and leadership-democratic pairs, the remaining paired comparisons do not seem to follow any single consistent pattern over time. If anything, there merely seems to be much less variation between the mean scores over the four time intervals. Any sizable differences that were found are associated with

Table 4. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Content Concepts for Experimental and Control Groups at Time One

	Condition		
Concept Pair	Criterion	Control	Experimental
Communication and Social Science	33.163	45.948	54.383
	(36.136)	(42.894)	(117.917)
	86	116	128
Communication and Leadership	54.124	30.137	29.669
	(132.191)	(53.357)	(80.191)
	89	117	133
Communication and Democratic	50.707	55.772	60.869
	(51.048)	(96.727)	(126.757)
	92	123	145
Leadership and Democratic	59.781	59.017	51.481
	(98.391)	(76.643)	(102.888)
	96	117	135
Communication and Good	31.552	48.155	40.556
	(40.390)	(101.061)	(99.977)
	87	116	124
Communication and Me	25.393	40.127	38.276
	(30.844)	(57.396)	(81.571)
	89	118	134

Table 5. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Content Concepts for Experimental and Control Groups at Time Two

	Condition		
Concept Pair	Criterion	Control	Experimental
Communication and Social Science	33.163	48.477	46.137
	(36.136)	(44.541)	(58.601)
	86	111	95
Communication and Leadership	54.124	34.638	30.379
	(132.191)	(48.621)	(51.682)
	89	105	95
Communication and Democratic	50.707	44.374	54.040
	(51.048)	(35.794)	(70.971)
	92	123	99
Leadership and Democratic	59.781	38.372	38.958
	(98.391)	(41.670)	(35.304)
	96	121	95
Communication and Good	31.552	52.018	42.354
	(40.390)	(95.650)	(92.557)
	87	111	96
Communication and Me	25.393	47.138	30.800
	(30.844)	(108.129)	(55.189)
	89	109	90

Table 6. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Content Concepts for Experimental and Control Groups at Time Three

Concept Pair	Condition		
	Criterion	Control	Experimental
Communication and Social Science	33.163	61.186	70.761
	(36.136)	(77.789)	(142.323)
	86	43	46
Communication and Leadership	54.124	57.167	40.733
	(132.191)	(146.246)	(81.536)
	89	48	45
Communication and Democratic	50.707	47.392	59.426
	(51.048)	(47.576)	(78.669)
	92	51	47
Leadership and Democratic	59.781	71.907	59.509
	(98.391)	(166.302)	(83.900)
	96	54	53
Communication and Good	31.552	52.553	71.522
	(40.390)	(143.002)	(156.934)
	87	47	46
Communication and Me	25.393	35.872	29.104
	(30.844)	(32.830)	(45.477)
	89	47	48

Table 7. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Content Concepts for Experimental and Control Groups at Time Four

Concept Pair	Condition		
	Criterion	Control	Experimental
Communication and Social Science	33.163	63.905	53.892
	(36.136)	(137.827)	(109.399)
	86	63	65
Communication and Leadership	54.124	41.077	23.986
	(132.191)	(123.873)	(39.771)
	89	65	70
Communication and Democratic	50.707	79.229	52.886
	(51.048)	(158.290)	(106.315)
	92	70	70
Leadership and Democratic	59.781	37.644	43.761
	(98.391)	(41.680)	(105.273)
	96	73	71
Communication and Good	31.552	42.769	36.970
	(40.390)	(66.732)	(64.134)
	87	65	66
Communication and Me	25.393	30.934	26.500
	(30.844)	(30.067)	(36.440)
	89	61	68

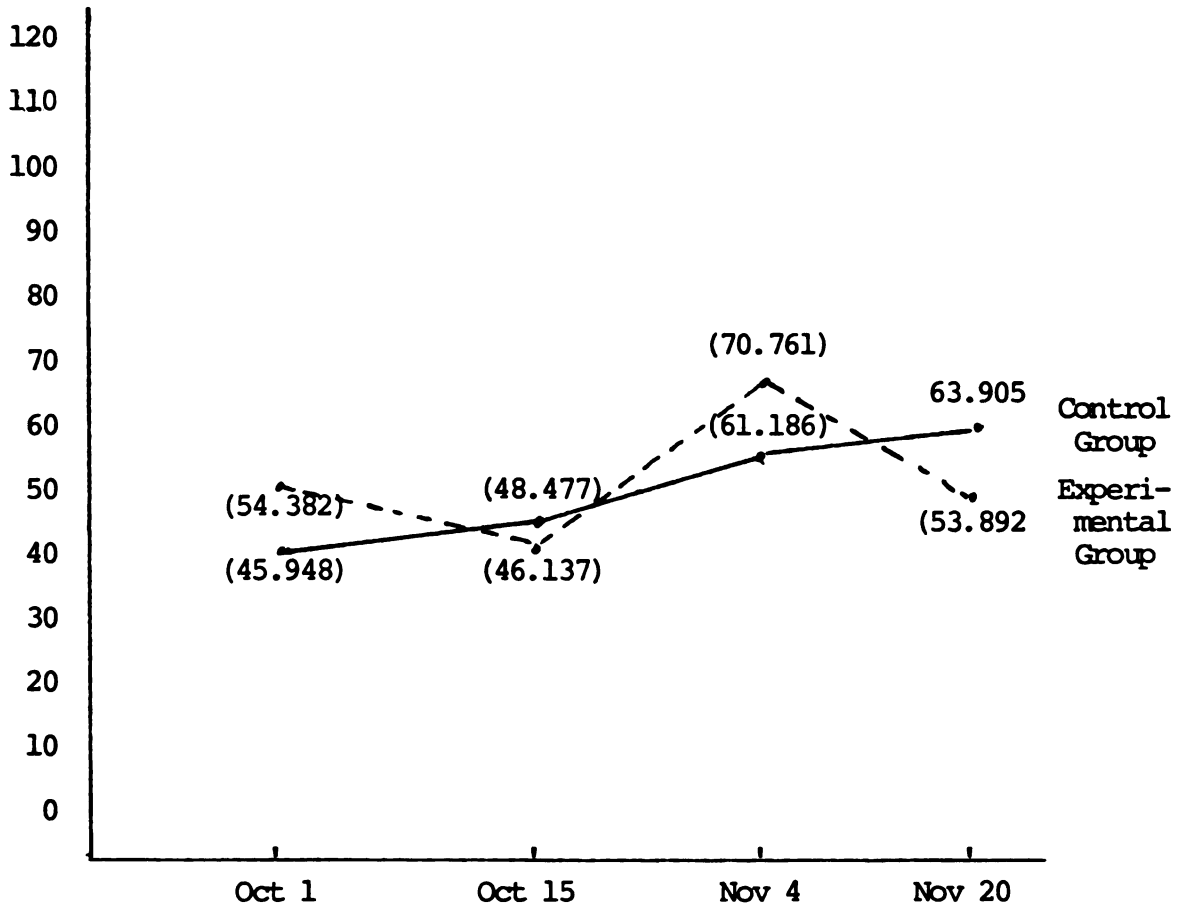

Figure 1. Plot of the Communication-Social Science Paired Comparison for Experimental and Control Groups Across Four Points in Time

Key: ——— Control Group

----- Experimental Group

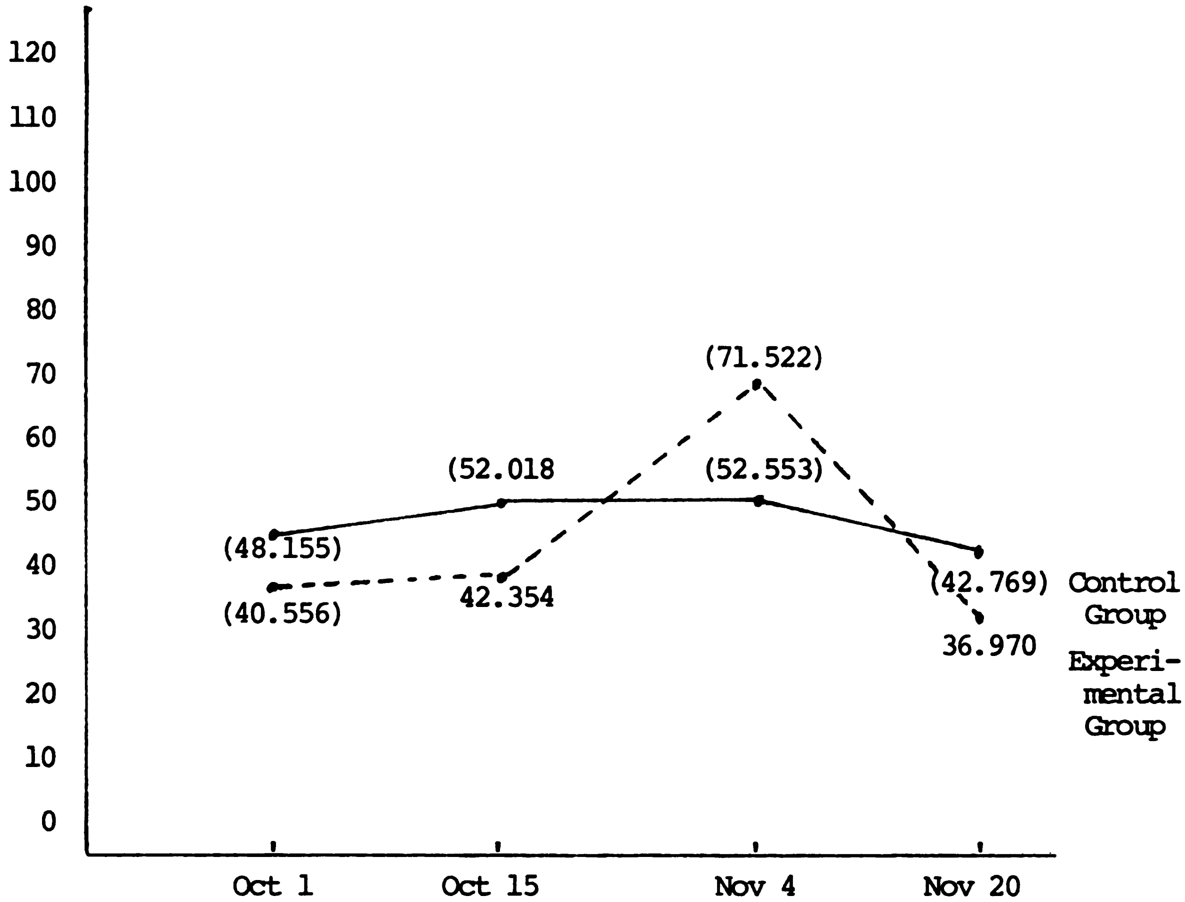

Figure 2. Plot of the Communication-Good Paired Comparison for Experimental and Control Groups Across Four Points in Time

Key: ——— Control Group

----- Experimental Group

pairs involving the concepts leadership and democratic, or, in other words, the more directly taught course material.

There are several plausible explanations for these findings. First, two of the concepts, leadership and democratic, were the only ones being examined which were actually taught to the students as specific course material. It is not surprising, then, to see pairs including these more directly manipulated concepts to experience more variation. Moreover, these concepts were not taught until just before or after the time of the third data collection. Consequently, the three paired comparisons which include at least one of these concepts would be expected to have remained relatively stable in the early part of the term. A second explanation deals with the all encompassing nature of several of the paired comparisons being examined. The relationship between communication and social science, for example, may not become particularly clear to students until they have had at least a minimal exposure to the course. In other words, it may have taken a while for the students to put specific course content into an overall assessment of where the field of communication fits into the larger academic picture. More specific content pairs should, then, show more immediate results. These findings may also have been a result of the execessive variability found in the data. The high standard deviations reported in Tables 4-7 suggests that the mean scores are probably not stable enough to obtain a clear picture of the differences between the experimental and control groups. A final contributing factor to these results may be found in the response burden placed on the student participants. Although participation was voluntary and the students presumably became more proficient in filling out the questionnaires, they may have become apathetic or even a little put out by the longevity of the project. This is at least somewhat reflected in the sizable decrease in sample sizes for the final two, and particularly the third, data collections. Moreover, the lack of participant awareness regarding the value of repeated measures may have also contributed to less care being exercised in filling out questionnaires.

