

5-30-1974

# Development of a Teacher Rating Instrument: Methodological Implications

Carole Gygi  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Education Commons](#), and the [Urban Studies Commons](#)

Let us know how access to this document benefits you.

---

## Recommended Citation

Gygi, Carole, "Development of a Teacher Rating Instrument: Methodological Implications" (1974).  
*Dissertations and Theses*. Paper 536.  
<https://doi.org/10.15760/etd.536>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

DEVELOPMENT OF A TEACHER RATING INSTRUMENT:  
METHODOLOGICAL IMPLICATIONS

by

CAROLE GYGI

A dissertation submitted in partial fulfillment of the  
requirements for the degree of

DOCTOR OF PHILOSOPHY  
in  
URBAN STUDIES

Portland State University  
1974


AN ABSTRACT OF THE DISSERTATION OF Carole Gygi for the Doctor of  
Philosophy in Urban Studies presented May 30, 1974.

Title: Development of a Teacher Rating Instrument:  
Methodological Implications.

APPROVED BY MEMBERS OF THE DISSERTATION COMMITTEE:

  
Dennis L. West, Chairman

  
Hugo Maynard

  
Cord B. Sengstake

  
Robert F. Powloski

In order to gain a more comprehensive understanding of the complexities which constitute an effective teacher and promote a positive learning climate, a field study methodology was employed to develop an evaluative instrument and to gain descriptive data.

For the purpose of developing an instrument to describe and measure effective teachers, 24 actual learning sessions (classes) were observed, extensive notes taken, and tape recordings were made in order

to isolate and describe the behaviors which seemed critical to the situation being studied. This was done by observing the teacher as he taught and the learners while they learned. The teacher was reputed to be an "effective" teacher and was chosen for this reason. At the end of the 24 sessions the perceptions, observations, and inferences of the investigator were measured against those of the learners in the same classroom situation to determine how congruent they were. The results of the study also confirmed the reputation of the teacher.

This information was utilized in the development of scale items. In addition, a second study was designed as a measure of the same teacher, teaching another course, and different students (92 second year medical students). A questionnaire was designed to test the overall effectiveness of the teacher, by the students, and just as important as the rating was the information elicited from the students regarding the teacher and the class process. This was done by way of open ended questions, and the coding of these questions by the investigator. This information too became useful in developing scale items.

Once the scale items were developed by the empirical method described above, the items were pretested on teachers teaching in the same department as the exemplar teacher. The results of the pre-test were statistically significant correlations between the scale items and a student rating of the overall effectiveness of the teachers being studied. The scale items were refined and tested on another larger and different sample of teachers.


The teachers in this sample were teachers at Portland State University in various departments. Thirty teachers participated in the testing of the instrument. This part of the study was designed to compare the instrument developed by the investigator using the process of a trained observer (direct observation) against a study designed and developed by the traditional survey method of scale development which utilized factor analysis to select scale items.

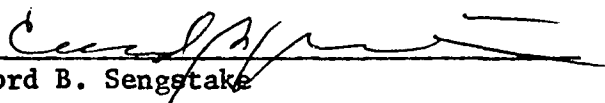
The correlation between the two scales (a split model design was used--one half the students answered the investigator's instrument, one half the students answered the criterion instrument) was nearly perfect.

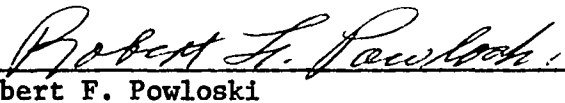
TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the dissertation of  
Carole Gygi presented May 30, 1974.

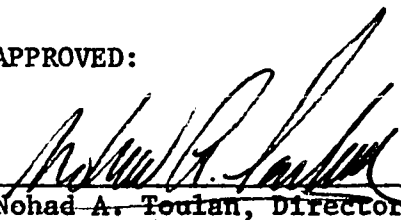
  
Dennis A. West, Chairman

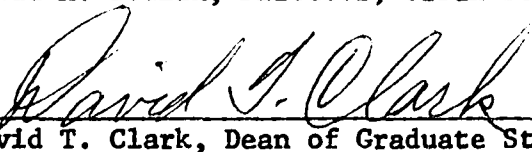
  
Hugo Maynard

  
Cord B. Sengstake

  
Robert F. Powloski

APPROVED:

  
Nohad A. Toulan, Director, Urban Studies

  
David T. Clark, Dean of Graduate Studies and Research

## ACKNOWLEDGMENTS

My Committee: Dr. D. West, Dr. H. Maynard, Dr. C. Sengstake,

Dr. R. Powloski

Faculty Members: Dr. J. Paulson, Dr. R. Jennings, Dr. D. Wrench,

Dr. J. Lansdowne, E. Lawrence, Dr. M. Weitman

My Friends: Joyce Lindhe, Janet Lahti, Nancy Chafin, Michael Davis,

Mary Ann Buchanan, Dennis Buchanan

Special Acknowledgments: Dr. G. Saslow, Dr. J. Blumel, Dr. C.

Goodmonson, Dr. J. Lansdowne, J. Behn,

Dr. D. Parker

this dissertation is dedicated to

the memory of

FRANZ KAFKA

and to

jerry lansdowne, courtney  
goodmonson, george saslow,  
hugo maynard, dennis west,  
kathleen gygi, brian gygi,  
renate gygi, who refuse to  
accept senseless paradigms.



TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS . . . . .	iii
LIST OF TABLES . . . . .	vii
CHAPTER	
I INTRODUCTION. . . . .	1
Traditional Methods of Social Research. . . . .	4
General Description of the Development of a Rating Scale by Traditional Methods . . . . .	9
General Description of the Dissertation . . . . .	11
II SURVEY OF LITERATURE. . . . .	13
Field Study Techniques. . . . .	13
Problem of Teacher Evaluation . . . . .	14
Description of Traditional Approaches . . . . .	14
Teaching Methods . . . . .	15
Teacher Attributes . . . . .	19
III METHODOLOGICAL DESCRIPTIONS. . . . .	29
Development of Scale Items . . . . .	29
Traditional Methods. . . . .	29
Dissertation Research Paradigm . . . . .	31

CHAPTER	PAGE
IV METHODOLOGY AND RESULTS . . . . .	42
Empirical Testing of Scale Items. . . . .	42
Introduction. . . . .	43
Subjects. . . . .	43
Procedure . . . . .	46
Results . . . . .	49
V DISCUSSION. . . . .	52
Availability of Observational Model . . . . .	53
Availability of Trained Observers . . . . .	54
Behavioral Unit or Act of the Observational System. . . . .	55
Strengths of this Method. . . . .	56
Implications for Research Areas Outside Education . . . . .	57
Summary and Implications. . . . .	58
REFERENCES. . . . .	60
APPENDIX	
A NARRATIVE DATA . . . . .	65
B PRELIMINARY QUESTIONNAIRE 1 . . . . .	76
C QUESTIONNAIRE DATA 2. . . . .	84
D PILOT TESTING OF PRELIMINARY SCALE ITEMS. . . . .	87
E TEACHER CHARACTERISTICS . . . . .	92
F DIMENSIONS OF AN EFFECTIVE TEACHER. . . . .	101

LIST OF TABLES

TABLE		PAGE
I	Progression from Raw Data to Major Categories . . .	37
II	Comparison of Traditional Methodology and Trained Observer Method of Teacher Evaluations. . . . .	40
III	Investigator's Scale. . . . .	43
IV	Berkeley Scale. . . . .	47
V	Correlation Among Means . . . . .	50

## CHAPTER I

### INTRODUCTION

The area of social research is faced with many difficult problems. The major problem is that social research seems to lack any clear-cut theory or generally accepted methodology. Therefore, there are many interpretations possible for any given social situation. There are other serious obstacles to overcome in trying to understand the behaviors of persons in a natural situation. To list a few, the sample of persons or of social behaviors to be studied must be identified. Some decision has to be made whether all behaviors will be observed, whether only selected behaviors will be studied, or whether time interval behaviors (i.e. periodic observations) will be observed. This observational unit (behavioral cues) determines the potential variables which become the data for the investigation. Once the sample data from the observational unit have been selected, the potential variables for study within that sample appear to be endless. On what basis does one choose from amongst the many potential independent variables? Since we lack any accepted theoretical framework one is left with having no real basis for selection of the important variables. Another serious problem is that unless one uses direct observation in the research design, the data are always of an antecedent nature. Such data are

notoriously unreliable because memories of past events are so easily distorted. Still another obstacle is that the sample used is virtually always of a non-random nature. Usually social situations being studied are comprised of persons who have allied themselves for a particular reason, self-selection for example, and it is difficult to break up these groups. This lack of randomness, of course, violates one of the cardinal principles of research in which cause-effect relationships are sought.

In view of these serious obstacles there are three readily identifiable social research needs. The first is a need for carefully conducted, rigorously designed, empirical studies of human functioning in natural settings to compliment the present heavy emphasis on laboratory research.

Behavioral science research methodology has reached a point in its development where, despite the greater number of uncontrolled variables, rigorous studies conducted in natural settings should match if not surpass laboratory research (Hutt and Hutt, 1970).

Without real world testing the artificial environment created by the laboratory situation may generate results which are peculiar to that laboratory setting. Perhaps persons involved in experiments view themselves as subjects in an experimental situation and may react differently than they would if the same stimuli was received by them in their own environment. This puts laboratory data in question. There are many interesting data which have not been tested outside of the laboratory and whether or not these data indicate needed changes in

terms of the real world is uncertain. Too often the study of social behaviors, however defined, has suffered from a lack of primary data (i.e. data that is collected as it occurs). Usually the data are in some degree, removed from the actual phenomena under investigation.

The second social research deficiency is the lack of research replication. In most instances even important behavioral research is not replicated. It is generally accepted that research must be replicable. This is true in social research as well as in other areas of scientific investigation.

A third major deficiency in the social research area is in the nature of the dependent variable. It is very difficult to publicly specify the dependent variable in social research. Most researchers rely upon three kinds of research vehicles--tests, questionnaires, and interview data. What then is one studying, the results of the test or questionnaire, or the social phenomena the data is supposed to reflect?

When test instruments are utilized the data are then based on a single performance measure which is easily affected by such variables as anxiety, illness, reaction to testing environment, etc. When data are so easily contaminated it is difficult to assess its reliability. It is a well known fact that a person seldom if ever obtains the same score twice on any test instrument. This includes projective techniques, scholastic achievement tests, and IQ tests. This raises the question of what is being measured. Are the data obtained from any

given measuring instrument the true measure of the property measured? If the measuring instrument varies from testing to testing it is then considered an unreliable indicator of the person's attributes being tested. Since the validity of the data generated cannot be higher than the reliability of the instrument being used in the data collection, reliability is a necessary condition of valid research results.

### Traditional Methods of Social Research

Social science research can be divided into four major categories: laboratory experiments, field experiments, field studies, and survey research. This breakdown is derived from two sources, the distinction between experimental and non-experimental research and between laboratory and field research. The basic difference between experimental and non-experimental or ex post facto research is simple. Experimental studies hypothesize if x, then y; if frustration, then aggression. The researcher working in the experimental model of research design can use some method to manipulate the independent variable. He then observes the dependent variable to see if concomitant variation (the predicted variation from the manipulation of the independent variable) occurs. In ex post facto research, the dependent variable is observed. Then a retrospective search for the independent variable ensues. Ex post facto research permits no control of the independent variable and randomization is therefore not possible. This kind of investigation must take things as they occur naturally. Therefore no statement concerning causal relations are justified from the

ex post facto research, only that a relationship does exist.

These four kinds of social research are described on the following pages.

1. A laboratory experiment is a research study in which the variance of influential independent variables not pertinent to the immediate problem of the investigation is kept at a minimum. This is done by isolating the research in a physical situation apart from the routine of ordinary living and by manipulating one or more independent variables under rigorously specified, operationalized, and controlled conditions (Kerlinger, 1964). This kind of research has the inherent value of control. In addition to situational control, laboratory experimenters can usually use random assignment and manipulate one or more independent variables at will. In this kind of research it is relatively easy to specify the operational definitions of the variables. The greatest weakness of the laboratory experiment is the lack of strength of the independent variable. Laboratory situations are different than real life situations. Only with great care can the data generated in laboratory situation be generalized beyond the laboratory.

In an example of this kind of research the subject was placed in a conflict situation (Milgram, 1963). Persons were asked to deliver levels of electric shock which could cause extreme pain to another person. Results of the experiment gave frightening evidence about the extent to which people will follow the commands of authority even when those commands may require them to violate moral standards. It is



possible, however, that people were willing to violate their moral standards in the exceptional situation of the laboratory setting and would not do so outside such an obviously contrived situation.