Prior to making more specific comparisons between the experimental and control groups, one last general finding should be briefly mentioned. It should, however, first be more explicitedly pointed out that the criterion scores are assumed to represent a standard that students appropriately learning course material should proceed toward. Of the six paired comparisons reflecting course content, only one, namely the communication-me pair (Figure 3), experienced relatively continuous movement toward the criterion score. While it is tempting to discuss the implications for this particular pair, it seems more appropriate to suggest that

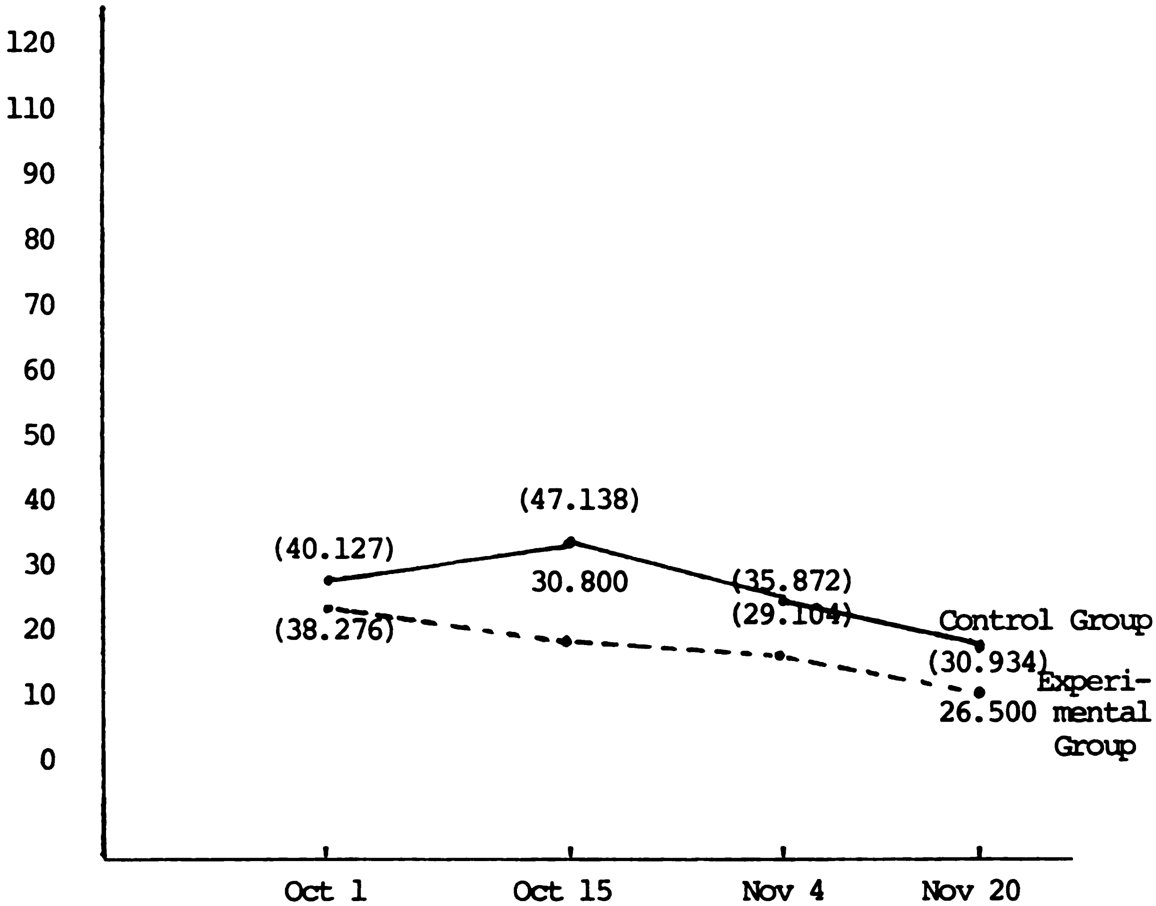

Figure 3. Plot of the Communication-Me Paired Comparison for Experimental and Control Groups Across Four Points in Time

Key: ——— Control Group

----- Experimental Group

these findings demonstrate the complexity of the learning process.

A more thorough examination of the data relative to specific criterion scores reveals several noteworthy distinctions between the experimental and control groups. The mean scores of three of the six pairs in the final data set for the experimental group were closer to their corresponding criterion scores than at any other point in time. However, one of the remaining scores was at that time further from the criterion score than it had ever previously been. For the control group, just two of the six pairs had final mean scores that were closer than at any of the three preceeding time intervals. Moreover, three of the other four paired comparisons had final scores that were then furthest away from their respective criterion scores.

At the initial data collection (Table 4), scores from only two of the six pairs for the experimental group were closer to the criterion scores than their counterparts in the control group. On the other hand, at the fourth and final point in time (Table 7), the scores from the experimental group were closer to their specific criterion score in five of the six content pairs.

Although these results seem to slightly suggest that students in the experimental group classes progressed more than students in the control group classes, they are, at best, only a weak indication in that general direction. It should again be pointed out that there were no significant differences found between the experimental and control groups and that the criterion scores were at least somewhat questionable. Thus, no conclusive distinction can be made regarding the extent of student learning that occurred in the experimental and control group classes.

Instructional Performance

The time one through time four data sets for the instructional performance paired comparisons are reported in Tables 8 through 11 for the control and experimental groups. Prior to examining these data, it should first be pointed out that the sample sizes shown in these tables usually exceed the number of student participants reported in Table 1. Although this appears to be an inconsistency, it is actually a result of several of the class sections having more than one teaching assistant. More specifically, two sections in the experimental group and four sections in the control group had two assistants. Consequently, each of the student volunteers in those sections responded to the teaching assistant-attribute pair twice, once for each teaching assistant.

The fact that several of the subjects made assessments of the instructional performance of more than one teaching assistant poses a problem with the analysis of these data. It was previously pointed out that the data across time was correlated because several subjects participated at more than

Table 8. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Instructional Concepts for Experimental and Control Groups at Time One

	Condition		
Concept Pair	Criterion	Control	Experimental
Teaching Assistant and Expert	56.586	116.045	92.924
	(82.305)	(193.677)	(159.082)
	251	137	105
Teaching Assistant and Leadership	35.498	50.351	53.610
	(77.221)	(61.432)	(116.175)
	243	165	164
Teaching Assistant and Democratic	63.134	54.242	65.050
	(112.021)	(41.522)	(125.295)
	245	153	158
Teaching Assistant and Confident	39.869	42.092	57.426
	(84.018)	(38.032)	(122.463)
	237	163	162
Teaching Assistant and Instructor	0.000	57.090	50.995
		(98.371)	(115.575)
		166	176

Table 9. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Instructional Concepts for Experimental and Control Groups at Time Two

Concept Pair	Condition			
	Criterion	Control	Experimental	p
Teaching Assistant and Expert	47.751 (66.915) 209	108.774 (207.141) 164	55.376 (40.401 133	.01
Teaching Assistant and Leadership	45.062 (102.259) 203	83.537 (185.220) 160	45.748 (57.982) 123	.02
Teaching Assistant and Democratic	46.333 (81.879) 207	94.023 (211.379) 164	60.114 (108.890) 122	ns
Teaching Assistant and Confident	33.792 (42.758) 197	74.032 (159.258) 157	45.168 (63.957) 125	.05
Teaching Assistant and Instructor	0.000	64.218 (151.292) 165	35.303 (59.600) 135	.05

Table 10. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Instructional Concepts for Experimental and Control Groups at Time Three

Concept Pair	Condition		
	Criterion	Control	Experimental
Teaching Assistant and Expert	70.649	99.356	66.507
	(133.570)	(183.896)	(57.965)
	97	73	73
Teaching Assistant and Leadership	48.536	89.780	58.203
	(80.246)	(183.896)	(76.486)
	97	73	69
Teaching Assistant and Democratic	58.980	91.218	74.042
	(92.059)	(180.133)	(134.857)
	103	78	71
Teaching Assistant and Confident	44.917	62.974	58.132
	(61.176)	(96.176)	(121.434)
	97	76	68
Teaching Assistant and Instructor	0.000	46.618	47.821
		(51.140)	(44.501)
		76	39

Table 11. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Instructional Concepts for Experimental and Control Groups at Time Four

	Condition			
Concept Pair	Criterion	Control	Experimental	p
Teaching Assistant and Expert	52.153 (93.507) 131	87.417 (147.213) 96	57.240 (68.060) 96	ns
Teaching Assistant and Leadership	33.075 (36.334) 139	60.071 (104.591) 99	33.848 (26.131) 92	.02
Teaching Assistant and Democratic	59.137 (137.246) 141	77.758 (138.894) 99	43.663 (29.432) 92	.02
Teaching Assistant and Confident	35.474 (60.381) 135	51.134 (76.479) 97	34.194 (30.817) 93	ns
Teaching Assistant and Instructor	0.000	49.320 (90.556) 97	39.638 (39.402) 94	ns

one point in time. The issue raised here is a somewhat different problem of correlated data. Rather than data being correlated across time, the data is instead correlated within each of the time intervals. The problem is essentially that each response to the teaching assistant-attribute pairs is not independent. The responses given by each subject who rated two assistants are interrelated and as such violate the statistical assumption of independent observations. Thus, the use of the Behrens-Fisher statistic in this context should be taken as an illustration of the analysis that should be performed when independent observations have been taken. It should be mentioned that most courses have only one instructor. This problem would, then, not be expected to be present in most educational situations.

As was the case in the previous set of analyses, it seems to also be informative here to first explore general trends in the data over time. The only distinctive trend that appears to hold across both groups is the presence of a consistent decrease in the mean scores from the third to the fourth point in time. Focusing attention on the experimental group data first, it seems that while in the previous analyses, the mean scores experienced relatively large variability over time, the scores for the instructional performance pairs are comparatively more stable. With the exception of a substantial decrease for the teaching assistant-expert pair from time one to time two, the rest of these data seem to follow the same general pattern as the content pairs. The changes that occur are, however, much less substantial. There are, then, minimal differences between time one and two scores, with a slight tendency for the latter to be smaller. An increase at time three is then followed by a decrease at the last data collection which makes the final mean score always smaller than its corresponding time one score; this can be seen in Figures 4 and 5.

The control group had more variation for the instructional performance pairs. There was, however, no uniform pattern for the six pairs over time. Probably the closest thing to an overall pattern is illustrated in Figure 4 where the mean score for the teaching assistant-leadership pair increases from time one to time two, then generally levels off between times two and three, and finally decreases from the third to the fourth data sets. Although there are noticeable deviations, the data from the six pairs, when taken together, roughly fit this pattern.

Differences between these general patterns for the control and experimental groups can be at least to some extent attributable to the training program. When the initial set of data was collected, only the first week and a half of the term had been completed and only the first of the four training sessions had been run. It is not surprising, then, to see that mean scores for the two groups are initially quite similar. Moreover, the differences

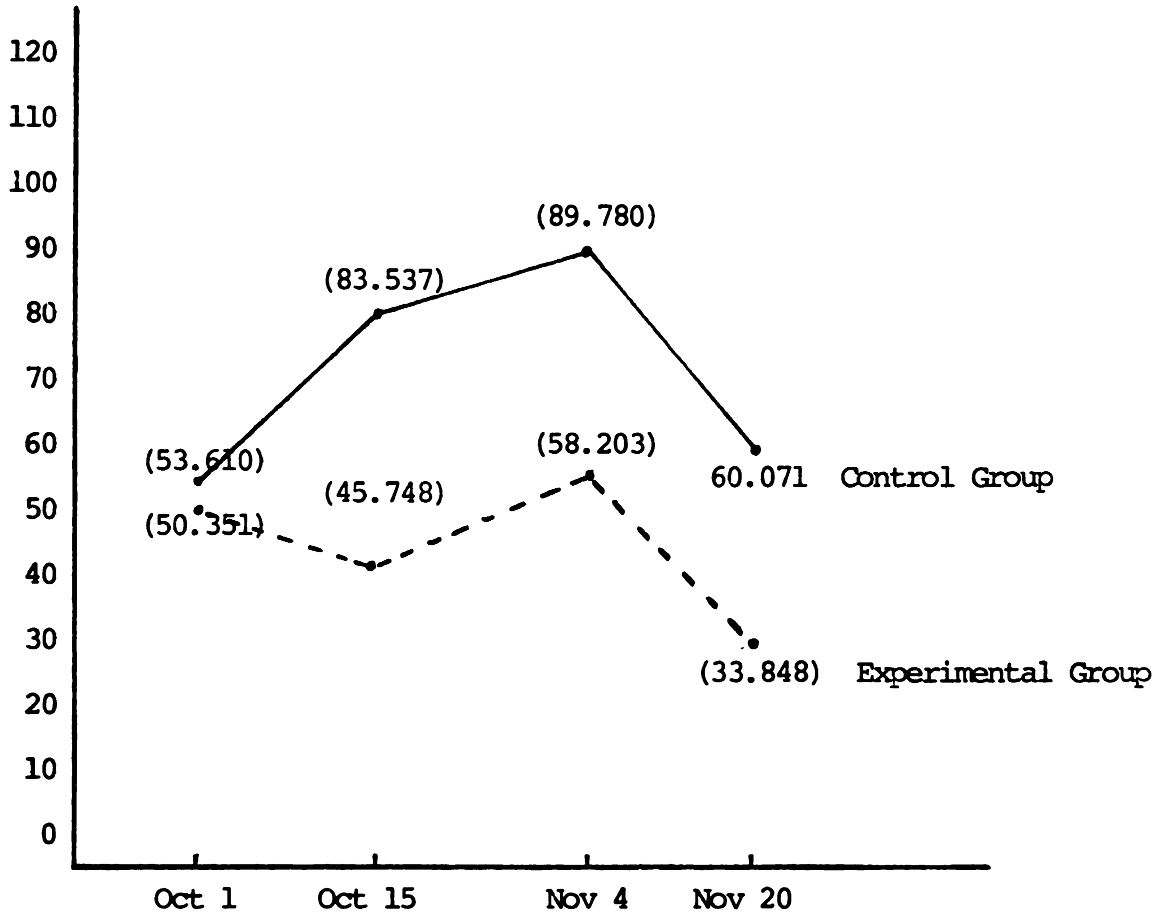

Figure 4. Plot of the Teaching Assistant-Leadership Paired Comparison for Experimental and Control Groups Across Four Points in Time

Key: ——— Control Group

- - - - Experimental Group

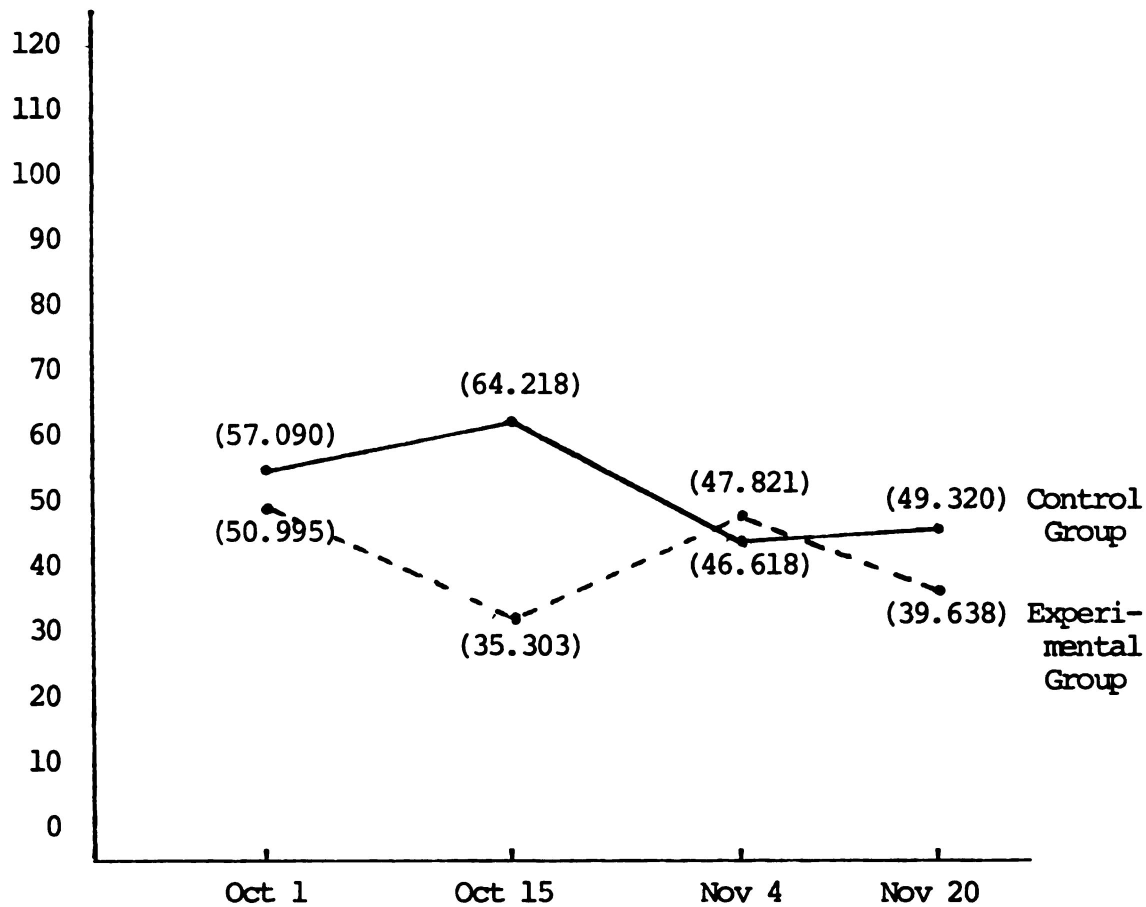

Figure 5. Plot of the Teaching Assistant-Instructor Paired Comparison for Experimental and Control Groups Across Four Points in Time

Key: ——— Control Group

----- Experimental Group

that had begun to develop several weeks later are quite expected. But the tendency for the trained teaching assistants to become more strongly associated with the specific attribute concepts was only slight at best. Instead, the difference between the experimental and control groups resulted from increases in the teaching assistant-attribute pairs for the control group. In other words, it may have been that rather than helping the trained assistants to be perceived more positively, the training program may have merely prevented them from becoming perceived more negatively. Students may have, at the early stages of the term, overestimated their teaching assistants. The training program could have enabled the trained assistants to quickly attain that level of performance while the lack of such training showed up in the reassessments made by the students in the second data set. Increases in mean scores for the experimental group from time two to time three may be a reflection of the training programs completion. That is, the teaching assistants in the training program may have reached a peak level of performance while they were learning to improve their classroom skills. A slight decrease in their performance may then occur until sufficient time has allowed them to improve by actually putting their skills to use. Along these same lines, the consistent decrease in the mean scores at the final point in time may merely be a function of experience. That is, all of the teaching assistants, whether they were trained or not, presumably improved their classroom skills as the term progressed. This would, of course, be expected to show more as they gained added experience. Thus, decreases in scores from the third to the final data sets for both groups may be accounted for by direct practical experience.

One additional finding should be mentioned before proceeding with more specific comparisons between the instructional performance pairs for the control and experimental groups. There is an exceptionally close resemblance between the mean scores at each point in time for the teaching assistant-leadership and teaching assistant-confident pairs. With the exception of the third data set in the control group, where the difference was 26.81, the scores are otherwise very similar. The differences range from .35 to 3.82 in the experimental group to a high of only 9.51 in the control group. The two pairs also follow identical patterns over time. While this similarity may merely reveal a strong covariation, it may also be an indication of the importance of building the teaching assistants' confidence in their leadership skills. In other words, at least perceptions of the teaching assistants' leadership ability may be in a large part dependent upon the confidence that they display in the classroom.

In discussing more specific differences between the control and experimental groups in relation to the criterion scores, it should first be recalled that the criterion score for each pair was allowed to vary over time. The instructor-attribute mean score for each point in time was used as the criterion. This enabled the criterion scores to adapt to the changing demands of the course as modeled by the instructor. Assuming that improvement in instructional performance should be reflected by mean scores converging on their respective criterion scores, the experimental group would appear to be the most improved over the term. That is, three of its five pairs were closest to the criterion score in the final data set. One of the remaining scores was at that time further than it had previously been in the preceding three data sets. None of the time four scores for the control group were either closest to or further from the criterion scores. These data additionally show that the mean scores for the experimental group were generally closer to the criterion scores than were their counterparts in the control group. Mean scores for the experimental group were closest for three of the five pairs in the initial data set (Table 8). Over the next three data collections (Table 9-11), the experimental group was closer to the criterion in fourteen of the fifteen comparisons.

The indication in these findings that the trained teaching assistants were perceived to be more closely identified with appropriate levels of teaching performance is further supported by several significant findings. However, these results must be interpreted with extreme caution since the Behrens-Fisher statistic does require independent random samples and has been previously mentioned, the data examined here is to some extent correlated. The discussion of these findings should thus be treated as a demonstration of what might result with a more appropriate data set. Of the six significant differences between the instructional performance pairs for the control and experimental groups, four occurred in the second time period (Table 9). In other words, there was a perceived difference between the trained and untrained assistants during the time in which the training program was being held. At this time, all but the teaching assistant-democratic pair yielded significant differences and even that pair approached significance. In all cases, the mean scores for the experimental group were closest to the criterion. The remaining two significant differences between the experimental and control groups occurred in the final data set (Table 11). It is especially informative to note that the teaching assistant-leadership pair (Figure 4) was the only pair that was significant in both the second and fourth data sets. At both times, the mean score was within a single point of the criterion score. It should be further pointed out that while there was no significant difference between the experimental and control group for this pair at time three (Table 10), there is still a sizable difference (31.58) between the mean scores. The teaching assistant-leadership pair is of particular importance because it reflects most directly on the training program which, of course, was specifically intended to improve the teaching assistants' leadership skills.

In summary, there were fifteen possible comparisons between the mean scores for the control and experimental group after the initial data set. Six were found to be significant and four others had substantial differences of over 30.00 which approached significance. In all but one case, significant or not, the mean for the experimental group was closest to the criterion. While these results are encouraging, the lack of additional significant findings suggests that caution against over interpretation should be exercised. However, two conclusions do seem to be warranted. First, the trained teaching assistants were perceived to be more closely identified with appropriate levels of teaching performance during their participation in the training program. Secondly, the trained teaching assistants were more strongly associated with the concept of leadership during and after their participation in the training program.

Adjunct Analyses

In the primary analyses, individual student responses were aggregated for all sections of the experimental and control groups. By using all available volunteers at each time interval, the samples were not random. Instead, the data, although treated as random, was correlated due to some subjects participating more than one time. In the analyses to follow, the data from students in each section was aggregated and then each section in turn individually contributed to the experimental or control group mean scores. While this procedure does not produce random samples, it should make the mean scores more stable. Only those sections who had students participating in all four data collections were used in these analyses (see Table 1). This procedure however, considerably reduced the sample sizes (experimental group, N=6 sections; control group, N=9 sections) and as such made any statistical comparison of means a relatively sterile endeavor. A discussion of general over time trends would, in any case, be informative.

Course Context

The effectiveness of the instructor, while expected to influence the rate and extent of student learning, would not necessarily be expected to also alter the overall learning process. In other words, the process by which students internalize course content would generally be expected to remain relatively constant. However, there appears to be no predominant trend that is entirely consistent for all six content pairs across the control and experimental groups (Tables 12 and 13). This is particularly evident in the control group data where changes over time seem to follow no trend whatsoever. Each of the pairs had their own unique pattern over time. On the other hand, each of the pairs in

Table 12. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Content Concepts for the Control Group Across Four Points in Time

Concept Pair	Condition			
	T_1	T_2	T_3	T_4
Communication and Social Science	46.461	50.029	65.061	53.479
	(9.126)	(21.647)	(33.486)	(36.379)
	9	9	9	9
Communication and Leadership	31.307	32.059	43.354	28.130
	(17.890)	(11.219)	(12.853)	(18.461)
	9	9	9	9
Communication and Democratic	53.803	46.467	62.314	75.424
	(18.721)	(14.421)	(29.550)	(62.000)
	9	9	9	9
Democratic and Leadership	60.112	43.674	52.619	39.991
	(20.087)	(10.348)	(21.920)	(23.561)
	9	9	9	9
Communication and Me	39.552	38.228	38.747	29.563
	(24.828	(12.970)	(18.980)	(14.509)
	9	9	9	9
Communication and Good	40.384	62.868	35.066	47.003
	(17.461)	(40.496)	(20.993)	(36.094)
	9	9	9	9

Table 13. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Content Concepts for the Experimental Group Across Four Points in Time

Concept Pair	Condition T_1	T_2	T_3	T_4
Communication and Social Science	58.710 (33.333) 6	45.957 (20.496) 6	47.875 (22.209) 6	45.347 (9.924) 6
Communication and Leadership	44.078 (49.542) 6	35.708 (13.046) 6	33.247 (15.335) 6	20.808 (8.681) 6
Communication and Democratic	69.620 (40.810) 6	57.905 (15.051) 6	54.028 (20.995) 6	46.775 (14.098) 6
Democratic and Leadership	54.205 (20.510) 6	41.235 (5.896) 6	57.805 (15.127) 6	32.767 (8.619) 6
Communication and Me	43.367 (32.370) 6	31.250 (13.670) 6	25.442 (9.789) 6	21.540 (5.791) 6
Communication and Good	54.430 (43.518) 6	41.923 (20.513) 6	47.135 (22.672) 6	33.973 (11.632) 6

the experimental group, with the exception of the democratic-leadership pair, tend to follow the same general tendencies. While the changes over time are not substantial, they are nevertheless quite consistent. For these pairs, there is a uniform decrease in the perceived discrepancy between the paired concepts from the first to the second time interval. There appears to be only minimal change from time two to time three. A decrease in mean scores then occurred from the third to the final points in time. Generally, this appears to suggest that for the experimental group, the concepts in the content pairs became more strongly associated with one another as the term progressed.