2. A field experiment is a research study in a realistic situation in which one or more independent variables are manipulated by the experimenter under as carefully controlled conditions as the situation will permit (Kerlinger, 1964). The manipulation of independent variables and the possibility of randomization are the most important strengths of the field experiment. The control of the field situation is not as tight as the control in the laboratory situation. In a field it is always possible that the independent variables are contaminated by uncontrolled environmental variables. The variables in a field experiment usually have a stronger effect than those of laboratory experiments. Realism increases the strength of the variables. The more realistic the situation, the more valid are generalizations to other situations likely to be. The field experiment is appropriate for studying complex social influences, processes, and changes in lifelike settings. The main weakness of the field experiment is that of practical difficulties. There is no theoretical reason why randomization cannot be used here, but there are obstacles to it like unwillingness to break up class groups, or to allow children to be assigned to experimental groups at random. A field investigator needs skills such as communication skills which will enable him to work with people, talk to them and convince them that they should participate in his study.

An example of this kind of research is a study done by Verplanck (1955) in which he was able to alter the extent to which his subjects expressed opinions during normal conversations without their being aware that they were being studied. Again one must be aware of the interaction of independent variables with the natural environment. This interaction cannot be controlled.

3. Field studies are ex post facto studies aimed at discovering the relations and interactions among sociological, psychological, and educational variables in real social structures.

Any study that systematically pursues relations and tests hypotheses (that are ex post facto) which are made in life situations such as communities, schools, factories, organizations, and institutions is an example of a field study. The investigator in a field study looks at an existing situation and then studies the relations among attitudes, values, perceptions, and behaviors of individuals and groups in the situation. Field studies are strong in realism, significance, strength of variables, theory orientation, and heuristic quality. The variance of many variables in actual field settings is large. Because of the strength of variables there is usually so much noise in the communication channel that even though the effects may be strong and the variance great, it is often hard for the researcher to separate the variables.

There is no artificiality in the field study. Of all types of studies discussed they are the closest to real life. However, the

field study is a scientifically weak study in comparison to the laboratory study. Its most serious weakness is its ex post facto nature. Thus statements of causal relations are not possible. Another weakness of this kind of study is the lack of precision in the measurement of field variables. Other problems in this kind of research are practical problems: feasibility, cost, sampling, and time. Field studies are very difficult. They require large amounts of time, energy, and skill. Questions to be considered are: Can the study be done with the facilities available to the researcher? Can the variables be measured? Will it cost too much? Will it take too much time and effort? Will the subjects cooperate?

An example of this kind of study is one done by Getzels and Guba (1954) in which the role conflict and role-taking effectiveness was studied. They studied officers who were also instructors at an air force school. The amount of conflict between the role of an officer as officer and his role as instructor was related to his rated effectiveness. It was found that the more acute this conflict became, the more ineffective the officer tended to be.

4. Survey research is that branch of social scientific investigation that studies large and small populations (or universes) by selecting and studying samples chosen from the populations to discover the relative incidence, distribution, and inter-relations of sociological and psychological variables. Surveys covered by this definition are often called sample surveys. Survey research has the advantage of

wide scope. Survey research information is accurate--within sampling error ranges, of course. There are weaknesses in survey research. The information may be superficial. Survey research is costly and time consuming. It is subject to sampling error. The survey interview can temporarily lift the subject out of his normal social context, which may contaminate the results. Survey research requires a good deal of research knowledge and sophistication. A good example of this kind of research is the Gallop Poll, and the Neilson TV Ratings.

One of the recurring and serious problems in measurement is to bridge the gap between behavior and construct. Competent observers and well-made observations can help bridge this gap. There is a need for descriptive studies. They can provide details about human and social behavioral patterns. Almost every other science is replete with catalogs and handbooks of facts about the phenomena it covers. Behavioral sciences have barely begun to accumulate and classify this kind of data. Without this descriptive information there is a gap. Inappropriate variables can be selected for study, meaningless hypotheses can be tested and erroneous inferences can easily be drawn (Brandt, 1973).

#### General Description of the Development of a Rating Scale by Traditional Methods

In studying and trying to understand what goes into making a positive learning climate, and producing an effective teacher, one must use some measure or measures. The usual process is to develop an instrument in which "a general picture of good teaching can be developed, and

characteristics of a good teacher emerge" (Miller, 1973).

For example, in a study done at the Center for Research and Development in Higher Education at Berkeley a selected sampling of students indicated from an extensive list of teacher traits, which of the traits they considered important in good teaching and in a good teacher. The study took three years to complete, involved more than 1600 students and faculty and made use of three questionnaires. One dealt with biographical information and students' academic backgrounds, college goals, and objectives they valued in teaching. The second asked the faculty to identify a best and a worst teacher among their colleagues and to answer, for each, questions about teaching activities observed outside the classroom, about in-class behavior, and about the presentation of talks and seminars. The third questionnaire dealt with the distribution of time among various academic pursuits. Questionnaires were distributed in May 1967 and 1968, and a follow-up validation study was distributed to fifty-one classes. The classes selected included, in about equal numbers, those instructors identified in 1967 as best teachers by three or more students or colleagues, those instructors identified as worst teachers, and those instructors not previously identified as either best or worst, and presumed to be teachers of intermediate effectiveness. The 1,015 respondents provided biographical data and answered questions about their college goals, various objectives of teaching, and the teaching of the given instructor. Ratings of the overall effectiveness of the teachers were also secured

(Center for Research and Development in Higher Education, University of California at Berkeley, 1971). This study was later used as the comparison study for this investigator's instrument.

Another example of this type, centering on criteria for effective teaching, was completed by the University of Toledo's Office of Institutional Research (1969). Thirteen thousand six hundred and forty-three responses from students, faculty, and alumni of the university were read and categorized into sixty effective teaching behaviors. These behaviors were then rated as to importance by 1,793 students, faculty and alumni (Miller, 1972).

#### General Description of the Dissertation

The dissertation which will be described was a study done in an attempt to deal with the apparent need for field studies conducted as rigorously as possible with the gap between behavior and construct closed by the utilization of direct observation. The study adequately dealt with problems of a practical nature. An actual social situation was utilized for intensive and extensive observation. The cost was minimal, requiring only one trained observer for 24 one-hour sessions and the extrapolation of the information gained therein into an instrument comparable in terms of reliability to one developed by a traditional questionnaire (factor analysis) method. The trained observer recorded, described, and measured the social situation being observed.

The study was done in the following way: a trained observer observed a social situation. Based on these observations she developed observational categories. These categories were formulated as the observer noted the behaviors, and then the classes of behaviors were subjectively assigned a name. These classes became summary statements which included what was considered to be the important content and context of the observed social interaction. These summary statements then became items of an evaluative/descriptive instrument. Each of these steps will be fully described as the study unfolds.

It is proposed that the observational system used in this dissertation can be adapted and used to observe, record, describe, and evaluate any social situation.

The methodological design of the study appeared to solve some of the problems encountered in traditional field studies. It seemed to offer an alternative to the traditional method of doing field studies and added descriptive information to the area of social research.

In brief, the teaching-learning situation was chosen as an ideal social interaction, occurring at specified times and contained within a specified time frame. It is also one of the most important social concerns of urban life. This research will demonstrate that observational methods are rigorous and can produce results at least as good as more traditional methods.

## CHAPTER II

### SURVEY OF LITERATURE

#### Field Study Techniques

The dissertation being reported uses the ex post facto method of a field study. As Kerlinger (1964) points out, in spite of weaknesses ex post facto studies must be done, and done as rigorously as possible because some areas such as psychology, sociology, and education do not lend themselves to experimental inquiry. For example, in education important variables such as intelligence, aptitude, home background, parental upbringing, teacher personality, school atmosphere--are not manipulable. It is possible and important to do careful and controlled inquiry into these areas. Improvements in educational ex post facto research are badly needed.

The need for careful inquiry was one of the reasons that the area of education was selected to illustrate the use of the trained observer approach to scale development. There were other important reasons. There is much unhappiness with education in this country. Education is an area of social concern that is being studied and studied, but the problems besetting the educational system are not yet clear and they continue to mount.



There were two other reasons that the area of education was chosen as an illustration of the trained observer technique. 1) There is an enormous amount of educational research available for background baseline data. 2) The classroom situation is a relatively long lasting social situation. The stability of the social situation in a classroom provides an observation opportunity for extensive and indepth observation.

### Problem of Teacher Evaluation

Teacher evaluation is not a new problem; yet it is one that has never been faced squarely. Today faculty evaluation is the most explosive area in higher education. Students want it. Legislators think it should be required. Administrators regard it as a necessity. Younger faculty are willing to take their chances, but older faculty regard it as a threat, crying unionization and academic freedom all the way. But union or no unions, tenure or no tenure, evaluation is inevitable. If faculty and administration cannot resolve the problem then outside agencies will. State legislators are already talking about it. Professional associations such as the National Educational Association are getting involved (Miller, 1972).

It is evident from Miller's statement that teacher evaluation is a pressing problem which social researchers should take as a serious problem. The time has come to do something more than theorize and hypothesize. Changes need to be made.

### Description of Traditional Approaches

While much research has been done in an effort to evaluate effective teaching, very little useable or conclusive information has

been obtained. A survey of the literature indicates the myriad approaches to the evaluation of teaching. The results are chaotic. The literature surveyed for this dissertation was chosen from the thousands of studies available and grouped into categories in order to give the reader an impression of this kind of research.

Work done in an effort to ferret out aspects of effective teachers and effective teaching methods is too voluminous to attempt a full review. The survey reflects, therefore, selected literature which seemed especially important to this dissertation.

Some of the studies cited are extremely broad and ill defined. Others are limited in scope and not well controlled. The literature is contradictory in nearly every phase of investigation. For many studies supporting a given notion or theoretical position, another can be cited to contradict it.

With this in mind, this study will turn to a review of the research in the areas of teaching methods, teacher attributes and the problem of teacher evaluation.

### Teaching Methods

Over the years a number of general teaching methods have evolved which imply particular patterns of teacher behavior as well as modifications in objectives and content. Among the best known of these are: the recitation method, lecture method, discussion method, laboratory (or project method) and the problem-solving method (Wallen and Travers,

1963). With respect to mastery of factual information, or for that matter or any specific goal in education, the findings indicate that there are no significant differences in effectiveness between lecture and discussion methods. This generalization holds for all levels of education. "When one asks whether lecture is better than discussion, the appropriate counter would seem to be, 'For what purpose?'"

(McKeachie, 1956). Studies by Ash, (1951); Bane, (1931); Bills, (1952); Carlsen, (1953); and Maloney, (1956), all support the notion that there is no difference between lecture and discussion.

Wallen and Travers (1963) thoroughly reviewed the research done in the area of analysis and investigation of teaching methods from 1933 to approximately 1960. They concluded that little has been done to develop teaching methods on the basis of scientific knowledge of learning. They further concluded that studies comparing teaching methods that were being pursued at the time their report was being written were only in the beginning stages of development.

In order to demonstrate the contradictions in this area of research, Bane (1931) and Rickard (1946) both found retention of factual knowledge to be superior in groups taught by the discussion method. Also Ward (1956) found greater retention of the "understanding" type of learning among students with greater academic ability if they were taught with techniques which used discussion procedures.

Proponents of the problem-solving method have indicated that it is the most effective way to teach. Patton (1955) did a dissertation

on the effects of this method on student responsibility and motivation and found that the problem-solving approach in teaching tended to encourage students to accept more responsibility for learning (McKeachie, 1954; and Perkins, 1950). All found that problem-solving teaching techniques encourage the students to use more psychological knowledge (i.e., to involve themselves more) and to interact better with others.

The laboratory method with its emphasis on direct experience with people and materials pertinent to the subject to be studied has been too little studied to allow one to draw any conclusions or evaluate its relative effectiveness. The studies that have been done usually report contradictory results, which suggest that this is probably the most idiosyncratic of the methods used. Some students probably thrive and learn best working this way. Others can't work so well. Balcziaik (1954) and White (1945) did studies at the college level and obtained the usual contradictory results.

There are other kinds of research which have at least a peripheral bearing on teaching techniques. Guidance is seen as the critical variable in effective teaching for Stones (1968). He states that it is possible to guide learning by careful reinforcement. He is an advocate of programmed learning and feels that such guidance and self-programming shortened the learning process (1968). He wrote a programmed text on programmed learning which he asserts proves his point. Greenhill (1963) maintained that there is a learning interaction between the learner and the instructional materials or resources which communicate the knowledge

to be learned, apart from the teacher.