This convergent tendency for the content pairs in the experimental group is even more apparent when comparing changes in the two groups. For five of the six pairs in the time one data, the perceived differentiation between the concepts was greater for the experimental group than for the control group. In contrast, on the final data collection, all six of the paired concepts were perceived to be closer for the experimental group than for the control group.

Before proceeding to an examination of the instructional performance data, it is of interest to note the similarity in over time patterns between the experimental and control groups for the communication-me and leadership-democratic pairs. While the remaining four pairs had quite different trends in the two groups, the parallel findings for these two pairs may be an indication of the importance of the particular course material being taught in the learning process. This may in fact explain why no overall tendency was found for the entire data set.

Instruction

Instructional performance pairs for the control and experimental groups (Tables 14 and 15) both seem to follow a quite consistent but different general pattern over time. Very little noteworthy change occurred in the control group over the four time intervals. In other words, student perceptions of instructional performance for the untrained teaching assistants remained relatively stable throughout the term. Conversely, the experimental group did experience some change. In particular, substantial decreases in the values for the teaching assistant-instructional characteristic pairs occurred from the first to the second time periods. A slight increase then took place from time two to time three which was followed by a relatively consistent decrease moving from the third to the final data sets. Thus, the trained teaching assistants appear to have generally become more strongly associated with the relevant instructional performance concepts.

In contrasting the two groups, the yet to be trained teaching assistants in the experimental group were furthest from the instructional performance concepts for four of the five pairs at the initial point in time. At the second and

Table 14. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Instructional Concepts for the Control Group Across Four Points in Time

Concept Pair	Condition			
	T_1	T_2	T_3	T_4
Teaching Assistant and Expert	74.239	67.204	64.873	73.098
	(26.460)	(19.016)	(51.584)	(33.274)
	9	9	9	9
Teaching Assistant and Leadership	52.543	47.464	50.156	47.080
	(18.576)	(22.168)	(21.555)	(26.748)
	9	9	9	9
Teaching Assistant and Democratic	61.366	57.958	54.316	60.161
	(18.352)	(17.839)	(32.206)	(24.937)
	9	9	9	9
Teaching Assistant and Confident	47.697	47.170	42.079	44.470
	(13.683)	(14.562)	(12.945)	(17.197)
	9	9	9	9
Teaching Assistant and Instructor	47.840	44.271	47.556	40.339
	(38.157)	(15.572)	(15.913)	(22.952)
	9	9	9	9

Table 15. Mean Distances, Standard Deviations, and Sample Sizes Among Pairs of Instructional Concepts for the Experimental Group Across Four Points in Time

Concept Pair	Condition			
	T_1	T_2	T_3	T_4
Teaching Assistant and Expert	134.377 (131.103) 6	58.625 (23.955) 6	62.630 (22.305) 6	58.723 (30.467) 6
Teaching Assistant and Leadership	43.982 (21.030 6	45.537 (14.788) 6	54.780 (21.740) 6	34.412 (12.702) 6
Teaching Assistant and Democratic	84.115 (87.067) 6	54.190 (19.621) 6	63.420 (33.424) 6	49.295 (15.672) 6
Teaching Assistant and Confident	95.057 (128.995 6	47.122 (23.346) 6	50.115 (26.123) 6	31.555 (14.342) 6
Teaching Assistant and Instructor	48.918 (28.129) 6	33.620 (13.018) 6	48.490 (15.587) 6	36.830 (12.637) 6

third time intervals, there appears to be little perceived difference between the untrained and the trained teaching assistants. In the last data set, the trained assistants had become more closely identified with all five of the instructional assessment concepts. These findings seem to suggest that there was at least a perception of improvement in the instructional performance of the trained teaching assistants.

Summary of Results

Preliminary analyses on the precision with which student respondents were able to use the measurement system revealed relatively high coefficients of variability, particularly for unmanipulated pairs that included educational concepts of a general nature (i.e., learning, thinking, and studying). Because the sample consisted of freshmen entering a new social and educational system, it was suggested that these variability coefficients reflected not only measurement error but also student exposure to a new environment. This noise in data may also at least partially account for the absence of a significant manipulation effect. That is, concepts associated with the experimental manipulations would be expected to experience more variability than concepts that were not attempted to be altered. Such a difference was, however, not found in this study.

In the major analyses, no significant differences were found in the extent of student learning between the experimental and control group classes. However, while the training program was in progress, the teaching assistants who participated in it were perceived to be more closely associated with effective levels of classroom performance than teaching assistants who did not participate in the program. Moreover, the trained teaching assistants continued to be more closely associated with effective leadership through the remainder of the term. It appears, then, that while the trained teaching assistants may have been perceived to be more effective in their classroom performance, their effects did not result in more student learning.

An adjunct analysis of the data over time revealed no specific overall pattern of change. There was, however, an indication that the pattern of learning seems to be dependent on the course material being taught. The over time trends for the instructional performance pairs appear to support the major analyses. That is, the trained teaching assistants were perceived to have improved on their classroom performance while no change at all was perceived in the instructional performance of the untrained assistants.

CHAPTER V

DISCUSSION, RECOMMENDATIONS, AND SUMMARY

Discussion

The results of this study did not reveal any significant differences in the extent of student learning that occurred in classes aided by trained teaching assistants as compared to classes that were aided by teaching assistants who did not participate in the training program. While the lack of more student learning by students taught by the trained teaching assistants may suggest that the training program was unsuccessful in substantially improving teaching effectiveness, there are several alternative interpretations for these findings. The responsibility for student learning in this course was placed on the instructors and not their assistants. The influence of the instructors undoubtedly contaminated differences in effectiveness between the trained and untrained teaching assistants. It may have also been the case that the students sampled in this study did not learn a great deal from the small group exercises in which they participated. The more important determinant of student learning may have instead been lectures or other types of more direct instructional techniques which are among the duties assumed by the instructors. The lack of significant findings may have also resulted from the inability of the measurement procedures to adequately assess student learning. In other words, differences in student learning may have actually occurred but were not able to be detected. Without additional information, it would be difficult to weight these possibilities. It is, however, important to acknowledge their presence and, in light of them, to discuss the implications of these results for the learning process, the measurement of learning, and the teacher training program.

The most predominant feature demonstrated throughout these data would appear to be the lack of any uniform pattern of student learning. Although the adjunct analyses revealed a slight tendency, in the experimental group, for the concepts in the content pairs to converge, the overtime sequence of mean scores for the content pairs was generally non-monotonic. Changes were not always in the same direction and, for this student sample, did not result in a reduction in the difference between the mean scores and their respective criterion scores. Moreover, the mean scores did not oscillate about the criterion scores. While there were similarities between some specific pairs, there was enough deviation to prevent any conclusion to be made regarding student learning.

An examination of the trace from the various spatial coordinates matrices (Tables 16-23 in Appendix B) may be useful in providing additional insight into the general nature of the learning process. It should first be recalled, from Chapter II, that the spatial coordinates matrix is constructed to be a square matrix consisting of orthogonal dimensions as columns and the projections of the concepts on each of the dimensions as rows. Variances of these projections are represented by eigenvalues which when summed across the matrix represents the total variance in the matrix referred to as the trace. In terms of learning, it would appear that effective teaching of specific course content to a class of students would yield a relatively uniform way of organizing that material in the cognitive structure. Consequently, variability should be low. It should also be pointed out, however, that if the goal of the course was to promote individualistic points of view, success in attaining this goal would be reflected by a relatively large trace. The data used here to obtain the spatial coordinates came from the primary analyses where individual student responses were aggregated across all sections for the experimental and control groups. The interpretation of these findings are, once again, restricted by the inappropriateness of this correlated data set. Consequently, the analysis primarily serves an illustrative function. In this particular study, the representation of the trace across four points in time for both the control and experimental groups (Figure 6) clearly illustrates the differences in the total variance

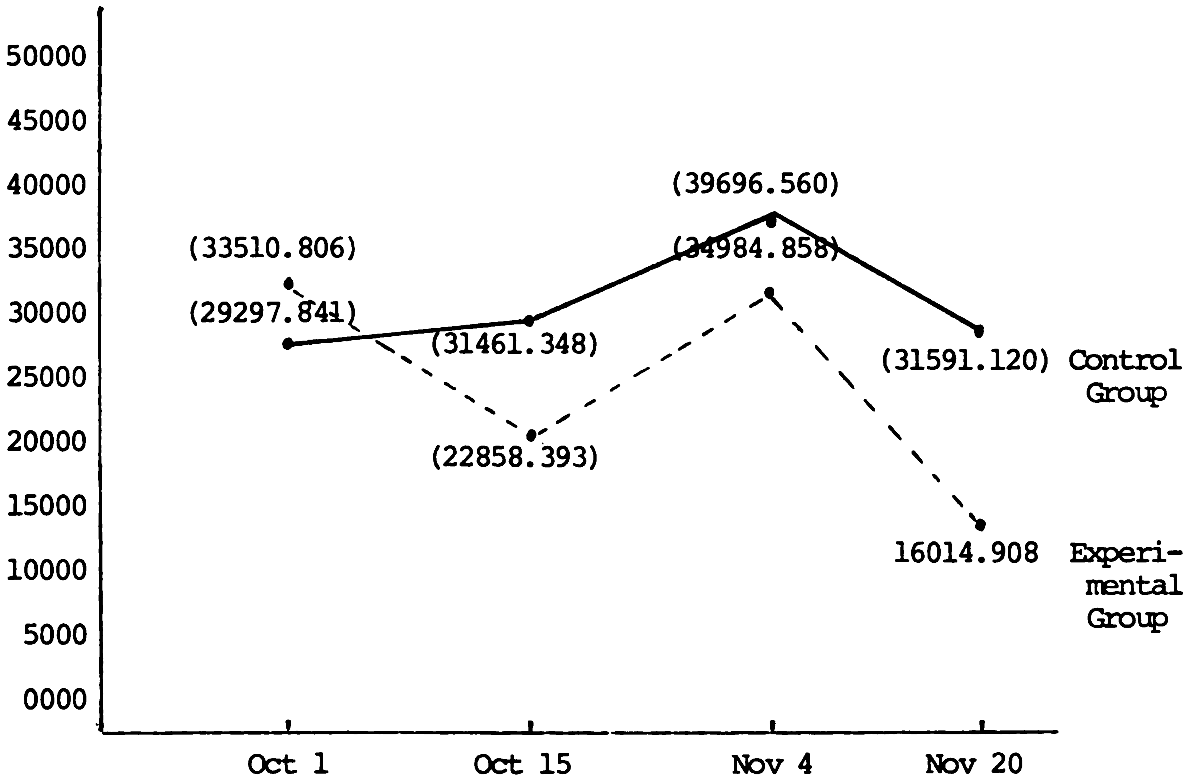

Figure 6. Plot of the Trace (Total Variance) from the Co-ordinates Matrix for Experimental and Control Groups Across Four Points in Time

Key: ——— Control Group

----- Experimental Group

for each of these matrices. It appears that after the initial data collection, the experimental group maintained less variability. This is particularly evident at the end of the term where the total variance for the experimental group (16,014.91) was approximately half of that for the control group (31,591.12). An examination of the trace would appear, then, to be informative and should be further pursued in future research.