Sprights (1967) used 239 students (two samples) to see if 1) high-achieving students have more favorable perceptions of instructor's teaching methods than low-achieving students, or 2) high achieving students view the personal attitudes of college instructors more favorably than low achieving students. The research indicated that students at the college level (of both high and low grade-point averages) think there is too much emphasis on the lecture method. Both also agree to a general lack of independent study. Above-average students favor greater use of audio-visual aids and more use of essay examinations.

Additional information about teaching techniques or methods have been reported. Flanders (1951), using a conception of teacher-centered versus learner-centered teaching, trained one teacher to manifest each pattern while teaching each of seven one-student "classes" in the interpretation of human behavior. Analysis of teacher statements and of Q sorts by students supported the adequacy of the role playing. Each student experienced both procedures. Flanders concludes that the teacher-centered role fostered more negative feelings, a greater concern with interpersonal as opposed to "learning" problems, higher physiological indices of anxiety, and less content mastery on the part of students.

In 1956 Ackerman did his dissertation on presentation of alternatives and its relation to set in problem solving. He found that intermediate grade students who were taught only one way of solving problems

were less flexible and saw fewer alternatives than those taught by two alternative approaches.

### Teacher Attributes

When teacher attributes are considered to be a reflection of personality, research attention centers on differences among teachers in ability, knowledge, attitude, temperament, and the like. Personality traits are inferences from relatively large sample of behavior. "Intelligence," "knowledge," and "authoritarianism" are some of the variables Gage (1963) used in studying personalities of teachers.

Clark (1950) studied 181 female elementary school teachers in a Midwestern city and a control group of college students to determine personality traits of teacher vs. non-teacher. Significant differences between the teachers and the comparison group were found as follows:

Teachers were above average in Objectivity, Agreeableness, and Cooperativeness. They were less introversive in their thinking, showed fewer signs of Depression and Cycloid Disposition, and had lower scores on General Activity and Ascendance-Submission than did the normative group.

Ryans (1960) directed a major project over a six year period which included 100 separate research projects in which more than 6,000 teachers in 1,700 schools and 450 school systems participated. Factor analysis of data from those parts of the project concerned with observation of teacher's behavior identified three personality factors which seemed to distinguish between good and poor teachers. These were:

1. Warm, understanding, friendly versus aloof, egocentric, restricted teacher behavior.

2. Responsible, businesslike, systematic versus dull, routine teacher behavior.

Pogue (1967) did a study on Students' Ratings of the "Ideal Teacher." He used the evaluation form prepared by Quick and Wolf at the University of Oregon. The sample included 307 students (entire student body) at Philander Smith College. He found that the most important characteristics of a good teacher are good knowledge of the subject and clear explanations.

Symonds (1955) did a study on "Characteristics of the Effective Teacher Based on Pupil Evaluations" and found that superior teachers liked children, were personally secure, self-assured, well integrated and possessed good personality organization, while inferior teachers disliked children, were insecure and personally disorganized. Symonds said in that study:

It seems evident that the verbalized teacher image is a widely shared and extremely stable stereotype. However, that the stability of this image is of little help in predicting teaching since the great variety of classroom behaviors among effective teachers seems to preclude the use of observation as a tool for distinguishing effective teachers; . . . the basic determinants are to be found in the personality structure of the teacher rather than in outward behavior.

Other investigators such as Getzels and Jackson (1963), also have posited that personality attributes of the teacher are the most critical variables whether or not a teacher is considered effective.

They asked high school teachers to rank these roles in order of importance and arrived at the following order: 1) mediator of the culture; 2) member of a community; 3) director of learning; 4) guidance and counseling person; 5) liaison between school and community; and 6) member of a profession.

In a series of studies Anderson (1946) determined that high teacher dominance tended to be associated with a variety of student behaviors which seemed undesirable in the learning situation such as-- failure to carry out requests, whispering, not paying attention, etc.

Other global attributes have been mentioned in the literature such as: teachers are salesmen (Graves, 1956); teacher should act as guide, philosopher and friend, always ready to help (Posthethwait, 1965); attributes of a good teacher are: the ability to relate to students; the ability to synthesize both conceptual and factual knowledge, the ability to communicate, to articulate and project ideas in a clear and colorful manner, is dramatic and melodramatic, and he likes to teach. These studies were noted by the investigator but were of no value because they seemed too general or lacked well designed controls.

The point here is that research in the area of teacher attributes, as in the case for teaching methods, is so global and contradictory as to render it useless as a guide to educational policy.

An Investigation of Student-Teacher Interaction as a Determiner of Effective Teaching done by Lewis (1964) was to determine whether



students and teachers tend to interact along measurable personality dimensions. He used three groups--groups one and two provided a control for sex (male) and variation of subject matter, while the third group provided a variation of sex. Each student, as well as selected instructors in various fields, completed two questionnaires--the Guilford-Zimmerman Temperament Survey, and a one-hundred item biographical inventory. The results indicated that effective teachers cannot be differentiated from less effective teachers on the basis of personality variables.

Students appear to prefer different types of college professors in the following order: teacher, researcher, socialite, and administrator. This information was secured from a study done by Yamamoto and Dizney (1966), in which 300 college students responded to a Likert-type inventory of eight types of college professors.

In an extended study done by Fishburn (1955), evidence was gathered to show that six areas represent some specialization of roles among teachers.

Logan Wilson, former President of the University of Texas, wrote over twenty-five years ago in The Academic Man:

Indeed it is no exaggeration to say that the most critical problem confronted in the social organization of any university is the proper evaluation of faculty services and the giving due recognition through impartial assignment of status.

Some people contend that teaching is an art and far too complex ever to be reduced to a set of questions or too subtle to be

perceptably observed by outsiders. Exhaustive research indicates that many variables interact in determining teaching effectiveness. But it is the very complexity of the teaching situation that makes every bit of empirical information the more precious (Miller, 1972).

How do we know that the professor who is highly rated by students is, in fact, an outstanding teacher? The question is a fair one in view of the inadequate research basis that has been built in this particular area, but we do have some research evidence. The Berkeley study indicated (Hildegard and Wilson, 1970): 1) There is excellent agreement among students and between faculty and students, about the effectiveness of given teachers; 2) Best and worst teachers engage in the same professional activities and allocate their time among academic pursuits in about the same way; 3) Eighty-five items are listed that characterize best teachers as perceived by students, and 54 items are listed that characterize best teachers as perceived by colleagues.

McKeachie (1969) on the basis of his numerous comparative studies on teacher evaluation, writes that student ratings do have some validity. Teachers rated as effective by students tend to be those teachers whose students learn most. Student ratings are by far the most prevalent procedure now used in faculty evaluations. This was proved in the work done at the Center for Research and Development in Higher Education, University of California at Berkeley (1971). There is however new and contradictory evidence.

In a study done by Rodin and Rodin (1972) it was discovered that students rate most highly teachers who taught the least: The study was done with 639 physics students and several instructors. Pre- and post-measures were taken on actual knowledge of physics. The results were correlated with student ratings of these various instructors. The correlation between the ratings of the teachers by the students and the knowledge actually gained was  $-0.729$ .

In another study Naftulin, Ware, and Donnelly (1973) tested the hypothesis that student ratings of educators depend largely on personality variables and not educational content, the authors programmed an actor to teach charismatically and nonsubstantively on a topic about which he knew nothing. The authors hypothesized that given a sufficiently impressive lecture paradigm, even experienced educators participating in a new learning experience can be seduced into feeling satisfied that they have learned despite irrelevant, conflicting, and meaningless content conveyed by the lecturer. The hypothesis was supported when 55 subjects responded favorably at the significant level to an eight-item questionnaire concerning their attitudes toward the lecture. The study serves as an example to educators that their effectiveness must be evaluated beyond the satisfaction with which students view them and raises the possibility of training actors to give "legitimate" lectures as an innovative approach toward effective education. The authors conclude by emphasizing that student satisfaction with learning may represent

little more than the illusion of having learned.

Based on their paper dealing with "Some Problems and a Proposal on Teacher Evaluation," Anderson and Hunka (1963) found that a major step toward better understanding of the problems relating to teacher competency may be the intensive and extensive study of teacher characteristics.

Hildebrand and Wilson (1963) did a study at the University of California at Davis which later became part of the study on Evaluating University Teaching. Analysis of 85 items in this study, characterizing good teachers as perceived by students produced five components of effective performance: analytic/synthetic approach; organization/clarity; instructor/group interaction; instructor/individual interaction; and dynamism/enthusiasm.

A study by Phillips (1964), tested the hypothesis that teacher effectiveness is measured by what happens in class, and found that students favor a highly structured class with "highly visible" tests over a highly motivating class with a strong emphasis on personal warmth.

While it is clear that both teacher and method are important to the learning process, Fishman (1967) after reviewing research stated that "We cannot say just what it is that the effective teacher is or does."

Borg (1957), looked at the degree of relationship between three separate criteria for evaluating instructor effectiveness

(a student evaluation, peer rating, and a supervisor rating) and a number of personality and interest variables. This study was done at Lackland Air Force Base. Eighty-nine instructors were selected at random. They were rated on a five-point scale by students, then ratings were made by peer groups, and thirdly by supervisors. Tests were administered which measured personality, vocational interests, etc. The findings were that students down-rated instructors with authoritarian tendencies. He also reported that teachers who can promote an awakened interest and awareness tend to be highly rated by students.

In 1967 Mayhew showed that the four reasonable sources for evidence determining effective teaching are the teacher himself, the student, someone who has seen teaching in progress and demonstrations of behaviors which the teaching was intended to modify.

In what is probably the most extensive review of research on the effectiveness of teachers, Mitzel (1963) arrived at four distinctive categories of evaluating effective teaching: 1) Teacher attributes such as attitudes, interests and abilities. 2) Classroom behaviors which include techniques, teacher-student interaction, etc. 3) Changes in pupil behavior or pupil growth in subject matter knowledge, in social skills, in appreciation of democratic values, attitudes or appreciations. 4) Contingency factors such as school location, school size, school organization and community economic factors.

Lundstedt (1963) in his study on "Criteria for Effective Teaching" says that teaching is basically communication. "The effective communicator is generally the effective teacher."

Gage (1963) gives some idea of the many attempts made in an effort to better understand teaching. He states "in the half century since research on teaching began, thousands of studies have been made." Yet with all these studies available there is very little clear cut, uncontested information regarding teaching. One major reason for these contradictory and conflicting results may be that most of the studies done on effective teaching are done on a theoretical basis, i.e., survey questionnaires. This results in the emergence of a "model teacher" and "ideal learning climates" which appear valid statistically but are not represented in the real world.

The dissertation which is described on the following pages presents an alternative approach to effective teaching research. This approach was used in an effort to develop a scale which would be effective as a descriptive and as an evaluative instrument. This illustration provides a good example of the methodology and resulted in some information regarding teachers which will be published in a separate study. The methodology then is described in Chapter III.

## CHAPTER III

### METHODOLOGICAL DESCRIPTIONS

#### Development of Scale Items

In order to facilitate understanding of the methodology used in this study and how it differs from scales developed by traditional methods, this chapter will describe the methodology of a widely used traditional method of scale development and then describe the direct observer methodology used by the investigator in this study.

#### Traditional Methods

Hildebrand, et al. (1971) have discussed traditional, large scale research methods and their applications to the development of teacher rating scales. More specifically, they also discuss the methods employed in the development of the Berkeley Scale which is used as the criterion for the validation of the dissertation scale.

Many researchers (Coffman, 1954; Crannel, 1953; Solomon, 1966; and Wherry, 1950) have identified basic components, dimensions, or scales of effective teaching by sorting individual items describing aspects of effective teaching into related groups. Teacher-rating forms developed by students commonly do the same. Scales have been determined by subjective examination of a list of items, or by factor analysis (which

mathematically establishes the tendency of responses to the various items to associate in clusters). The number of scales developed in these studies ranges from four to thirteen, with four or five particular scales (knowledge, presentation, relation with students, enthusiasm) appearing rather consistently, even though the terminology differs.

In the study done at the Center for Research and Development in Higher Education, Berkeley, 1971, scales characterizing effective teaching as perceived by students were established by factor analyzing 91 items describing the teaching of 338 teachers identified as best by respondents to an earlier survey. Items were eliminated from the original list of 158 if: 1) they did not discriminate between best and worst teachers at the .001 level; 2) 25 percent or more of the respondents could not reply yes or no to them; 3) they were descriptive of virtually all best teachers, of few best or worst teachers, or of most best and worst teachers; or 4) if they were applicable only to small classes, or related to examinations and assignments.

But this merely hints at the abstract and complicated nature of factor analysis. The following discussion further indicates something of the extent of data analysis utilized in the development of the Berkeley Scale.