In the previous chapter, it was suggested that the all encompassing nature of some paired comparisons may have resulted in changes being delayed until students had absorbed enough course material to draw some general conclusions. Information of this kind was not specifically taught to students as course content but should eventually have been affected by it. On the other hand, pairs including concepts (leadership and democratic) that were directly covered by course materials showed more immediate changes. This would seem to suggest that different methods of teaching (i.e. direct versus indirect) may have resulted in entirely different patterns of learning.

In addition to bringing up the potential affects that various teaching methods may have had on student learning, it may also be informative to consider factors which influence how concepts change in the cognitive structure. Researchers have, while using procedures similar to those used in this study, found that the accumulation of concept relevant information is inversely related to the change in that concept over time (Danes, 1976; Saltiel and Woelfel, 1975; Woelfel and Saltiel, 1974). Put in the current context, the more information that students had regarding the content of this particular course, the less change would be expected to result from exposure to relevant course material. Student learning may have been dependent upon the amount of previous information that the students brought with them to the classroom. The particular method used in presenting concept relevant course material and the susceptibility or resistance of those concepts to change in the cognitive structure would appear, then, to be important considerations in assessing learning.

The results of this study have several additional implications which focus more directly on the training program. Five of the seven significant differences between the control and experimental groups for the instructional performance pairs occurred during the time in which the teaching assistants were actually participating in the training program. At that time, they were perceived to be more closely associated with effective levels of instruction. However, these differences did not hold up in succeeding time intervals. This would seem to indicate that while it was important to provide the teaching assistants with early training that they could put to immediate use, the training may have been prematurely terminated. Additional, less lengthy, sessions held periodically through the remainder of the term may have been helpful in improving or at least maintaining the levels of teaching performance attained during the early part of the training program.

The need for establishing some sort of standard for assessing teaching performance is clearly demonstrated by the teaching assistant-democratic pair in the final data set (Table 11). For this pair, the criterion score is 59.14 with the mean score for the control and experimental groups being 77.76 and 43.66 respectively. The criterion score is near the center of the difference between the control and experimental groups. The teaching assistants from the two groups are then almost equally distant from what is considered to be the optimal level of teaching effectiveness for this particular attribute. However, improvement for the untrained teaching assistants would require them to become more democratic. The trained teaching assistants, on the other hand, appear to have appropriated too much of this characteristic and as a result have begun to become counter productive. Improvement for them would call for a less democratic style and possibly more authoritative leadership. Without such a standard, as is typically the case in current evaluation procedures, the changes that are necessary for improvement are not clearly defined. In the current example, the position of the trained teaching assistants could easily be misinterpreted as being superior to that of the untrained teaching assistants. Improvement may then be associated with becoming more democratic which for the experimental group would lead to less effective instruction. Thus, the establishment of a standard representing optimal levels of teaching performance for relevant attributes would appear to be an essential ingredient in the evaluation of instruction.

The results from the major analyses seem to demonstrate the most important deficiency in current evaluation practices. That is, current procedures primarily focus on the assessment of instructors in regards to certain specific teaching characteristics (Genova et al., 1976) which is essentially the same as the determinations being made with the teaching performance items in the current study. If this was all that was being used to determine teaching effectiveness, it might generally be concluded, with some limitations, that the training program was successful in producing teaching assistants that were at least for a short time perceived to be more effective in the classroom. By additionally examining student learning, however, it can be seen that these perceived differences in effectiveness between the trained and untrained teaching assistants are not reflected in corresponding variations in the extent to which students learned course material. It would appear, then, that evaluations based strictly on teaching performance may be very misleading. That is not to say that examining essential teaching characteristics is unimportant and should be eliminated from evaluation procedures. To the contrary, such assessments are informative, particularly for diagnostic purposes. However, the examination of teaching performance in evaluation should be used in conjunction with measures of learning.

Future Research

While possibilities for future research were implicitedly made in the previous discussion, this section will provide some more specific suggestions. The most immediate concern for researchers should be to thoroughly examine the reliability and validity of the evaluation system. Reliability can be assessed by simply including several paired comparisons more than once in the instrument and computing the correlations between the responses to the same pairs. Concepts not manipulated and whose meaning would not be expected to change could also be included in the instrument in order to examine the stability of the system over time. Additionally, control groups may be employed to examine manipulation effects. An indication of the participants ability to adequately use the measurement procedures can be determined by inserting the criterion pair in the instrument. Substantial deviations from the score provided for the subjects use would seem to be grounds for assuming that subjects weren't able or willing to properly fill out the questionnaire. This would appear to be justification for eliminating such cases from the subject pool. Validity can be addressed by partialling students according to the grade they received in the course. A strong positive relationship would be expected between the grade received by the students and their success in learning the appropriate meanings for the domain of instruction. Moreover, a negative relationship should exist between student grades and the perceived difference between the actual and ideal instructors.

Beyond this preliminary work, the research possibilities for this system are enormous. The teaching-learning relationship is the foundation of formal education. A precise, reliable and valid measure of teaching effectiveness which is based on this relationship can be very useful to educational researchers. One specific application that should be mentioned is the use of this system to improve the quality of instruction. The use of specific instructional performance characteristics in conjunction with some ideal standard has previously been shown to be able to: (1) locate optimal levels at which such characteristics would be most effective, (2) determine the deviation of instructors from those optimal levels, and (3) suggest what might be done to help the instructor become more effective. In addition, various teaching styles (made up from combinations of these characteristics) can be identified. By additionally video taping and analyzing a sample of instructors, specific behavioral indicators for each characteristic and style can be found. Used with measures of learning, this information can lead to an assessment of the relationship of various teaching characteristics and styles with student learning. Training programs could then be devised in accordance with this information.

This system would appear to be most useful in the multi-sectioned course such as the one examined in this study. Averaging the data across all sections would enable the directors of such courses to obtain general information regarding the specific strengths and weaknesses of their staffs. Training sessions could then be established in light of this information. The addition of each instructor as a row and column to the original means distance matrix allows feedback to be obtained for each instructor in regards to their relative position to the ideal standard as well as in relation to their colleagues. If data were collected early enough, it would be possible to improve teaching effectiveness in time to have a positive influence on student learning in the same term.

Two specific studies that are already in progress should be briefly mentioned. In the first study, the most essential concepts from an undergraduate research methods course were incorporated into a measuring instrument similar to the one used in the present study. Each student in the class randomly selected, from the student directory, one student that was not in the course and willing to participate in the study. These selected students served as the control group. With the exception of weekends, four students from both the class and the control group filled out a questionnaire on a daily basis. Specific days for filling out the questionnaire were randomly chosen. Every student participated twice during the term. Taking into account the specific dates in which concept relevant material was covered in class, an examination of these data across time should produce some very meaningful findings in regards to the learning process.

The second study is an actual comparison between the evaluation procedures proposed in this report and more traditional methods of assessing teaching effectiveness. The same teaching performance characteristics are represented in both instruments. In addition, different levels of each of these characteristics were incorporated into what has been pretested to be effective and ineffective video taped lectures on leadership. The results will be examined in order to determine the extent to which these instruments are able to detect these manipulated differences. This study also examines the susceptibility of the two instruments to the halo effect. For this purpose, subjects were either told that the lecturer was extremely cooperative in projects that would eventually benefit students or that the lecturer was very uncooperative and was not particularly interested in determining means for improving the quality of classroom instruction. A neutral condition in which nothing was said regarding the lecturer was also included. The presence of the halo effect would be reflected in dissimilarities of responses across these three conditions. These two studies represent attempts to shed additional light on the procedures presented in this report such that their usefulness can be further maximized.

Summary of Report

This report begins by pointing out the controversial history associated with the evaluation of instruction. Some of the major theoretical and methodological problems with current practices in evaluating instruction are reviewed. It is argued that such evaluation should be based on a theory of the teaching-learning relationship. It is further suggested that a cognitive approach to student learning would be the most fruitful for the purpose of evaluating instruction. Such a perspective based on the definition of learning as the evolution of meaning is then provided. On the basis of this perspective, a multidimensional scaling technique for precisely evaluating instruction is presented.

The current study evaluates a training program for teaching assistants by examining the teaching effectiveness of a group of trained assistants in contrast to a group of assistants that were not trained. Student learning and perceptions of actual classroom performance were used as the criteria for the evaluation. Data was gathered from student volunteers at four different times during the term. The participation of some subjects in more than one data collection was in violation of the statistical assumption of independent random samples. The analyses were performed, in any case, for illustrative purposes.

No significant differences were found in the extent of student learning between the classes aided by trained teaching assistants as opposed to those classes aided by untrained teaching assistants. However, while the training program was in progress, the teaching assistants who participated in it were perceived to be more closely associated with effective levels of classroom performance than teaching assistants who did not participate in the program. Moreover, the trained teaching assistants continued to be more closely associated with effective leadership through the remainder of the term. These findings suggest that while the trained teaching assistants may have been perceived to be more effective in their classroom performances, their efforts did not result in more student learning.

APPENDIX A

Questionnaire

MICHIGAN STATE UNIVERSITY

Department of Communication

Fall, 1975

Dear Participant:

This term we are engaged in a project involving Communication 100. All twenty sections will be assisting us in this endeavor. Periodically, we will be asking each of you for your cooperation in giving us the information necessary to make the project a success. It is our hope that with your help, this information will lead to improvements in the COM 100 course.

We appreciate your response to our initial request two weeks ago. We are grateful to those of you who participated and ask for your continued support of this project. We wish that those of you not participating initially will join our project at this time. Your cooperation will greatly contribute to the project's success.

Should you have any questions regarding the project, please feel free to call one of us or stop in at 535 South Kedzie.

Thank you,

Kim Kanaga	353-3237 or 487-1641
Ilene Benison	355-5557
Marianne Mnich	353-2824
David Palmer	353-0577
Donna Paquette	353-0274
Jean Riker	355-0436

Instructor = ____________
UTA #1 = ____________
UTA #2 = ____________

1) ID#________________ 2) Name__________________________

3) Local Address____

4) Date________________ 5) Telephone___________ 6) Age___

7) Sex________________ 8) Major_______________ 9) GPA___

10) Class:Fr__________ Soph________ Jr________ Sr_______

11) COM 100 Section #________________

12) COURSE STATUS (Circle one and answer the adjoining questions)

a) Student

a1) In the last two weeks, how many different times was <u>your</u> classroom activities (i.e. exercises, group discussions, lectures) lead by:
UTA #1_____________________ times
UTA #2_____________________ times
Instructor_________________ times

a2) How many minutes during the past two weeks were <u>your</u> classroom activities (i.e. exercises, group discussions, lectures) lead by:
UTA #1_____________________ minutes
UTA #2_____________________ minutes
Instructor_________________ minutes

a3) In the past two weeks, what percentage of class time was spent on:
Lectures___________________ %
Discussions________________ %
Exercises__________________ %
Other______________________ %

b) Undergraduate Teaching Assistants (UTA)

b1) During the past two weeks, how many different times were you responsible for leading some classroom activity (i.e. exercises, group discussions, lectures)?
___________________________ times

b2) How many minutes during the past two weeks were you responsible for leading some classroom activity (i.e. exercises, group discussions, lectures)?
___________________________ minutes

b3) When your class was divided into groups, what was the average size group (during the past two weeks) that you were responsible for leading?

___________________________ people

b4) Prior to this term, have you ever been a teaching assistant or had any other similar experience?

Yes_______________ No_____________

b5) Are you participating in the UTA workshop offered by the Communication Department?

Yes_______________ No_____________

If yes, how many sessions have you attended?

____

c) Instructor, Jr._______ Sr.____________

c1) During the past two weeks, how many different times were you responsible for leading some classroom activity (i.e. exercises, group discussions, lectures)?

___________________________ times

c2) How many minutes during the past two weeks were you responsible for leading some classroom activity (i.e. exercises, group discussions, lectures)?

___________________________ minutes

c3) When your class was divided into groups, what was the average size group (during the past two weeks) that you were responsible for leading?

___________________________ people

c4) Have you had any teaching experience prior to this term?

Yes________________ No_____________

c5) In the past two weeks, what percentage of class time was spent on:

Lectures___________________ %
Discussions________________ %
Exercises__________________ %
Other______________________ %

c6) Class size:________________ students
________________ UTA's
________________ instructors

c7) Dates on which small group communication was taught (to be filled out <u>only</u> after this area has been completed)

____

c8) Dates on which leadership was taught (to be filled out <u>only</u> after this area has been completed)

____

Just as we can measure the distance between two physical objects (in terms of inches, yards, miles, etc.), we can also measure the distance between concepts or ideas. This questionnaire asks you to make judgments about how different (or in other words "far apart") certain concepts are from each other. Differences between concepts are measured in units, such that the more different two concepts are, the more units apart they are from each other. Two concepts that are identical in meaning, then, would be zero (0) units apart.