The method used was a principal-components analysis with a varimax rotation. After several analyses, a five-factor solution was selected as giving the maximum number of distinct and interpretable components of effective teaching. Items which had greater than .40



correlation were retained and analyzed further by pre-set cluster analysis to determine the consistency and reliability of the scales and their intercorrelations. The items were then re-analyzed with data from a validation survey. An alpha reliability was carried out on the data to determine the internal reliability of the scale.

For clarification purposes a description of factor analysis follows. This is the method commonly used for determining the number and nature of the underlying variables among large numbers of measures. Factor analysis tells us what tests or measures can be added and studied together rather than separately. The actual process is as follows: The data are correlation scores between a variety of tests, which are usually designed to measure a variety of processes. Each of a large number of individuals receives a score from each test. These scores are then intercorrelated yielding a correlation matrix. Factor analysis uses mathematical methods to compute the correlation of each of the tests with a few factors. Correlations between test scores and factors are known as factor loadings. Once the factors are determined that account for most of the intercorrelations, the factors can be interpreted by studying the content of the tests most highly weighted on each factor. The factor analysis itself is strictly a mathematical process. The naming of the factors, however, is a subjective process by the investigator who notes the tests that are in each factor and arbitrarily assigns a name to each of the factors.

A factor is a construct, a hypothetical entity that is assumed to underlie tests and test performance. It is extremely important to be aware of the fact that we can get nothing more out of a factor analysis than is put into it. We must also remember that we are not always aware of everything that is put into it. Any analysis, especially factor analysis, cannot produce factors that were not in the original items being analyzed. The human element is the crucial element no matter how abstract the statistical procedures become. Bad data well analyzed does not give good information. The data and the analysis of the data are no better than the original items from which they come (Kerlinger, 1964; Bales, 1950).

#### Dissertation Research Paradigm

The method used in this dissertation was quite different from that described above, via direct observation of a teacher while he taught and students as they learned. The direct observational model employed in this research began with the empirical evidence (direct observation) and moved through the development of a rationale, i.e., classified and named the observed phenomena, and then validated the observational data by using a student questionnaire.

It is important to note that the first and perhaps the most important consideration in any observation system is to know clearly what is being observed (Kerlinger, 1964; Bales, 1950). For this dissertation it was an actual classroom which constituted the total

universe of possible observational data. Within this exhaustive universe it was essential to have an observational unit. For this study the observational unit for sampling was delineated by the use of a behavioral cue.

A behavioral cue is a non-verbal message which transmits a powerful message. Behavioral cues are defined as things such as touch, facial expressions, body movement, speech patterns and change in speech patterns, posture, laughter, etc.

Since it was apparent that all the behaviors of the teacher and the students were too large and too vague a task, the situation was broken down into sub-sets. This was done by noting transactions that were preceded, accompanied, and/or followed by some behavioral cue. These were the sub-sets that became the raw data for this work.

This manner of sampling is known as event sampling and was chosen because of its appropriateness in field studies (Kerlinger, 1964; Guilford, 1956). Event sampling has several advantages, for example: the events are natural life-like situations and thus possess an inherent validity not possible by other kinds of sampling procedures; an integral event possesses a continuity of behavior that the more piecemeal behavioral sets of sampling such as time sampling (the selection of behavioral units for observation at different points in time) do not possess; and lastly many important events are important but infrequent. To illustrate, one may be interested in the ultimate steps in problem solving, or teacher's disciplinary methods (which may be a variable in

problem solving). As such, they can be easily missed by time sampling, they therefore require event sampling.

The actual process was as follows:

1. The investigator selected a class taught by a teacher reputed to be excellent. She secured permission from both the teacher and the participants in the class to observe them. The observational period consisted of 24 one-hour sessions. This observation period was judged to be a period long enough for pertinent events to be noted. The observer attended all 24 class sessions. For the first three days she took notes of all interactions occurring which fulfilled the criterion of the behavioral act (a behavioral cue was part of the interaction). This was done to see if the data needed to do the study were apparent and could be clearly specified. During all observational sessions the investigator made audio tape recordings of the entire period and made written notations describing the behavioral and contextual aspects of the cued interactions. For each isolated interaction, the footage on the tape recorder was noted. Later the same day the observer combined the verbal interaction and the behavioral notes, and recorded them in written form for consideration in the final scale item data. For this dissertation a transaction was considered to be any interaction between teacher and student(s) which contained a behavioral cue. It began with noun-verb (subject-object) problem statement and ended when verbal closure was reached

by mutual consent between teacher and students. Usually the teacher would say something such as: "Is everyone in agreement and ready to go on?" After three observational sessions the data were carefully examined for recurring patterns.

With careful examination of the data (three observational sessions) four sub-sets appeared to emerge. These were loosely labeled by the observer as: 1) positive learning climate, 2) teacher attributes and behaviors, 3) knowledge (techniques), and 4) interaction. It seemed to the observer that all the data (events) being gathered could be subsumed in one of these classes of behaviors.

The next phase of the study, the practical and fundamental job of assigning behaviors to categories, required inference on the part of the observer. Researchers who work with observational data are always required to use some degree of inference. Kerlinger (1964) says "observational systems differ on the amount of inference required by the observer." Systems with higher degrees of inference required of the observer are most commonly used and are probably more useful in most research, especially in educational research, because more information can be utilized. The high inference observation system used in this dissertation was considered to be within the training, educational, and experiential scope of the observer. The observer was trained in clinical psychology. She has the observational skills necessary to do clinical work, that is, extensive training and experience in looking at, listening to, and reporting observations in a systematic way. This

is, of course, the major task of a psychological clinician.

Data considered for inclusion in the final scale items were those interactions which had occurred at least three times, were not already represented in a category of the collected data, and were judged by the observer to be significant to the situation being observed.

One observed behavior (risk) needs special mention since it is not normally mentioned in the literature which covers teaching and the evaluation of teachers. Risk taking, which could be considered a sub-set of personal characteristics of the teacher, was apparent at the outset of the study and was judged by the observer to be one of the critical variables in the effectiveness of this teacher.

The observer developed a "fairly detailed but not too detailed definitions of occurring behaviors" (Kerlinger, 1964). She recorded some examples of the behaviors she observed in extensive, detailed, narrative form (verbatim from the tapes and notes). These are included as Appendix A. With these data available there should have been good correspondence, presumably, between what was measured and what was intended to be measured. Given these definitions and descriptions other investigators should be able to apply the terms in a similar way to the same or similar types of behaviors. In his book on Behavioral Research, Kerlinger (1964) states that reporting data in this manner renders the observational schema replicable and thus acceptable for research purposes.

An example of the behaviors and the step wise progression of the study is contained in Table I.

In summary, the investigator sat in the classroom and observed teacher-pupil and pupil-pupil interactions. She noted and recorded specific behaviors according to the research design which utilized behavioral cues to determine the observational event; she noted and recorded in written form specific behaviors judged by her as important to the situation being observed; she formed these into sub-sets of the actual situation. She began with a defined observational unit (or event), and she considered behaviors which occurred and recorded a variety of like behaviors under one rubric or category. The sub-categories (or sub-sets) were by necessity inferred by the observer. A machine cannot infer. If it could, it would be the preferred observer for research. These sub-sets were then further inferred by the observer as being subsumable under more general words. She chose four global terms (umbrella words) to label the behaviors she saw. These words were: interaction, personality, techniques, risk-taking. These words were too general to be useful as scale items. Thus, the sub-categories became the material for the final scale items (Table I).

One additional item was added. This was the item which noted that the teacher seemed to fit naturally into the role of teacher. This information was obtained from personal experience and the personal opinion of the observer. This information was not easily specifiable from the observational data--rather it came from behaviors of the

TABLE I

PROGRESSION FROM RAW DATA TO MAJOR CATEGORIES

Illustrative Statements or Behavior	Sub-Categories	Major Categories
<p>A. Problem solving; psychodrama (<u>doubling</u>): good natured-feeling level exchanges: confronts, uses problem situations for learning; gives adequate teaching responses to class fluidity (changing enrollment) employes vertical learning (able to teach students with different levels of competence in the same class); shares his philosophical views.</p>	<p>A. Creates positive learning climate</p>	<p>Interaction (between teacher and students)</p>
<p>B. Eclectic; uses all schools of psychiatric thought with facility; cites current research; demonstrates interviews; admitted to loving teaching in an interview; teaches from choice; acts happy; believed patient when staff disbelieved; discussed dilemma of psychiatric who is subpoenaed; sees everyone on a first come basis; shows anger, hostility, frustration, joy, etc.</p>	<p>B. Shows extensive preparation in his field; Uses correct language; euphonious; enjoys teaching; trust himself and others; believes in democracy; is egalitarian; gives appropriate affect.</p>	<p>Personal characteristics of teacher</p>



TABLE I--Continued

Illustrative Statements or Behavior	Sub-Categories	Major Categories
<p>C. Uses modelling; interviews, real life situations; applies semantic principles to behavioral description (never labels); insures understanding through paraphrasing, summarizing; repetition (saying it 3 times); and other emphatic devices; uses positive and negative reinforcement (verbal-non-verbal); asks for closure; explores alternative hypothesis and then shapes solutions accordingly; states that learning is idiosyncratic; lets people do their own thing; keeps discussions up-to-date through timely in-class interviews with students; observes and comments on student behaviors.</p>	<p>C. Demonstrates knowledge, reaches closure; communicates clearly; gives immediate feedback; uses divergent approach to learning; determines learners' level of competence; reinforces learning; experiential learning.</p>	Techniques
<p>D. Admits when wrong; can take criticism-- invites criticism; uses problems for learning as they arise; allows students to deal with real patients in second year of training; takes responsibility for student mistakes while guarding against their repetition and helps students learn from the mistake.</p>	<p>D. Not motivated by self-esteem; Lets students make own mistakes; Protects students as they learn; Uses self-correcting devices; Demonstrates empathy and support.</p>	Risk-taking

teacher such as teaching when he was not required to do so (he was department head of a major division of a teaching hospital), by the statement of the teacher that he "really liked to teach," and from the general feeling of enjoyment which he seemed to bring to the classroom on most occasions.

There were two additional steps in the development of scale items:

1. The investigator developed a questionnaire format which was designed to see how congruent her observations and perceptions of the situation were with other learners in the class. This preliminary questionnaire was administered to the same students who were in the classroom that was observed. It included both numerical and descriptive information. This questionnaire is included as Appendix B.

2. Further corroborative information was elicited by way of a second and different questionnaire which was administered to a different class, but taught by the same teacher. This questionnaire asked for information regarding the effectiveness of the teacher and the general learning climate created by the teacher. This questionnaire is included as Appendix C.

All information collected in the steps described above was utilized in development of the scale items.

Table II presents in tabular form the method used by the investigator as compared to traditional methodology (factor analysis) used to develop teacher evaluation scales.

TABLE II

COMPARISON OF TRADITIONAL METHODOLOGY AND TRAINED OBSERVER  
method of teacher evaluations

	FACTOR ANALYZED SCALE	TRAINED OBSERVER SCALE
Origin of Scale Items	<u>Rational.</u> Statements which seem pertinent to quality of teaching (i.e., they have face validity) collected from various sources or made up.	<u>Empirical.</u> A trained observer observed actual class of criterion teacher and analyzed incidents of student-teacher interaction.
Creation of Items To Go on Final Instrument	<u>Empirical.</u> Students were given whole pool of items to describe teachers. Results factor analyzed to produce "main factors" which become the final scale items (i.e., a summary statement made to include the <u>content</u> of those original items which define each factor--that summary statement is the scale item on final instrument).	<u>Rational.</u> The observer collects incidents into classes of related incident or sub-categories. Each class is <u>named</u> according to content of the group of sub-categories. The <u>name</u> is the "major category." The items on the final instrument are summary statements which include the <u>content</u> of the sub-categories.

As Table II demonstrates the traditional method requires a rationale, then proceeds to the collection of empirical data. On the other hand, the research paradigm proceeds in the opposite direction; that is, the collection of empirical data which is later supported by rationale.

## CHAPTER IV

### METHODOLOGY AND RESULTS

#### Empirical Testing of Scale Items

##### Introduction

The results of the data analysis of the preliminary scale items (Appendix D) lent support to the general categories developed by the investigator. The results of any ex post facto investigation (such as this study) must always be cautiously and carefully considered. Perhaps a good rule to follow is to be wary of any ex post facto research that does not test at least one hypothesis. In this research, it was hypothesized that an instrument developed by the non-traditional method of direct observation would be at least as accurate as an instrument developed in a traditional way, and would be more useful in several ways. The increased usefulness would lie in the descriptive information about the context of a given situation. There is reason to believe that context is nearly as an important and powerful determinant of behavior as is the content of the situation (Barker, 1967). Much information can be gathered during the development of the scale items. This is available as feedback material to anyone and everyone involved in the situation being observed and measured.

A scale developed in this manner does raise the question of validation. A generally accepted method is to validate the newly developed scale against another, more widely replicated one. For example, the well-known Stanford-Binet Intelligence Test was cross-validated with the Wechsler Intelligence Scale. This technique was used to validate this investigator's scale in that the scale that emerged from the direct observer data was validated against the more traditionally developed scale discussed in Chapter III (Berkeley Scale). In order to accomplish this, the investigator's scale and the Berkeley Scale were administered as indicated in the following discussion.

### Subjects

Ratings were taken on thirty (30) teachers in a state university. These teachers were from various departments. The departments represented were: Psychology, Economics, Sociology, Urban Studies, Anthropology, Physics, Mathematics, Foreign Languages, English, Art, Political Science, Business Administration, History, Geography, Music and Philosophy.

From these thirty sampled classes eight hundred and forty-eight (848) students answered the rating scale questionnaire and returned it. The sizes of the classes ranged from 6 to 150 students.

### Procedure

The investigator prepared packets of thirty questionnaires

which contained fifteen questionnaires developed by the people at Berkeley and fifteen questionnaires developed by the investigator. Since there was no way to be sure of how many students would be in each class, thirty was chosen as the reasonable number between a large and a small class. There were additional questionnaires available if needed. These too were alternately arranged for ease of distribution. The alternate arrangement of the scales insured that half the class would answer the traditional criterion scale and half would answer the investigator's scale. All questionnaires were anonymous. There were no identifying marks or codes on any of the questionnaires or packet containers.

Examples of the two questionnaires are contained in Tables III and IV.

The classes sampled were selected in the following way: the investigator and a fellow graduate student went to various departments at varying times of the day. The teachers who were approached and asked to participate in this study were those who happened to be present at these times. The investigator introduced herself and explained that she was doing research for a doctoral dissertation. The dissertation was explained. The person was then asked if he/she would participate in the final part of this dissertation by 1) answering a questionnaire,<sup>1</sup> and by 2) administering the questionnaires

<sup>1</sup>The results of the questionnaire prepared for teachers are not included in this study. They will be presented as a separate study. A copy of this questionnaire data is included as Appendix E.

TABLE III

INVESTIGATOR'S SCALE

STUDENT DESCRIPTION OF TEACHERS

1. The following items reflect some of the ways teachers can be described in and out of the classroom. For the instructor of this class, please circle the number which indicates the degree to which you feel each item is descriptive of him or her. In some cases, the statement may not apply to this individual. In these cases, check Does not apply or don't know for that item.

	Not at all Descrip- tive	1	2	3	4	5	6	7	Doesn't apply or don't know
1. The teacher is able to explain concepts in a way I can understand. He can demonstrate these concepts. He relates this subject to other areas of knowledge. He has a good knowledge of his subject.	1	2	3	4	5	6	7	( )	
2. The teacher makes clear what I'm expected to learn. The teacher is aware of what stage I am at in the learning process. He gives me prompt feedback. He helps me move to the next step in the learning process in a way that makes good sense.	1	2	3	4	5	6	7	( )	
3. The teacher allows me to make a try at the material to be learned with a minimum fear of penalty for making an error. If while learning I should make a mistake, I feel the teacher would support me and help me learn from the mistake.	1	2	3	4	5	6	7	( )	



TABLE III--Continued

	Not at all Descrip- tive			Very Descrip- tive		Doesn't apply or don't know			
	1	2	3	4	5	6	7	( )	
4. The teacher take some personal risk in allowing me to make mistakes. He involves himself, his knowledge, his feelings in the learning process.	1	2	3	4	5	6	7	( )	
5. The teacher deals honestly with me and with what is taking place at the moment in the classroom. He seems not to care how I learn the material but rather that I learn the material.	1	2	3	4	5	6	7	( )	
6. The teacher fits naturally into the teacher role.	1	2	3	4	5	6	7	( )	
7. Rate the overall effectiveness of this teacher for you.	1	2	3	4	5	6	7	( )	

TABLE IV

## BERKELEY SCALE

---



---

 Student Description of Teachers
 

---

I. The following items reflect some of the ways teachers can be described in and out of the classroom. For the instructor of this class, please circle the number which indicates the degree to which you feel each item is descriptive of him or her. In some cases, the statement may not apply to this individual. In these cases, check Does not apply or don't know for that item.

	Not at all Descrip- tive	1	2	3	4	5	Very Descrip- tive	6	7	Doesn't apply or don't know
1. Has command of the subject, presents material in an analytic way, contrasts various points of view, discusses current developments, and relates topics to other areas of knowledge.	1	2	3	4	5	6	7	( )		
2. Makes himself clear, states objectives summarizes major points, presents material in an organized manner, and provides emphasis.	1	2	3	4	5	6	7	( )		
3. Is sensitive to the response of the class, encourages student participation, and welcomes questions and discussion.	1	2	3	4	5	6	7	( )		

TABLE IV--Continued

	Not at all Descrip- tive			Very Descrip- tive			Doesn't apply or don't know	
4. Is available to and friendly towards students, is interested in students as individuals, is himself respected as a person, and is valued for advice not directly related to the course.	1	2	3	4	5	6	7	( )
5. Enjoys teaching, is enthusiastic about his subject, makes the course exciting, and has self-confidence.	1	2	3	4	5	6	7	( )
6. Rate your overall effectiveness of this teacher for you.	1	2	3	4	5	6	7	( )

to one of his/her classes. The questionnaires were displayed and discussed. The teachers were told that each student was to receive only one questionnaire and that these questionnaires were already alternately arranged for ease of distribution.

The teachers were asked to fill out the questionnaire designed for teachers at the same time the students were answering the questionnaire designed for students. They were told that if they wanted feedback on the results of this rating they could mark the envelope their questionnaires (teacher and students) were contained in so that they could be retrieved for special analysis and feedback

purposes. If the teacher agreed to participate, he was given a packet, and additional questionnaires if he needed them. Nearly all teachers approached accepted the packet and indicated their intent to carry out the procedure.

The packets were gathered by the investigator from the departmental secretaries a week later. These questionnaires were administered to classes the next to last week of the term.

Thirty-four packets were given out. Thirty were answered and returned.

### Results

A total of 848 questionnaires were included in the final analysis. Of these, 432 were the Berkeley Scale and 416 were the investigator's scale.

Four means were calculated for each class: 1) the overall mean rating of each teacher on the Berkeley Scale (items 6), 2) the overall mean rating on each teacher on the investigator's scale (items 7), 3) the mean for the other 4 items on the Berkeley Scale and 4) the mean for the other 4 items on the investigator's scale. For final data analysis Item 5 on the investigator's scale was discarded because the correlation of this item with the overall teacher ratings done in the pilot testing of the scale was not significant. Item 5 was not discarded earlier because of investigator's interest in the item. Six correlations were done and are included as Table V.

TABLE V  
CORRELATION AMONG MEANS

ITEM	X	Y	R	DF
1. Davis overall mean with investigator's overall mean	1	2	.79745	58
2. Davis overall mean with the Davis items mean	1	3	.89191	58
3. The Davis overall mean with the investigator's items mean	1	4	.73608	58
4. The investigator's overall mean with the Davis items mean	2	3	.64870	58
5. The investigator's overall mean with the investigator's items mean	2	4	.89358	58

Means

1. = Berkeley overall mean of effective teaching
2. = Investigator's overall mean of effective teaching
3. = Berkeley mean of scale items
4. = Investigator's mean of scale items

The six pairs of means correlated in Table V were: 1) the overall ratings of the teachers on both scales (means 1 and 2), 2)

the overall ratings of the teachers on the Berkeley Scale with means of the 4 items on the Berkeley Scale (means 1 and 3), 3) the overall rating of the teachers on the Berkeley Scale with the 4 items on the investigator's scale (means 1 and 4), 4) the overall ratings of the investigator's 4 scale items with the 4 scale items on the Berkeley Scale (means 2 and 3), 5) the overall ratings of the teachers on the investigator's scale with the 4 items on the investigator's scale (means 2 and 4), and 6) the means of the 4 items on the Berkeley Scale with the means of the 4 items on the investigator's scale (means 3 and 4).

A Kuder-Richardson test for estimating the internal reliability of the investigator's scale was performed. The result was an  $r$  of .79. The internal reliability of the Berkeley Scale ranged from .81 to .89. Because both the scales are somewhat unreliable a correction for attenuation was done in order to better assess how closely the accurate part of the Berkeley Scale correlated with the accurate part of the investigator's scale. This analysis resulted in a correlation of 1.01. It is not unusual for this analysis to over correct. This simply means that some of the underlying assumptions for the analysis were not met. In this analysis the assumptions are not critical to the results being valid.

## CHAPTER V

### DISCUSSION

This study was based on the idea that a trained observer, paying attention to behavioral cues, and studying one good teacher, could develop a rating scale. It was hypothesized that one good example of a teacher while he taught, and students while they learned could supply enough information to develop scale items. Further, these scale items could be formed into a teacher evaluation scale which would be as good as a teacher rating scale developed by traditional methods, i.e., factor analysis.

The results are in accordance with predictions made in the hypothesis. While much of the variance of effective teaching is accounted for (.79 on the observer's scale and .81-.89 on the Berkeley scale) there is still variance unaccounted for. The coefficient of correlation between the two scales, after being corrected for attenuation, approached a perfect correlation.

As both scales are somewhat unreliable (scales reported in the literature are also somewhat unreliable), it may well be that it is not now possible, and will never be possible, to totally isolate all the variables operating in effective teaching. Some variables such as knowledge, etc. hold constant from teaching situation to teaching

situation, and appear on most evaluation scales (including the two scales discussed in this study). However, it appears that there are going to be unique variables operating in each teacher-student situation that cannot be captured in the form of constant scale items. It is apparent that a single hypothetical model of an effective teacher is not to be found in the real world of teaching and teachers.

As was noted earlier, it is a generally accepted practice in psychology to validate one scale against another. Since the dissertation scale has an extremely high correlation with the criterion scale, the acceptability of the procedure used to develop the scale items in the observer's scale is supported.

Some discussion of the problems and strengths of the observational method seems in order. The problems reside in three major areas:

a) availability of observational model, b) the availability of trained observers, c) the difficulty and importance of selecting an adequate observational unit or act. A discussion of these problems follows:

#### Availability of Observational Model

This study was possible because the investigator was able to secure permission from both teacher and students to observe, in great depth and for an extensive period, the teacher as he taught and students as they learned. The observer approached three teachers before finding one who was reputed to be excellent and would consent to being observed in the manner dictated by the research design. This reluctance on the



part of teachers to be observed is well known (Midley and Mitzel, 1963). It would, no doubt, be difficult and challenging to submit oneself to close scrutiny and hence evaluation. As educators become more familiar with evaluation and more comfortable with the idea of evaluation, hopefully they will view direct observation with less resistance than they do now. Part of the process of educating teachers to the usefulness of direct observation and making them more easy with the process of evaluation would seem to include studies such as this one. Since this study is explicit as to what direct observation entails, perhaps, teachers will better understand - and hence be more willing to serve as observational models. Perhaps, they can ultimately be convinced that evaluation is a productive rather than a destructive force in improving education. This circumstance, the reluctance of teachers to be observed, may be the biggest problem encountered in attempting to use the method described in the dissertation.

#### Availability of Trained Observers

In virtually all studies involving direct observation Kerlinger's statement with regard to observers has been taken as a categorical imperative --- "The major problem of behavioral observation is the observer himself. . . In behavioral observational techniques the observer is both the crucial strength and the crucial weakness. Why? The observer must digest the information derived from his observations and then make inferences from them about constructs. . . (pp 505-507)."

It stands to reason, and is logical that a trained observer can develop a more accurate instrument than a non-trained observer can.

With the exception of a totally exhaustive observational system (all behaviors recorded without classification), all observational systems require some inference on the part of the observer. Since recording all behaviors is a vague and difficult task, it is seldom used. (Kerlinger, 1964). A major limitation of observational systems which require some inference on the part of the observer seems to be the lack of trained observers. (Madron, 1969, Kerlinger, 1964). However, it seems to this researcher that while the lack of trained observers is indeed a problem, it can be remedied. Useful models for this kind of training (observation of behaviors) are readily available, can be used with any size group, and can be done in an expeditious manner. (Feinstein, 1970, Engel, 1972, Saslow, 1974).

The method used in the study being described required a good deal of inference on the part of the observer. The reason for choosing this system and the observer's capabilities for making reliable inferences were discussed earlier in the paper.