To help you know how big a unit is, Red and White are 100 units apart; that is, imagine that the difference (distance) between the colors Red and White is 100 units. We would like you to use this idea of distance in the comparison of the concepts on the next few pages.

You are supposed to tell us how many units apart the concepts on the next few pages are from each other. Remember, the more different the two concepts are from each other, the larger the number of units apart they are. If you think that any of the pairs of concepts are more different than red and white, write a number larger than 100. If you think they are not as different, use a smaller number. Remember, the more different the concepts are from each other, the higher the number you should write.

There are no correct or incorrect responses, only your perceptions of the differences between concepts. Consider each pair carefully and indicate the number of units that you feel separate the concepts.

Your cooperation is most appreciated.

Thank you for your help.

REMEMBER: RED AND WHITE ARE 100 UNITS APART. THE MORE DIFFERENT THE CONCEPTS ARE FROM EACH OTHER, THE HIGHER THE NUMBER YOU SHOULD WRITE. WRITE ANY NUMBER YOU WANT.

Instructor = ________
UTA #1 = ________
UTA #2 = ________

How far apart are	Units
Expert and Instructor	______
Expert and Studying	______
Expert and Thinking	______
Expert and Me	______
Expert and UTA #2	______
Expert and Physical Science	______
Expert and Leadership	______
Expert and Learning	______

How far apart are	Units
Expert and Democratic	______
Expert and Humanities	______
Expert and Good	______
Expert and Confident	______
Expert and Social Science	______
Expert and UTA #1	______
Expert and Communication	______
Instructor and Studying	______

How far apart are	Units
Instructor and Thinking	______
Instructor and Me	______
Instructor and UTA #2	______
Instructor and Physical Science	______
Instructor and Leadership	______
Instructor and Learning	______
Instructor and Democratic	______
Instructor and Humanities	______

REMEMBER: RED AND WHITE ARE 100 UNITS APART. THE MORE DIFFERENT THE CONCEPTS ARE FROM EACH OTHER, THE HIGHER THE NUMBER YOU SHOULD WRITE. WRITE ANY NUMBER YOU WANT.

Instructor = ________
UTA #1 = ________
UTA #2 = ________

How far apart are	Units
Instructor and Good	______
Instructor and Confident	______
Instructor and Social Science	______
Instructor and UTA #1	______
Instructor and Communication	______
Studying and Thinking	______
Studying and Me	______
Studying and UTA #2	______

How far apart are	Units
Studying and Physical Science	______
Studying and Leadership	______
Studying and Learning	______
Studying and Democratic	______
Studying and Good	______
Studying and Confident	______
Studying and Social Science	______

How far apart are	Units
Studying and UTA #1	______
Studying and Communication	______
Thinking and Me	______
Thinking and UTA #2	______
Thinking and Physical Science	______
Thinking and Leadership	______
Thinking and Learning	______
Thinking and Democratic	______

REMEMBER: RED AND WHITE ARE 100 UNITS APART. THE MORE DIFFERENT THE CONCEPTS ARE FROM EACH OTHER, THE HIGHER THE NUMBER YOU SHOULD WRITE. WRITE ANY NUMBER YOU WANT.

Instructor = ________
UTA #1 = ________
UTA #2 = ________

How far apart are	Units
Thinking and Humanities	______
Thinking and Good	______
Thinking and Confident	______
Thinking and Social Science	______
Thinking and UTA #1	______
Thinking and Communication	______
Me and UTA #2	______
Me and Physical Science	______

How far apart are	Units
Me and Leadership	______
Me and Learning	______
Me and Democratic	______
Me and Humanities	______
Me and Good	______
Me and Confident	______
Me and Social Science	______
Me and UTA #1	______

How far apart are	Units
Me and Communication	______
UTA #2 and Physical Science	______
UTA #2 and Leadership	______
UTA #2 and Learning	______
UTA #2 and Democratic	______
UTA #2 and Humanities	______
UTA #2 and Good	______
UTA #2 and Confident	______

REMEMBER: RED AND WHITE ARE 100 UNITS APART. THE MORE DIFFERENT THE CONCEPTS ARE FROM EACH OTHER, THE HIGHER THE NUMBER YOU SHOULD WRITE. WRITE ANY NUMBER YOU WANT.

Instructor = ______
UTA #1 = ______
UTA #2 = ______

How far apart are	Units
UTA #2 and Social Science	______
UTA #2 and UTA #1	______
UTA #2 and Communication	______
Physical Science and Leadership	______
Physical Science and Learning	______
Physical Science and Democratic	______
Physical Science and Humanities	______
Physical Science and Good	______

How far apart are	Units
Physical Science and Confident	______
Physical Science and Social Science	______
Physical Science and UTA #1	______
Physical Science and Communication	______
Leadership and Learning	______
Leadership and Democratic	______
Leadership and Humanities	______
Leadership and Good	______

How far apart are	Units
Leadership and Confident	______
Leadership and Social Science	______
Leadership and UTA #1	______
Leadership and Communication	______
Learning and Democratic	______
Learning and Humanities	______
Learning and Good	______
Learning and Confident	______

REMEMBER: RED AND WHITE ARE 100 UNITS APART. THE MORE DIFFERENT THE CONCEPTS ARE FROM EACH OTHER, THE HIGHER THE NUMBER YOU SHOULD WRITE. WRITE ANY NUMBER YOU WANT.

Instructor = __________
UTA #1 = __________
UTA #2 = __________

How far apart are	Units
Learning and Social Science	
Learning and UTA #1	
Learning and Communication	
Democratic and Humanities	
Democratic and Good	
Democratic and Confident	
Democratic and Social Science	
Democratic and UTA #1	

How far apart are	Units
Democratic and Communication	
Humanities and Good	
Humanities and Confident	
Humanities and Social Science	
Humanities and UTA #1	
Humanities and Communication	
Good and Confident	
Good and Social Science	

How far apart are	Units
Good and UTA #1	
Good and Communication	
Confident and Social Science	
Confident and UTA #1	
Confident and Communication	
Social Science and UTA #1	
Social Science and Communication	
UTA #1 and Communication	

APPENDIX B

Spatial Coordinates Matrices for the Experimental and Control Groups at Four Points in Time

Table 16. Spatial Coordinates of First Eight Dimensions for Control Group at Time One

Concepts	1	2	3	4	5	6	7	8
Expert	59.246	-28.861	-32.078	-2.938	13.944	2.756	-.552	1.441
Instructor	1.974	-3.037	-31.182	11.320	-5.557	-2.058	1.768	2.761
Studying	27.744	-8.427	28.586	4.397	-7.436	-2.066	-.399	12.898
Thinking	16.043	-4.600	16.863	-3.231	5.770	5.742	-.817	-8.447
Me	-29.142	-38.180	34.597	-14.171	-17.663	7.565	.413	-1.751
Teaching Assistant #2	-37.420	35.355	-.856	-17.469	13.705	-17.439	-4.715	.588
Physical Science	49.491	47.203	28.352	3.799	5.081	5.042	.082	-5.661
Leadership	-3.683	-3.616	-14.496	-11.132	-6.141	-2.003	-19.878	-9.571
Learning	8.422	1.074	9.018	-4.616	1.518	4.096	5.149	9.456
Democratic	-36.747	-6.138	-5.361	.913	25.314	30.661	-5.587	4.543
Humanities	-22.111	-12.724	9.389	18.545	22.159	-27.166	4.921	3.459
Good	-4.020	-6.449	-5.724	-17.193	5.904	1.041	26.412	-11.141
Confident	6.767	-.543	-15.160	-23.040	-20.236	-15.774	-1.039	4.475
Social Science	-10.024	-10.528	.902	40.196	-12.581	-4.611	-.490	-9.372
Teaching Assistant #1	-21.344	34.322	-19.975	10.278	-21.164	15.664	8.735	4.314
Communication	-5.196	5.149	-2.875	4.343	-2.647	-1.453	-9.005	2.008

Table 16 (cont'd.)

Eigenvalues (Roots) of Eigenvector Matrix								
	11808.692	7441.234	6183.362	3743.639	3067.811	2653.326	1393.201	751.113
Number of Iterations to Derive the Root								
	7	13	71	13	13	6	40	11
Percentage of Distance Accounted for by Individual Vector								
	31.351	19.756	16.416	9.939	8.145	7.044	3.699	1.994
Cumulative Percentages of Real Distance Accounted for								
	31.351	51.107	67.523	77.462	85.606	92.651	96.350	98.344
Cumulative Percentages of Total (Real and Imaginary) Distance Accounted for								
	40.306	65.704	86.809	99.587	110.058	119.115	123.870	126.434

Trace 29297.841

Number of Dimensions in Real Space 11

Table 17. Spatial Coordinates of First Eight Dimensions for Experimental Group at Time One

Concepts	1	2	3	4	5	6	7	8
Expert	40.198	-54.406	17.339	-28.542	1.019	-3.065	2.516	-.804
Instructor	-6.210	-15.753	.147	-1.193	.931	-9.648	-19.111	-8.072
Studying	19.871	8.875	-13.959	-7.083	25.403	27.279	-2.497	-3.355
Thinking	17.537	7.212	.920	-2.525	13.974	3.587	-.330	7.282
Me	-32.046	65.079	-3.176	-13.706	9.662	.003	-2.798	4.039
Teaching Assistant #2	-1.241	-24.112	-10.640	39.258	2.105	4.577	-9.456	23.895
Physical Science	79.429	28.670	-8.500	12.317	2.443	-14.698	6.825	-1.601
Leadership	-27.725	-13.214	5.844	-13.541	-5.344	-11.769	3.287	8.295
Learning	10.809	6.955	-13.855	-.895	1.093	-3.569	-2.466	-9.405
Democratic	-36.173	-5.437	27.224	11.823	36.640	-15.228	13.300	-4.664
Humanities	-.459	8.951	17.995	-1.602	-25.293	14.531	26.163	6.360
Good	-10.029	13.984	-19.482	.563	-31.133	-11.987	.151	-3.367
Confident	-24.024	-18.491	-31.296	-36.149	-.330	5.892	-.237	3.712
Social Science	1.999	14.740	47.237	2.197	-16.990	15.091	-18.607	-4.804
Teaching Assistant #1	-24.472	-24.984	-17.103	32.820	-5.584	12.732	8.210	-18.008
Communication	-7.462	1.931	.306	6.256	-8.595	-13.728	-4.951	.998

Table 17 (cont'd.)

Eigenvalues (Roots) of Eigenvector Matrix

13224.823	10692.953	5878.594	5508.049	4321.345	2447.452	1838.233	1316.467

Number of Iterations to Derive the Root

8	9	19	12	11	9	7	5

Percentage of Distance Accounted for by Individual Vector

28.577	23.106	12.703	11.902	9.338	5.289	3.972	2.845

Cumulative Percentages of Real Distance Accounted for

28.577	51.682	64.385	76.287	85.625	90.913	94.885	97.730

Cumulative Percentages of Total (Real and Imaginary) Distance Accounted for

39.464	71.373	88.916	105.352	118.248	125.551	131.037	134.965

Trace 33510.806

Number of Dimensions in Real Space 11

Table 18. Spatial Coordinates of First Eight Dimensions for Control Group at Time Two

Concepts	1	2	3	4	5	6	7	8
Expert	-32.788	14.160	39.529	6.169	-10.038	3.006	3.811	4.347
Instructor	15.607	3.693	17.586	9.416	8.956	6.861	-12.012	-9.207
Studying	-9.323	19.228	-10.711	-24.544	-10.380	-13.427	4.374	-2.670
Thinking	-38.448	-7.346	-2.563	-6.462	6.243	12.970	-7.950	-13.145
Me	-17.392	-40.249	-33.907	-7.969	-6.034	-9.412	-3.010	-.513
Teaching Assistant #2	83.778	27.298	-8.051	6.811	-7.778	-9.433	.893	1.725
Physical Science	-30.656	54.979	-12.295	-5.755	21.510	-2.688	-4.992	3.666
Leadership	-5.726	-16.091	9.838	14.956	-15.288	2.100	2.042	8.840
Learning	-11.006	4.095	-4.120	-8.635	4.435	-6.897	-3.129	7.020
Democratic	-10.251	-9.337	-15.854	42.652	13.924	.060	-4.429	5.031
Humanities	8.535	-2.228	-5.792	-1.966	9.934	20.745	30.336	-10.734
Good	-10.069	-13.082	6.279	.569	7.305	-14.792	13.993	16.461
Confident	-1.565	-15.286	21.670	-8.444	-7.211	-23.636	-2.228	-11.910
Social Science	6.047	2.860	-7.220	-14.181	-25.459	29.763	-8.379	12.165
Teaching Assistant #1	45.809	-29.349	13.764	-19.949	27.170	6.969	-7.879	3.702
Communication	7.449	6.656	-8.152	17.332	-17.289	-2.189	-1.440	-14.778

Table 18 (cont'd.)