#### Behavioral Unit or Act of the Observational System

Much of the reliability of the coding of the system depends on the decision as to what constitutes a unit of behavior. In this study, the unit of behavior was determined by the use of a behavioral cue. The statistical results of this study indicate that this method of sampling

is a good one. As noted earlier, the behavioral cue is not usually used in sampling but is used extensively in clinical work. The use of behavioral cues requires some training - but as noted above - this training can be easily obtained.

#### Strengths of this Method

The anticipated problems which usually accompany field studies, i.e., cost, time, availability of observer, feasibility, etc., were not encountered in this study. Not only was it a very reasonable task, but the method used in this study yielded a great deal of information for relatively reasonable inputs of time, energy, and other resources. An example of some of the information obtained by this study is the narrative data contained in Appendix A. While narrative data does not lend itself easily to quantification, nevertheless, it is important in terms of making alternative and additional kinds of information about human behavior available.

Additional research possibilities became evident to the observer during the observational period. Subtle variables were noted that seemed to be important in the situation. Some of these observations have already been used as pilot work for further research into the area of teaching effectiveness; for example, personality variables of teachers is now being researched in another study by the same researcher. In line with many researchers (Hutt and Hutt, 1970, Brandt, 1972), this researcher advocates intensive study of specific

aspects of behavior with joint field and laboratory experiments to clarify more complex relationships. These should be complimentary methods for either the development or confirmation of hypotheses. They also provide the basis for a research cycle which allows the investigator to begin with a natural setting and end with a series of new questions about the behaviors in the same situation.

An especially interesting aspect of this work is the speculative model of an effective teacher that emerged. This model describes the character of the teacher, his behavior, his personality, and the learning climate he created. This information is included as Appendix F.

#### Implications for Research Areas Outside Education

Recently there has been increasing emphasis upon the utilization of behavioral science concepts and methods for meeting social problems. Every day there are new programs in education, public health, medical care, urban redevelopment, and social welfare. Because of the enormous amount of public monies spent on these kinds of programs, people are wanting to know if this money is being well spent and if the programs they are paying for are effective. This means that these programs must be evaluated. People resist evaluation unless they can be assured of a positive outcome. This is not difficult to understand because people have vested interests and are reluctant to have their programs and the values underlying these programs questioned or judged. However, many Federal Programs cannot be funded without a research

component being included in the proposal. This has become the norm in Federal funding.

The method used for this study could be especially useful for evaluative research because of the feedback aspect of the observational procedure. Much information, besides statistical measures, can be made available to agencies or programs that might like to use this method for evaluative purposes. Gage, et al, 1963, support the idea that feedback improves behavior considerably.

#### Summary and Implications

Briefly stated, the hypothesis of this study was that a trained observer and one example of effective teaching could provide enough useful information for the generation of scale items which would result in an evaluation scale as good as the traditional factor analysis method.

All in all, the results have the following theoretical and practical implications: the direct observation method was given support by the high correlations between the observer's scale and a traditional scale; the scale developed in this dissertation was not only as good an instrument as the criterion scale, but it supplied additional and useful information.

For practical purposes, our results suggest that this method can be used to develop teacher evaluation scales; and it is speculated that it would be useful in other social areas as well. This is a question for further investigation.

Finally, a couple of notes to researchers are in order. First, the use of the behavioral cue was an effective and an interesting way to sample social-interaction data. The observer not only collected data, but learned something about the subject being dealt with in the observational situation.

Second, if this method is to be used, it is important that the ground be prepared for observation and evaluation. This can be done by careful education and consideration of the people involved in the situation being studied.

## References

- Ackerman, W. I. 1956. "Presentation of Alternatives and its Relation to Set in Problem Solving," Unpublished doctoral dissertation, Harvard University.
- Anderson, C. C., and Hunka, S. M. 1963. "Teacher Evaluation: Some Problems and a Proposal." Harvard Educational Review, 33, 74-95.
- Anderson, H. H., and Brewer, J. E. 1946. "Studies of Teachers' Classroom Personalities. II. Effects of Teachers' Dominative and Integrative Contacts on Children's Classroom Behavior." Applied Psychology Monograph, No. 8(a).
- Anikeeff, Alexis M. 1953. "Factors Affecting Student Evaluation of College Faculty Members," Journal Applied Psychology, 37, 458-460.
- Asch, M. J. 1951. "Nondirective Teaching in Psychology: An Experimental Study," Psychology Monograph, 65, No. 4 (Whole No. 321).
- Baloziak, L. W. 1954. "The Role of the Laboratory and Demonstration in College Physical Science in Achieving the Objectives of General Education." Dissertation Abstract, 14, 502-503.
- Bales, R. F. 1950. "A Set of Categories for the Analysis of Small Group Interaction," American Sociological Review, XV, pp 258.
- Barker, R. G. 1967. "Naturalistic Methods in Psychological Research," Human Development, 10: 223-229.
- Bane, C. L. 1931. The Lecture in College Teaching. Boston: Badger.
- Bills, R. E. 1952. "An Investigation of Student Centered Teaching," Journal of Educational Research, 46, 313-319.
- Borg, Walter R. 1957. "Personality and Interest Measures as Related to Criteria of Instructor Effectiveness," Journal of Educational Research, 50, 701-709.

- Brandt, R. M. 1972. Studying Behavior in Natural Settings, New York: Holt, Rinehart, and Winston.
- Carlson, C. R. 1953. "A Study of the Relative Effectiveness of Lecture and Directed Discussion Methods of Teaching Tests and Measurements to Prospective Air Force Instructors," Dissertation Abstracts, 13, 1112-1113.
- Coffman, W. E. 1954. "Determining Student's of Effective Teaching from their Ratings of Instructors," Journal of Educational Psychology, 45, 227-286.
- Cohen, H., et al. 1968. Training Professionals for the Establishment of Educational Environments, New York: Faculty Press, 248-254.
- Crannell, C. W. 1953. "A Preliminary Attempt to Identify the Factors in Student-Instructor Evaluation," Journal of Psychology, 36, 417-422.
- Flanders, N. A. 1951. "Personal-Social Anxiety as a Factor in Experimental Learning Situations," Journal of Educational Research, 45, 100-110.
- Fishburn, C. E. 1955. "Teacher Role Perception in the Secondary Schools of one Community," Dissertation Abstracts, 15, 1798-1799.
- Fishman, J. 1967. "Cross-Cultural Perspective on the Evaluation of Guided Behavioral Change," The Evaluation of Teaching, Washington, D.C.: Pi Lambda Theta, 9-13.
- Gage, N. L. 1963. "Paradigms for Research on Teaching," In N. L. Gage (Ed), Handbook of Research on Teaching, Chicago: Rand McNally, 418-428.
- Gage, N. L. 1963. "The Teacher and Characteristics," In N. L. Gage (Ed), Handbook of Research on Teaching, Chicago: Rand McNally, 506-582.
- Gage, N. L. 1961. "The Appraisal of College Teaching," Journal of Higher Education, 32, 17-22.
- Getzels, J. and Jackson, P. 1963. "Occupational Choice and Cognitive Functioning: Career Aspirations of Highly Intelligent and Highly Creative Adolescents," Journal of Abnormal and Social Psychology, 61, 254-277.
- Getzels J. and Guba E. 1954. "Role Conflict and Effectiveness," American Sociological Review, 19, 164-175.



- Graves, J. 1956. "Use of Tapes in Medical Teaching," P. G. Medical Journal: 16, 52.
- Greenhill, L. P. 1963. "Communication Research and the Teaching-Learning Process," Journal of Medical Education: 38, 495-502.
- Guilford, J. P. 1956. Fundamental Statistics in Psychology and Education, New York: McGraw Hill, 395.
- Hildebrand, M. and Wilson, R. C. 1970. Effective University Teaching and its Evaluation, Berkeley, Calif.: Center for Research and Development in Higher Education, 28 pp.
- Hildebrand, M. and Wilson, R. C. 1963. Evaluating University Teaching, Davis: Unpublished Manuscript, Center for Research and Development in Higher Education, 18 pp.
- Hutt, S. J. and Hutt, C. 1970. Direct Observation and Measurement of Behavior, Springfield, Ill.: C. H. Thomas.
- Kerlinger, F. N. 1964. Foundations of Behavioral Research, New York: Holt, Rinehart and Winston, Inc., pp 275-444.
- Lewis, E. C. 1964. "An Investigation of Student-Teacher Interaction as a Determiner of Effective Teaching," Journal of Education Research, 57, 360-363.
- Lundstedt, W. 1963. "Criteria for Effective Teaching," In N. L. Gage (Ed.), Handbook of Research on Teaching, Chicago: Rand McNally.
- Madron, T. W. 1969. Small Group Methods, Evanston, Ill.: Northwestern University Press.
- Maloney, R. M. 1956. "Group Learning Through Group Discussion," Journal of Social Psychology, 43, 3-9.
- Mayhew, L. B. 1967. "A Tissue Committee for Teachers," Improving College and University Teaching, 15, 5-10.
- McKeachie, W. J. 1965. "Student-centered vs Instructor-centered Instruction," Journal of Educational Psychology, 45, 143-150.
- McKeachie, W. J. 1963. "Research on Teaching at the College and University Level," In N. L. Gage (Ed.), Handbook of Research on Teaching, Chicago: Rand McNally and Co.: 1118-11172.

- McKeachie, W. L. 1969. "Student Ratings of Faculty," AAUP Bulletin, 55, 439-444.
- Medley, D. M. and Mitzel, H. E. 1963. "Measuring Classroom Behavior by Systematic Observation," in N. L. Gage (Ed.) Handbook of Research on Teaching, Rand McNally and Co. Chicago, 247-328.
- Miller, R. I. 1972. Evaluating Faculty Performance, Jossey-Bass, Inc.
- Naftulin, D., Ware, J. E., and Donnelly, F. A. 1973. "The Doctor Fox Lecture: A Paradigm of Educational Seduction," Journal of Medical Education, 48, 630-635.
- Patton, J. A. 1955. "A Study of the Effects of Student Acceptance of Responsibility and Motivation on Course Behavior," Dissertation Abstracts, 27, 637-638.
- Perkins, H. V. 1950. "The Effects of Climate and Curriculum on Group Learning," Journal of Educational Psychology, 41, 268-286.
- Phillips, B. N. 1964. "The Individual and the Classroom Group as Grades of Reference in Determining Teacher Effectiveness," Journal of Educational Research, 58, 128-131.
- Pogue, F. G., Jr. 1967. "Students' Ratings of the 'Ideal Teacher,'" Improving College and University Teaching, 133-136.
- Posthethwait, S. N. 1965. "Audio-Tutoring: A Practical Solution for Independent Study," Medical and Biological Journal. 1, 345-347.
- Rickard, P. B. 1946. "An Experimental Study of the Effectiveness of Group Discussion in the Teaching of Factual Content," North-western University Summaries of Doctoral Dissertations, 14, 72-77.
- Rodin, and Rodin, B. 1972. "Student Evaluations of Teachers," Science, 77, 1164-1166.
- Ryans, D. G. 1960. Characteristics of Teachers, Washington, D.C.: American Council on Education.
- Saslow, G. 1974. Practice of Selected Communication Skills, Mimeograph, Conjoint 413, University of Oregon Medical School.
- Saslow, G., and Grant, R. 1969. "Maximizing Responsible Decision Making or How Do We Get Out of Here?" Paper presented at Conference on New Hospital Psychiatry, Madison, Wisc.

- Solomon, D. 1966. "Teacher Behavior Dimension, Course Characteristics and Student Evaluations of Teachers," American Educational Journal, 3, 35-47.
- Sprights, E. 1967. "Students Appraise Teachers' Methods and Attitudes," Improving College and University Teaching, 15, 15-17.
- Stones, E. 1968. Learning and Teaching, London: Wiley and Sons.
- Suchman, E. A. 1967. Evaluation Research, New York: Russell Sage Foundation.
- Symonds, P. M. 1955. "Characteristics of the Effective Teacher Based on Pupil Evaluation," Journal of Experimental Education, 23, 289-310.
- Verplanck, W. 1955. "The Control of the Content of Conversation," Journal of Abnormal and Social Psychology, 51, 668-676.
- Wallen, E. W. and Travers, E. M. W. 1963. "Analysis and Investigation of Teaching Methods," In N. L. Gage (Ed.), Handbook of Research on Teaching, Chicago: Rand McNally and Co., 448-505.
- Ward, J. N. 1956. "Group vs. Lecture Demonstration Method in Physical Science Instruction for General Education College Students," Journal of Experimental Education, 24, 197-210.
- Wherry, R. J. 1950. Control of Bias in Rating: Instructor Rating Scales: Washington, D. C.: Personnel Research Section, AGO, U. S. Department of Army.
- White, J. R. 1945. "A Comparison of the Group-Laboratory and the Lecture Demonstration Methods in Engineering Instruction," Journal of Engineering Education, 36, 50-54.
- Wilson, L. 1943. The Academic Man, New York: Oxford University.
- Yamamoto, K., and Dizney, H. F. 1966. "Eight Professors," Journal of Educational Psychology, 57, 146-150.

## APPENDIX A

## NARRATIVE DATA

The variables which emerged from analysis of the tape recordings, observations by the experimenter of the classroom situation, and notes taken by the experimenter during conversations with the teacher being studied are described in the following pages.

1. Core of Learning Climate

The basic core of the learning climate seemed to be in the use made of innate human potential--both the teacher's and the student's. In an effort to clarify this idea, which seemed to the observer the critical variable in this situation, the following analogy seems appropriate. The concept of synergy, used in Chemistry, is defined as the combined and correlated action of two or more elements or constituents. This idea suggests that the total is greater than the sum of the parts. For example: the properties of chrome-nickel alloy are greater than the sum of the individual properties of chrome plus nickel (Cohen, et. al., 1968). The effect of human interaction cannot be predicted or controlled by the examination of discrete parts. Behavior must be considered in relation to the total environment.

The basic dynamics for this teacher's effectiveness came from the

kind of involvement of both the teacher and the students. For example: in one group learning session the members and the teacher were discussing a very difficult patient. One of the students said, "I'm so confused about this patient. First we are instructed to treat her one way, then we are to treat her another way." The teacher jumped up and ran behind the student's chair, and, using a technique known in psychodrama as doubling, said, "If you guys go on giving me these cockeyed contradictory instructions I'll be crazy too." The student in charge of the patient went on to try to explain his position further. The teacher then said, "Don't go too fast. You go too fast for me. Have you finished yet describing all your observations about her problem? If not, do it step by step." The student began again. Another member of the class went to double for the student and said, "They don't want to hear about what my feelings are about this patient, and I have very strong feelings about the patient and the position I'm put in as a result of having inherited this problem." The teacher then said with good nature, "At the moment I don't give a goddamn what his feelings are." The student then said, "And I don't give a goddamn what your feelings are either." There was much laughter and the teacher said, "Well, let's all go out and get some coffee." This was at the beginning of a very productive problem-solving session (Tape recording, March 30, 1969). These and other data led the experimenter to conclude that from this alloy of teacher, students, and problem, there emerged a spontaneous, supportive, viable, exciting aura which imparted to the classroom

situation immediacy, validity, and worthwhileness. Here is an educational climate which seems flexible and honest, where people can really put their knowledge (or search for knowledge) and their feelings into the moment. The teacher does the same. The stage is set for wide-open productive experiencing. The climate of excitement prevails even though the students range in training and experience from aides through residents, and even though the personnel was constantly changing. New people such as aides were entering the class whenever they became members of the staff. The minimum number of sessions attended by any one student was fifteen, the maximum was twenty-four. (Only people who had attended at least fifteen sessions were used in this study.)

## 2. Personality

The personality of the teacher emerges as a significant variable in the present study. In the first place he is extremely well-educated and erudite. He appears interested in a great many areas of human knowledge and experience. In one hour the experimenter noted seven references drawn from non-medical fields such as music, literature, politics, which were used to illustrate a point he was making (Tape, April 1, 1969). He has a thorough knowledge of theories and theorists in his own field. In trying to help students understand and work through problems, he uses any and all schools of psychiatric thought with facility. In one hour he went by such varied routes as cognitive

dissonance, oedipus complex, and adequate data base, (Festinger, Freud, Weed), to give students the best possible information concerning the problem they were discussing (Tape, April 1, 1969). He is, however, a task-oriented teacher and his aim is always to let the students learn how to learn for themselves. This teacher has amazing zest. He is extremely quick in all areas. He enjoys living and takes the task of living seriously. He gets off the elevator on the third floor rather than the second floor in order to get a better view of Mount Hood as he walks to his office. This involves walking three sets of stairs rather than two (Investigator's Notes). He seems to exalt and inspire the students. In a learning session on June 3, 1969, he and the students together invented a whole new treatment approach for a patient being discussed. None of the ideas can be found in books. His comment was, "So little is known that I am willing to experiment a bit--from things we already know to things we might know" (Investigator's Notes). His demeanor and attitudes are those of an extraordinarily decent human being. He has basic trust in, hope for, and concern about people. The students once expressed the feeling that they were being "garbaged" by an alcoholic patient who reported a week-end with a decreased use of alcohol and an increased attempt to a more meaningful kind of existence. The teacher said, ". . .You people are really deep-dyed skeptics. I don't know where you get the idea that people are malingering all the time. That's a strange notion to me. I must give you that thing I wrote up last year--in which I included the principle that each person does the best

he can. I think it's nothing remarkable if a person feeling better-- begins to act better. I think we can be optimistic, and quite genuinely so, even if it doesn't last. . .I think we must be careful so as not to make it even harder for a fragile person's determination to keep on working hard for another way of life" (Tape, April 21, 1969). This teacher is idealistic. He believes that the good fight, win or lose, is the best way to live. In one taped session the experimenter noted 10 expressions of idealism, such as: ". . .I don't expect you not to make mistakes. The only concern I would have is--would you be doing an injustice to the one person who is the most helpless in this whole totem pole, the patient. That's the one thing I would be very concerned about. This is pernicious in its influence. . ." (Tape, April 2, 1969) In a discussion concerning the predicament of psychiatrists who are subpoenaed into court as witnesses involving patients whom they have seen, the teacher repeatedly stressed the delicate balance between two responsibilities--protecting the patient, who has given them information in trust, and their responsibility to defend and protect our democratic way of life (Investigator's notes).

The emotional and physical involvement of teacher and students is very apparent. This teacher possesses unusual analytical skill. He brings to every situation a tremendous apperceptive mass. He listens carefully and seems to have the ability to sustain patience at any length, as long as he feels it is producing anything worthwhile or usable for anyone. He refuses to be involved in situations that he considers worthless. Once he said, "I haven't any fancy notions about



whom I give my time to, but I do have fancy notions about wasting my time" (Investigator's notes). It is impossible to cover adequately the exciting interchanges, the calm and methodical working through of problems, the fellowship and support, and the flavor and fun of the classes.

His speech pattern is rapid but precise. His language is euphonious and correct. He gives to his students a double message which is, "I do not have all the answers. I was not put on this earth to protect you from every unhappiness, nor to live your life for you. At the same time, I am willing to do whatever is in my power to enable you to live your life in the best possible way--to this end I will give of any strengths and knowledge I possess. I will support and protect your right to learn and to live, with any resources I have, even if it means significant risk on my part" (Investigator's Notes). When he tells students, "I back your right to learn", he means it (Tape, April 30, 1969). He expresses the idea that he regards as sacred both one's autonomy and one's right ultimately to decide what in life makes the best sense for him (Tape, April 1, 1969). He believes and exemplifies the attitude that "The dilemmas of living are opportunities for growth" (Grant and Saslow, 1969).

### 3. Techniques

The major teaching techniques emerging from analysis of the tapes appeared to be:

a. Affling: which means guided discovery plus fellowship (Gygi, C., 1967). In this teaching technique the ultimate answer is not presented immediately, but the present state is defined and then modified in such a way that there is shared understanding or closure on everyone's part. The problem or dilemma then is rephrased to present a question about which alternative hypotheses can be set up. The response is gradually shaped to approximate the ultimate requirements of the problem situation. In this method the student gets immediate reinforcement, so he knows on what points he is incorrect in the total context of the problem being studied. This teacher is extremely apt at setting up an educational environment in which the student seems willing to make a try at the information to be learned, without fear of making errors. If mistakes are made, they become grist for the intellectual mill, and become a vital part of the learning process. This teacher feels that the "Most important form of learning. . .is the examination of the consequences of any action or decision rather than the emphasis on the 'right decision' and an attempt to prevent 'mistakes'" (Saslow and Grant, 1969).

b. Repetition: The experimenter noticed upon analyzing the tapes that each idea brought up for discussion was repeated three times: (A) by the student initiator, (B) again by the teacher, who at that time added any additional information he might have, (C) in summary by teacher or by both teacher and students as all ideas are presented and closure reached. It is insidious and effective. The observer did not

become aware of this while sitting in the classroom--only when the analysis revealed it as a pattern. One session (April 1, 1969) was transcribed using only the teacher's remarks. Names of students presenting ideas, outside references bearing on the problem, and other pertinent information implying more than one person, were apparent in the summarizing statements made by the teacher.

#### 4. Flexible Language Use

The teacher's language is geared to an ongoing, dynamic and precise but flexible use of words, which always invites the people involved to add to what is being said. It is an effective technique for involving people. He works with the idea that time and movement are inexorable, and that in a fraction of a second the situation being dealt with and described has already begun to change. He believes that language influences behavior. A student who, in discussing a patient who was neurologically not blind but said that she could not see, kept referring to her "blindness." The teacher commented as follows:

". . . Suppose we throw out that word blind and you give me some behavior descriptions that I can understand. She is not blind; I happen to know that. So you can't use that word because it prevents us from seeing clearly what is going on. . . . So., as soon as she has somebody to lean on she becomes "blind." Can't you see the treachery of using such a word? Mustn't you train yourself not to use such words--She's not blind!" (Tape, April 29, 1969). Another student kept using the

word plateau to describe a patient. The teacher commented, "Plateau-- that word doesn't mean anything to me. A plateau can be a foot wide, a mile wide, 200 miles wide, I don't understand that word used in this context." This teacher feels, as do semanticists such as Korzybski, that a behavioral functional description is much more useful in terms of the ongoing nature of the universe and man, than are such labels as "A is A." These static labels are useless in his particular time and movement context. He repeatedly calls attention to words that are meaningless or misleading. He distrusts the use of global labels. He expresses the idea that certain questions are worthless (here he uses the only example of punishment that can be documented from the tapes, the refusal to answer some questions until they were rephrased) and unanswerable, and he simply refuses to waste his time on them. Instead of assertions such as, "It is known", he would be more likely to say something such as, "It appears that.", or "One might look at an idea such as", etc. If too many global words are used by a student, he plays dumb and invites the student to "tell me what that word means." Sometimes he really does not understand the words being used and again invites the person to "tell me more," until it is clear to him exactly what the implications of the word symbols are to that person. Except in the summary part of the hour, his use of the "to be" verb is almost absent. Then he resorts to a little more "hard-lining" in an effort to stamp in more effectively the ideas that have been discussed and clarified

during the class.

### 5. Reinforcement

His use of non-verbal reinforcement is very skillful. He uses a touch of the hand, a movement towards a person, or a smile as positive reinforcement, as well as verbal reinforcement. In discussing a 10-minute interview with a patient, a student commented: "We could have spent the whole hour on this one person, but I thought we got enough material in 10 minutes to work with." The teacher responded: "I don't think it was necessary to use more time, and I like the way you are trying to get people a little more willing to have short interviews." (Tape, April 16, 1969). His affects are immediate and appropriate. He can display anger, frustration, and hostility by bodily movements (non-verbal reinforcement). His manner of gesturing adds to his presentation of a point. Sometimes the gestures are tense and straight-lined, indicating seriousness and urgency. Often they seem round and slow, indicating an easy supportive fellowship. His hands are a graphic and delightful adjunct to his material and to the whole milieu of the class. The experimenter observed that many of the students imitated the gesturing pattern of the teacher.

### Divergent Approach:

All his teaching techniques are aimed toward a divergent rather than a convergent approach to problem solving. He constantly encourages

the students to think of many possible alternative solutions to the immediate problem, rather than directing them towards one and only one right answer. In one session he stated: "I have never known any one way that everybody learned anything." (Tape, April 2, 1969). An example of this teacher's style is as follows: Someone presents a problem and asks, "What do I do now?" or says, "I don't know what to do." The teacher then starts a process which is essentially: "Before we try to answer your question, let us ask ourselves the following questions: A. 'What are all the observations and information we now have about this problem?' B. 'What additional observations or data do we need to see this question more clearly in a broader context?' C. 'What's the context now?' D. 'How does the original question look at this point?' "

Furthermore, he models divergency by being always willing to look for new solutions for old problems. He doesn't care where the answers come from. If it makes sense, he uses it. This teacher works with the moment and signs his name to it as life spent. His whole approach to teaching is one of immediacy, validity and risk. He assumes no hard and fast rules for life, and will listen to and consider any variation on the theme of existence no matter how great a deviation from the norm it seems. Here, the observer feels, is a most important aspect of this teacher's effectiveness.