Eigenvalues (Roots) of Eigenvector Matrix

13771.030	7712.193	4574.526	4006.296	3223.810	2788.399	1564.924	1361.120

Number of Iterations to Derive the Root

5	6	13	32	14	21	14	11

Percentage of Distance Accounted for by Individual Vector

34.736	19.453	11.539	10.105	8.132	7.033	3.947	3.433

Cumulative Percentages of Real Distance Accounted for

34.736	54.189	65.727	75.833	83.964	90.998	94.945	98.378

Cumulative Percentages of Total (Real and Imaginary) Distance Accounted for

43.771	68.284	82.825	95.559	105.806	114.668	119.643	123.969

Trace 31461.348

Number of Dimensions in Real Space 10

Table 19. Spatial Coordinates of First Eight Dimensions for Experimental Group at Time Two

Concepts	1	2	3	4	5	6	7	8
Expert	27.534	46.182	12.340	-14.743	5.615	2.538	-1.589	.470
Instructor	-8.666	10.414	9.437	-11.727	-9.435	.920	.004	4.610
Studying	22.092	-7.357	-22.922	-22.444	-8.822	-2.401	5.467	3.197
Thinking	4.723	-5.716	.123	-5.518	-6.550	8.593	-.716	6.482
Me	-31.119	-27.145	-26.717	7.179	-8.985	-4.878	-2.926	6.493
Teaching Assistant #2	-11.365	13.778	32.380	21.869	-.343	4.354	3.511	13.510
Physical Science	56.765	-12.235	-15.316	22.457	7.743	9.874	1.535	-2.710
Leadership	-19.252	12.117	-2.211	-2.254	15.334	.307	-18.501	-8.006
Learning	11.797	3.524	-12.321	7.161	-3.818	-11.651	-5.350	6.682
Democratic	-34.274	-17.327	.445	-4.697	2.883	27.872	1.975	-4.308
Humanities	-3.266	-35.336	18.491	-12.922	23.365	-16.303	3.606	4.313
Good	-10.747	9.034	-11.463	3.584	17.538	-1.285	20.886	-5.651
Confident	-9.968	30.115	-20.798	.849	2.664	1.292	-7.102	2.260
Social Science	25.601	-32.349	24.591	-5.403	-13.670	4.483	-4.723	-7.044
Teaching Assistant #1	-15.786	19.409	6.067	4.679	-20.919	-10.322	11.518	-10.942
Communication	-4.069	-7.110	7.874	11.931	-2.400	-13.395	-7.596	-9.355

Table 19 (cont'd.)

Eigenvalues (Roots) of Eigenvector Matrix

8494.626	7574.812	4529.942	2375.594	2129.826	1714.968	1143.331	745.536

Number of Iterations to Derive the Root

17	6	13	18	12	15	15	8

Percentage of Distance Accounted for by Individual Vector

28.938	25.805	15.432	8.093	7.256	5.842	3.895	2.540

Cumulative Percentages of Real Distance Accounted for

28.938	54.743	70.175	78.267	85.523	91.365	95.260	97.800

Cumulative Percentages of Total (Real and Imaginary) Distance Accounted for

37.162	70.300	90.117	100.510	109.827	117.330	122.332	125.593

Trace 22858.393

Number of Dimensions in Real Space 12

Table 20. Spatial Coordinates of First Eight Dimensions for Control Group at Time Three

Concepts	1	2	3	4	5	6	7	8
Expert	-36.794	6.629	35.768	-20.800	-11.897	18.399	-5.972	-10.283
Instructor	30.276	-13.568	22.316	7.805	-17.085	-22.614	10.369	12.011
Studying	8.694	19.269	-19.212	-16.151	-13.096	24.416	.476	5.075
Thinking	-25.802	19.352	7.277	-9.201	5.381	-4.527	2.892	.066
Me	-22.421	8.691	-34.444	6.346	27.943	-3.381	-18.409	3.274
Teaching Assistant #2	115.109	5.415	-8.048	-3.809	-.414	-5.233	-3.424	-2.630
Physical Science	6.889	78.697	14.924	13.247	1.124	-5.811	-10.555	1.443
Leadership	-12.700	-18.573	25.833	-7.379	32.316	-11.453	6.891	-2.108
Learning	-31.893	-10.116	-2.584	-24.067	-7.645	-2.430	-1.991	24.037
Democratic	-22.259	-34.726	-7.412	18.326	-15.980	-6.957	-20.837	10.333
Humanities	-17.726	4.780	-23.230	-9.921	-24.279	-20.308	8.233	-18.856
Good	4.117	-20.584	-23.537	-19.603	5.942	5.862	7.318	-10.253
Confident	15.964	-6.334	2.741	-12.131	18.261	-1.674	12.052	3.506
Social Science	-34.264	10.683	-12.745	39.271	-.651	10.909	29.209	3.386
Teaching Assistant #1	39.230	-26.836	12.557	18.088	.759	29.560	-3.440	.139
Communication	-16.417	-22.778	9.797	19.980	-.680	-4.757	-12.812	-19.140

Table 20 (cont'd.)

Eigenvalues (Roots) of Eigenvector Matrix

22054.558	10764.788	5919.534	4937.372	3733.993	3189.569	2395.182	1835.281

Number of Iterations to Derive the Root

5	10	20	8	14	10	11	6

Percentage of Distance Accounted for by Individual Vector

39.419	19.241	10.580	8.825	6.674	5.701	4.281	3.280

Cumulative Percentages of Real Distance Accounted for

39.419	58.660	69.240	78.065	84.739	90.440	94.721	98.001

Cumulative Percentages of Total (Real and Imaginary) Distance Accounted for

55.558	82.676	97.587	110.025	119.432	127.466	133.500	138.124

Trace 39696.560

Number of Dimensions in Real Space 11

Table 21. Spatial Coordinates of First Eight Dimensions for Experimental Group at Time Three

Concepts	1	2	3	4	5	6	7	8
Expert	-19.910	54.941	10.001	-10.300	-12.553	-3.228	-7.865	-15.712
Instructor	-14.712	13.894	-12.982	-14.226	19.478	-3.538	-17.299	26.581
Studying	12.860	17.782	.806	11.259	13.705	-27.746	-12.897	-4.734
Thinking	4.144	4.215	-11.627	-2.499	-1.314	-12.254	10.883	-4.028
Me	-1.662	-44.263	2.508	30.647	.599	-9.958	-3.924	-1.596
Teaching Assistant #2	6.789	-.999	-9.284	-18.101	11.711	-4.770	10.304	-16.376
Physical Science	74.503	7.506	-22.372	10.851	-4.075	3.850	9.920	-1.044
Leadership	-7.501	-12.224	-21.495	-.549	-13.133	36.166	-12.907	-3.773
Learning	-1.771	16.533	-6.495	21.567	-13.482	-7.562	1.897	14.474
Democratic	-48.431	-35.892	-18.376	-12.757	-14.512	-17.361	-.048	-7.190
Humanities	8.796	-23.126	34.693	-12.431	30.789	3.583	5.476	-.657
Good	-15.747	4.456	36.721	18.772	-22.016	.693	11.239	7.340
Confident	-24.722	13.360	4.772	27.196	23.389	24.832	-3.109	-9.660
Social Science	40.417	-17.204	27.802	-22.957	-19.731	3.647	-21.158	.233
Teaching Assistant #1	-11.118	7.986	4.768	-20.338	-2.847	9.830	25.864	11.372
Communication	-1.935	-6.968	-19.439	-6.133	3.992	3.816	3.625	4.769

Table 21 (cont'd.)

Eigenvalues (Roots) of Eigenvector Matrix								
	11497.317	8414.592	5587.626	4679.122	3838.517	3500.720	2331.763	1839.055
Number of Iterations to Derive the Root								
	7	8	16	12	14	18	12	8
Percentage of Distance Accounted for by Individual Vector								
	26.535	19.420	12.896	10.799	8.859	8.079	5.382	4.244
Cumulative Percentages of Real Distance Accounted for								
	26.535	45.956	58.852	69.651	78.510	86.589	91.971	96.215
Cumulative Percentages of Total (Real and Imaginary) Distance Accounted for								
	32.864	56.916	72.887	86.262	97.234	107.240	113.905	119.162

Trace 34984.858

Number of Dimensions in Real Space 12

Table 22. Spatial Coordinates of First Eight Dimensions for Control Group at Time Four

Concepts	1	2	3	4	5	6	7	8
Expert	-.602	-8.404	-16.143	-25.536	24.075	17.614	26.109	8.076
Instructor	16.731	9.940	-18.437	-.893	9.179	-26.941	1.807	12.619
Studying	-24.004	3.567	-7.247	-1.137	14.527	9.794	-6.874	-4.980
Thinking	-6.025	1.279	2.270	-10.678	-1.822	-8.822	5.089	-9.013
Me	-17.925	-36.180	9.568	-12.406	-19.160	-2.558	-20.822	-1.463
Teaching Assistant #2	59.979	28.098	13.972	1.501	.005	6.903	-9.792	-6.377
Physical Science	-44.484	59.923	-.740	-13.098	-16.455	-3.971	-2.448	-1.713
Leadership	22.349	-20.609	-.403	-17.496	-4.011	-15.503	11.673	-9.713
Learning	-16.517	-3.436	.157	-7.328	12.359	-14.998	-11.529	18.778
Democratic	4.481	-12.745	41.702	1.085	-27.641	-4.687	16.267	7.874
Humanities	-46.223	-17.031	-17.307	32.376	-4.602	2.064	9.008	-9.401
Good	5.809	-2.239	1.041	7.562	-12.106	31.482	-7.931	18.723
Confident	17.996	1.434	-23.268	-16.763	-16.241	13.046	1.885	-12.512
Social Science	-8.727	6.497	47.349	17.352	32.032	3.288	4.130	-7.328
Teaching Assistant #1	26.379	5.141	-23.065	44.468	-8.744	-6.075	6.616	2.969
Communication	10.782	-15.237	-9.450	.991	18.606	-.634	-23.988	-6.438

Table 22 (cont'd.)

Eigenvalues (Roots) of Eigenvector Matrix

10965.240	7070.008	6389.899	5122.775	4329.186	2980.724	2600.842	1586.104

Number of Iterations to Derive the Root

7	19	9	16	43	17	67	6

Percentage of Distance Accounted for by Individual Vector

25.993	16.759	15.147	12.144	10.262	7.066	6.165	3.760

Cumulative Percentages of Real Distance Accounted for

25.993	42.753	57.900	70.044	80.306	87.372	93.537	97.297

Cumulative Percentages of Total (Real and Imaginary) Distance Accounted for

34.710	57.090	77.316	93.532	107.236	116.671	124.904	129.925

Trace 31591.120

Number of Dimensions in Real Space 11

Table 23. Spatial Coordinates of First Eight Dimensions for Experimental Group at Time Four

Concepts	1	2	3	4	5	6	7	8
Expert	-1.924	38.551	-16.064	4.519	-.802	-13.061	1.736	3.717
Instructor	-22.929	5.547	-2.395	1.899	-17.899	.582	6.404	-14.225
Studying	37.251	7.031	-6.546	-11.209	-6.686	3.830	-8.551	.947
Thinking	7.075	8.479	-.268	-1.893	7.929	-.591	-1.649	-.584
Me	4.259	-23.860	8.855	-16.082	13.781	-4.109	-8.407	-4.556
Teaching Assistant #2	-4.492	-9.024	4.209	7.038	.086	13.417	2.267	.078
Physical Science	27.496	15.858	34.364	12.625	5.335	-.052	5.835	-.190
Leadership	-18.717	-5.260	-8.276	-6.861	-5.295	-1.976	-8.025	1.977
Learning	2.406	7.199	11.213	-14.021	-4.065	-2.023	-4.240	-3.584
Democratic	-1.399	-21.203	12.286	20.976	-16.813	-14.699	-10.073	2.656
Humanities	19.098	-20.664	-17.534	-11.187	-3.288	-12.895	18.756	4.019
Good	-26.885	-6.011	8.202	6.232	19.634	-6.531	9.108	2.166
Confident	-25.030	13.542	-1.082	-9.030	7.666	.098	-8.734	4.454
Social Science	13.905	-6.830	-30.886	22.442	12.746	7.697	-5.272	-6.929
Teaching Assistant #1	-6.540	-4.722	-1.118	1.867	-8.221	16.622	2.704	17.060
Communication	-3.575	1.366	5.040	-7.316	-4.107	13.690	8.139	-7.006

Table 23 (cont'd.)

Eigenvalues (Roots) of Eigenvector Matrix								
	5082.471	3786.879	3285.363	2110.576	1676.194	1338.800	1025.943	690.843
Number of Iterations to Derive the Root								
	7	8	8	23	14	12	8	208
Percentage of Distance Accounted for by Individual Vector								
	25.472	18.979	15.465	10.578	8.401	6.710	5.142	3.462
Cumulative Percentages of Real Distance Accounted for								
	25.472	44.451	60.916	71.494	79.894	86.604	91.746	95.208
Cumulative Percentages of Total (Real and Imaginary) Distance Accounted for								
	31.736	55.382	75.896	89.075	99.542	107.901	114.307	118.621

Trace 16014.908

Number of Dimensions in Real Space 12

APPENDIX C

Descriptive Statistics of Students from Experimental and Control Group Sections

Table 24. Descriptive Statistics of Students from Experimental and Control Group Sections.

	Control	Experimental
Class Activity		
Mean Percentage of Time on Lectures	29.75	33.04
Mean Percentage of Time in Discussions	29.75	29.80
Mean Percentage of Time in Exercises	39.25	31.87
Mean Age	19.02	19.19
Mean Grade Point Average	3.02	3.05
Number of Students in Sections		
Mean	62.5	62.5
Minimum	38	34
Maximum	71	76
Percentage of Female Students (Mean)	62.5	65.4
Percentage of Male Students (Mean)	37.5	34.6
Percentage of Communication Majors	6.9	11.4
Percentage of No Preference	26	19
ercentage of Freshman	57.9	49.2
Percentage of Upper Division (Jr and Sr)	28.3	27.7

REFERENCES

REFERENCES

Association for Educational Communications & Technology. Educational technology: Definition & glossary of terms. Washington, D.C., 1977.

Astin, A.W., & Lee, C.B.T. Current practices in the evaluation and training of college teachers. Educational Record, 1966, 47, 361-364.

Ausubel, D.P. Educational psychology: A cognitive view. New York: Holt, Rinehart & Winston, 1968.

Bourbour, A. & Goldberg, A. Interpersonal communication: Teaching strategies and resources. Falls Church, Va., Speech Communication Association, 1974, 85.

Barker, L.L. Listening behavior. Englewood Cliffs, N.J.: Prentice Hall, 1971.

Blackburn, R.T., & Clark, M. An assessment of faculty performance: Some correlates between administrator, colleague, student and self ratings. Sociology of Education, 1975, 242-256.

Bolles, R.C. Learning theory. New York: Holt, Rinehart and Winston, 1975.

Bruner, J.S. Toward a theory of instruction. Cambridge: Harvard University Press, 1966.

Bryant, P.T. By their fruits ye shall know them. Journal of Education, 1967, 38, 326-330.

Campbell, N.R. An account of the principles of measurement and calculation. London: Longmans Green, 1928.

Cantwell, Z.M., & Doyle, H.A. Instructional technology. Metuchen, N.J.: Scarecrow Press, 1974.

Centra, J. Effectiveness of student feedback in modifying college instruction. Journal of Educational Psychology, 1973, 65, 395-401.

Cody, M. An application of the multiple attribute measurement model: Measurement and manipulation of source credibility. Unpublished M.A. thesis. Michigan State University, East Lansing, Michigan, 1976.

Cohen, A.M., & Brawer, F.B. Measuring faculty performance. Washington D.C.: ERIC Clearinghouse for Junior College Information/American Association of Junior Colleges, 1969.

Costin, F., Greenough, W.T., & Menges, R.J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1972, 41, 511-535.

Cushman, D.P. The message-attitude-behavior-relationship. Bureau of Business Research, Michigan State University Press, forthcoming.

Deshpande, A.S., Webb, S.C., & Marks, E. Student perceptions of engineering instructor behaviors and their relationships to the evaluation of instructors and courses. American Educational Research Journal, 1970, 7, 289-305.

Doyle, K.O., Jr. Student evaluation of instruction. Lexington: D.C. Heath, 1975.

Doyle, K.O., Jr., & Whitely, S.E. Student ratings as criteria for effective teaching. American Educational Research Journal, 1974, 11(3), 259-274.

Eble, K.E. The recognition and evaluation of teaching. Salt Lake City, Utah. Project to improve college teaching, 1970.

Einstein, A. Relativity: The special and general theory. New York: Crown, 1961.

Fenker, R.M. Teacher evaluation at TCU: An analysis of the perceived roles of faculty by students, administrators, and faculty. In the Evaluation of College Teaching. R.I. Miller. New York: Jossey-Bass, 1974.

Frey, P.W. Student ratings of teaching: Validity of several rating factors. Science, 1973, 182, 83-85.

Friedrich, G.W., Galvin, K.M., & Book, C. Growing together: Classroom communication. Columbus, O.: Charles E. Merrill, 1976.

Genova, W.J., Madoff, M.K., Chin, R., & Thomas, G.B. Mutual benefit evaluation of faculty and administrators in higher education. Cambridge, Mass.: Ballinger Publishing, 1976.

Gessner, P. Evaluation of instruction. Science, 1973, 180, 566-570.

Gillham, J., & Woelfel, J.P. The galileo system: Preliminary evidence for precision, stability, and equivalence to traditional measures. Human Communication Research, Fall, 1976.

Gray, C.F. The teaching model and evaluation of teaching effectiveness. Improving College and University Teaching, 1970, 18, 275-278.

Grush, J.E., & Costin, F. The student as consumer of the teaching process. American Educational Research Journal, 1975, 12, 55-66.

Gustad, J.W. Politics and practices in faculty evaluation. Educational Record, 1961, 42, 194-211.

Hammond, P.E., Meyer, J., & Miller, D. Teaching versus research; sources of misperceptions. Journal of Higher Education, 1969, 40, 682-690.

Hebb, D.O. A textbook of psychology. Philadelphia: W.B. Saunders, 1958.

Hebb, D.O. The American revolution. American Psychologist, 1960.

Helm, C.E., Messick, S., & Tucker, L. Psychophysical law and scaling methods. Princeton: Educational Testing Service, 1959.

Hildebrand, M. How to recommend promotion for a mediocre teacher without actually lying. Journal of Higher Education, 1972, 43, 44-62.

Hilgard, E.R., & Bower, G.H. Theories of learning. New Jersey: Prentice-Hall, 1975.

Howe, H. Less teaching, more conversation. In C.B.T. Lee (ed.), Improving College Teaching. Washington, D.C.: American Council on Education, 1967.

Jacobi, C.G.J. Uber ein leichles Verfahren die in der Theorie der Sacularstorvngen vorkomnenden Gleichuagen numerisch aufzulosen. Jouranl fur die Reine und Angewandyte Mathematik, 30:51-94.

Johnson, D.W. Reaching out: Interpersonal effectiveness and self-actualization. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Kauffman, J., Hallahan, D., Payne, J., & Ball, D. Teaching/learning: Quantitative and functional analysis of education performance. Journal of Special Education, 1973, 7, 261-268.

Keast, W.R., & Macy, J.W., Jr. Faculty tenure. San Francisco: Jossey-Bass, 1973.

Kent, L. Student evaluation of teaching. Educational Record, 1966, 47, 376-390.

Koffka, K. Principles of gestalt psychology. New York: Harcourt-Brace, 1935.

Kohler, W. Gestalt psychology. New York: Liveright, 1947.

Lewis, L.S. Scaling the ivory tower. Baltimore, Maryland: The John Hopkins University Press, 1975.

Liska, A.E. The consistency controversy. New York: John Wiley, 1975.

Luthans, F. The faculty promotion process and analysis of the management of large state universities. Iowa City, Iowa: Bureau of Business and Economic Research, College of Business Administration, the Unviersity of Iowa, 1967.

McKeachie, W.J., Lin, Y., & Mann, W. Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal, 1971, 8, 435-445.

Miller, R.I. Evaluating faculty performance. San Francisco: Jossey-Bass, 1972.

Morine, H., & Morine G. Discovery: A challenge to teachers. Englewood Cliffs, N.J.: Prentice Hall, 1973.

Mortensen, C.D. Communication: The study of human interaction. New York: McGraw-Hill, 1972.

Nobel, C. An analysis of meaning. Psychological Review, 1952, 59, 421-30.

Osgood, C.E. A behavioristic analysis of perception and language as cognitive phenomena. In Contemporary Approaches to Cognition, Cambridge: Harvard University Press, 1957.

Pittenger, O.E., & Gooding, C.T. Learning theories in educational practice. New York: John Wiley and Sons, 1971.

Pfeiffer, J.W., & Jones, J.E. A handbook of structured experiences for human relations training (3 vols.). Iowa City, Iowa: University Associates Press, 1971.

Rippey, R. Student evaluations of professors: Are they of value? Journal of Medical Education, 1975, 50, 951-954.

Serota, K. Metric multidimensional scaling and communication: Theory and implementation. Unpublished M.A. thesis. Michigan State University, East Lansing, Michigan, 1974.

Sharon, A.T. Eliminating bias from student ratings of college instructors. Journal of Applied Psychology, 1970, 54, 278-281.

Snelbecker, G.E. Learning theory, instructional theory, and psychoeducational design. New York: McGraw-Hill, 1974.

Stevens, S. Mathematics, measurement, and psychophysics. In S. Stevens (ed.), Handbook of Experimental Psychology (1st ed.). New York: Wiley, 1951.

Suppes, P., & Zinnes, J.L. Basic measurement theory. In R. D. Luce, R.R. Bush, & E. Golanter (Eds.), Handbook of Mathematical Psychology. New York: John Wiley and Sons, 1963, 1-76.

Tolman, E.C. Purposive behavior in animals and men. New York: Appleton-Century-Crofts, 1932.

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

Tyler, R.W. The evaluation of teaching. In R.M. Cooper (Ed.), The Two Sides of the Log. Minneapolis: University of Minnesota Press, 1958.

Wakshlag, J., & Edison, N. A model for the validation of the source credibility construct. Paper presented at the Speech Communication Association Annual Convention, December 1975.

Wertheimer, M. Productive thinking. New York: Harper, 1959.

Williams, R.E. Communication: Definitions, concepts and models. In J. Shubert (Ed.), Human Communication: Concepts, Principles and Skills. East Lansing, Michigan: Michigan State University, 1974.

Woelfel, J. Sociology and science. Unpublished manuscript. Michigan State University, East Lansing, Michigan, 1972.

Woelfel, J. Metric measurements of cultural processes. Paper presented to the Speech Communication Association, Chicago, 1974.

Woelfel, J. Foundations of cognitive theory. In D.P. Cushman (Ed.), The Message-Attitudes-Behavior Relationship, Bureau of Business Research, Michigan State University Press, forthcoming.

Woelfel, J., & Danes, J. New techniques for the multidimensional analysis of communication and conception. In P. Monge and J. Cappella (Eds.), Multivariate Analysis in Communication Research, forthcoming.

Woelfel, J. Cody, M., Gillham, J., & Holmes, R. Basic premises of multidimensional attitude change theory-an experimental analysis. Unpublished manuscript. Michigan State University, East Lansing, Michigan, 1977.

Woelfel, J., & Saltiel, J. Cognitive processes as motions in a multidimensional space: A general linear model. Unpublished manuscript. Department of Communication, Michigan State University, East Lansing, Michigan, 1974.

Wotruba, T.R., & Wright, P. How to develop a teacher-rating instrument. Journal of Higher Education, 1975, 46, 653-663.

Young, G., & Householder, A. Discussion of a set of points in terms of their mutual distances. Psychometrika, 1938, 3, 19-22.