## APPENDIX B

## PRELIMINARY QUESTIONNAIRE 1

This preliminary questionnaire was administered to the same students who were in the classroom that had been observed. The questionnaire provided not only numerical data but verbal responses to open ended questions. Both of these sources of data supported the major categories initially developed and the variables isolated as being important in the observed learning situation. The comments (verbal data) were coded by the observer. The coding system was then checked by a fellow graduate student who randomly selected one-fourth of the 25 questionnaires and was able to locate their representation in the coding schema of the observer.

The questionnaire developed to corroborate the accuracy of the observational system of the observer, and to elicit information from other learners in the situation is as follows:

## QUESTIONNAIRE 1

N = 25

This is a study of the way a particular teacher affects those with whom he works. Your replies to this questionnaire will be part of the study. Analysis of tape recorded group teaching sessions with the teacher present or absent will be another part of the study. Your assistance with the questions below will be most helpful to me.

Thank you.

Name  
Graduate Student Psychology  
Portland State University

1. Rate Dr. X's class on its interest to you. (interaction)

1 2 3 4 5 M = 4.71  
least \_\_\_\_\_ most

2. Rate the effectiveness of his teaching. (Personality)

1 2 3 4 5 M = 4.80  
low \_\_\_\_\_ high

Comment:

3. Rate the degree to which Dr. X's ideas and values have helped you in re-examining your own life and your ideals. (Personality, Technique)

1 2 3 4 5 M = 3.96  
little \_\_\_\_\_ much

Comment:

4. Rate the degree to which you are influenced by Dr. X's to seek out non-medical sources of information regarding Psychiatry. (Techniques, Personality)

1 2 3 4 5 M = 3.97  
not at all \_\_\_\_\_ a great deal

5. Rate the degree to which you personally involve yourself in Dr. X's class. How? (Interaction)

1 2 3 4 5 M = 3.81  
not at all \_\_\_\_\_ a great deal



5. Comment:

6. Rate the degree to which you think the other participants involve themselves in Dr. X's class. (Interaction)

1 2 3 4 5 M = 3.80  
Not at all \_\_\_\_\_ a great deal

7. Rate the degree to which you feel you have attempted to influence other people (non-patients) to understand and employ ideas, values, and techniques as taught by Dr. X. (Personality)

1 2 3 4 5 M = 4.00  
Not at \_\_\_\_\_ a great deal  
all or \_\_\_\_\_ or often  
never

Comment:

8. Rate your effectiveness in influencing others with regard to Dr. X's ideals, values, and techniques. (Personality)

1 2 3 4 5 M = 3.50  
little \_\_\_\_\_ much

Comment:

9. Rate the degree to which Dr. X supports and protects your right to learn. (Risk)

1 2 3 4 5 M = 4.52  
little \_\_\_\_\_ a great deal

10. Is there an element of risk in his protection and support of your learning? How important is this risk in terms of accomplishing your goals? (Risk)

1 2 3 4 5 M = 4.00  
none \_\_\_\_\_ a lot

Comment:

11. What do you consider the most valuable payoff for having known Dr. X as a teacher.

Comment:

The experimenter and the learners were in agreement on a surprising number of ideas regarding this teacher. Some of the comments, even more than the statistical results, are interesting from the point of view of gathering information. A few comments regarding the questionnaire will help with further interpretation of the results. In the first place the questionnaire was a difficult one. It continually asked for illustrations to support the ratings given. Question Number 10 was the most difficult and only fifteen of the twenty-two returned forms contained answers and comments on this question. The unanswered forms said something to the effect - "I don't understand this question. It seems very vague and abstract to me." The people who understood the question were able to support very effectively the idea that the elements of support, protection and risk were important aspects in the effectiveness of this teacher. This idea will be explored later.

To support the idea that the critical variable for this teacher's effectiveness has something to do with emotional and behavioral involvement and his ability to transmit this to the learners, questions Number 1 (interest), Number 5 (personal involvement) and Number 6 (involvement of others) were formulated. The mean score on Number 1 was 4.71. The mean score on Number 5 was 3.81. The mean score on Number 6 was 3.80.

Some of the comments were as follows: "On passive involvement (non-verbal) I would rate a 5 and on active involvement (verbal) I

would rate 4." "Part of the agenda for my behavioral change. . .includes not participating overactively and sitting on my own agenda while listening to others. Actually this is my personal involvement." "I'm always listening, the amount of verbal output varies." This person checked a 5+ on the question regarding the degree to which he or she attempted to influence others to use this teacher's ideas. "I never feel that I would prefer doing something else. These sessions have helped me become more sensitive to my own responses and reactions with everyone I come in contact with." "I feel very involved in these sessions and I have learned that my own learning is increased by my more active participation. As a result I miss very few of these sessions. These sessions are about 80% of my reason for working on the Psychiatric Ward in this Hospital."

Questions Number 3 (values), Number 7 (influence), and Number 8 (effectiveness in influencing others), were designed to corroborate the idea that the personality of the teacher is an important variable. The mean score on Number 3 was 3.96. The mean score on Number 7 was 4.00. The mean score on Number 8 was 3.50. All of these means were above an average three. While most ratings tend to be at the high end of this scale the comments were more emphatic than the statistics. Some of the comments were as follows: "My years with this teacher have influenced my whole life style. I rely heavily on principles taught by this teacher, and they have been extremely rewarding to me in terms of results." "I appreciate his own perspective of moral and ethical values and his respect for the

difficulty people have in trying to change. . ." "The modeling behavior he exhibits is important to me. He strikes me as being a mildly confrontative person who is dedicated to helping me learn." "He shows much interest and enthusiasm. He models defining problems in behavioral terms." "Although I had been familiar with this teacher prior to these teaching sessions, I feel that personal and regular contact with him in this way has prompted some efforts of my own at behavioral change." "He sees everyone in the order that they ask for an appointment--not excluding anyone. He assumes everyone is trying his best." I interpreted attempts to acquaint others with this teacher's ideas as significant (Nos. 7 and 8) and some comments here were: "I am currently involved in a special teaching project. My major reason for involving myself in this project is to get this teacher's ideas in operation." "I feel almost a compulsion to teach others what I have learned. "I try." "I have attempted to teach wife and friends this teacher's ideas. I have had average success." "I have attempted to increase spontaneity and communication among members of a religious book study group. I hope to improve in influencing others in this regard." This person rated 3 on both question 7 and 8.

Questions Number 4 and Number 11 were designed to get the learner's impressions of the teaching techniques employed by this teacher. Question Number 4 (non-medical sources--divergence) had a mean score of 3.59. Question Number 11 was an open-ended question with no rating required. Some of the comments here were: "I can learn more in five

minutes with this teacher than I can in an hour with regular instruction. I am able to generate more ideas and apply them to different situations."

"This teacher is very frank yet somehow avoids offending others. This is very refreshing." "I like his ability to summarize and discriminate when dealing with apparently divergent and conflicting points of view."

"He is very skillful at rephrasing questions to make them answerable. He uses positive and negative reinforcement." "His unrelenting pursuit of alternative ways of approaching a concept and his continued high expectations are the most valuable payoff for having known this teacher."

"To be exposed to a teacher who displays at every meeting a systematic, step-by-step, clear, concrete, concise and interesting, down-to-earth approach to every problem is wonderful." "Learning the process of how he handles problems, rather than the content of each particular problem has helped me." "I've learned to be less threatened by someone who is much more knowledgeable than myself because he encourages my participation and shows a willingness to listen, even though many times he does not agree with me." "He has taught me an effective and comprehensive way of evaluating and analyzing problems." "He has encouraged me to think for myself and to experience more awareness of the responsibility for my own decisions." "He has helped me to realize that I have the right to be myself, which inspires me to do better." "He has taught me that sometimes it is necessary to make decisions and I should try to make the best ones, realizing that they won't all be good ones." "He has given

me the freedom to try new approaches to difficult problems which are contrary to traditional thinking." "He slaps me down with one hand but picks me up with the other. I respect him."

The ideas of protection and risk as important elements in this person's effective teaching are interesting ones. These ideas were picked up early in the study as the experimenter saw the teacher walk into class, say "Let's go. Any follow-ups, problems, etc.?" Then, as he would consider with the class what the students had been doing on their own, he selectively rewarded them for good work, pointed out possible errors, and supported and encouraged them to do the best they possibly could. He seemed to value an honest attempt even if it resulted in an honest mistake. He seemed always happy to protect and support them in their learning attempts. Questions Number 9 and Number 10 were designed to look at these ideas of protection and risk. The mean score on question 9 was 4.523. The mean score on Number 10 was 4.000. Some of the comments were as follows: "The element of risk is the freedom to make mistakes. This risk-taking has been a model for me in all areas of my life." "There is definitely an element of risk. Taking this allows the student to run himself differently than before--learn more about how his approaches are seen by others." "Risk is: he allows us to make mistakes. This is essential to learning." "Since most learning experiences of mine have been directed and structured, the idea of protection and support toward practical approaches of my own design is new."

## APPENDIX C

## QUESTIONNAIRE DATA 2

Introduction

In an effort to secure further information toward generating scale items a second questionnaire was designed by the investigator. The questionnaire asked for information regarding the effectiveness of the teacher and the general learning climate created by this teacher. One question was designed specifically to support the categories initially developed by the observational technique. This was an open ended question which asked the students why they rated the teacher as they did.

Subjects

The subjects were 92 second-year medical students taught by the same teacher.

Procedure

A questionnaire was administered to the students in this class at the end of the term. The questionnaire was accompanied by a stamped, self-addressed envelope for mailing of the completed questionnaire to the investigator. The questionnaires were anonymous. This questionnaire was coded and scored by the investigator. The coding system

was verified by another graduate student who randomly sampled a number of the questionnaires and was able to locate the responses in the coding system developed by observer.

### Results

Eighty-two questionnaires were answered and returned to the investigator.

The open ended question was coded and the results are as follows:

1. Dr. X is: human, sensitive, open minded, enthusiastic, well educated, well informed, direct, insightful, dynamic, innovative, liberal, good role model, adaptable, idealistic, mature, experienced, sincere, fair, talented, honest, realistic, broad based, good sense of humor, communicates well, impressive, thorough, logical.

Forty-four (44) questionnaires contained either one or a number of these responses. The mean rating of the teacher by these 44 students was 4.3. The investigator interpreted these responses as falling into the general category of personality of the teacher.

2. Dr. X knows what is important in learning process, pertinent handouts, good examples, incredible memory for material, interesting, well organized, synthesized well, good presentation, stimulating, well prepared, projects well, inspiring in the classroom, makes learning a joy, has right idea on traditional garbage, teaches more than most professors.

Thirty-seven (37) questionnaires contained one of a number of



these responses. The mean rating of the teacher by these students was 4.6. The investigator interpreted these responses as falling into the general category of techniques used by the teacher.

3. Dr. X is concerned for the students and for the education of the student, gives students credit for having some sense regarding their own learning styles, pays attention to student's feelings and suggestions, can see student difficulties quickly and gives immediate positive or negative feedback for correction and improvement, has students actually perform, encourages meaningful questions and discussions.

Twenty-seven (27) questionnaires contained one or a number of these responses. The mean rating of the teacher by these students was 4.3. The investigator interpreted these responses as falling into the general category of interaction--core of learning climate.

The other questions added no new data to the study. Thus, the results are not included.

## APPENDIX D

## PILOT TESTING OF PRELIMINARY SCALE ITEMS

Introduction

Based on the preceding information (Chapters III and IV) which the investigator interpreted to be inter-observer corroboration, 15 preliminary scale items were developed. One question was included by the investigator which was not taken from observational data per se, nor from questionnaire data. This question emerged from 1) a statement by the teacher that he liked to teach, 2) the personal characteristics of the teacher, 3) the observer's impression that the classes were pleasant to attend, and 4) the fact that this teacher chose to carry a heavy teaching load even though he was head of a major hospital department. The remaining questions were developed by extrapolating from the behaviors observed and from responses to the open ended questions of the two prior questionnaires.

Subjects

Eighteen teachers from the same department as the exemplar teacher were rated on this preliminary scale by their students. These teachers were teaching the second term of the class described in the previous study. The second term of the course utilized the small group format

by dividing the students from the previous term into 18 groups of approximately 5 students each.

Ninety-six second year medical students in the small groups and who had been taught by the exemplar teacher the previous term were given this preliminary scale and asked to rate their group instructor

### Procedure

The tentative scale items were administered to 96 students. The scale is as follows:

(Small Group Interviewing)

Instructor's name \_\_\_\_\_  
Date \_\_\_\_\_

Please place the appropriate number next to the question and answer all questions.

- |   | Never |   |   |   | Always |
|---|-------|---|---|---|--------|
|   | 1     | 2 | 3 | 4 | 5      |
| 1. The instructor makes clear to me what I'm expected to learn.   |       |   |   |   |        |
|   | _____ |   |   |   |        |
| 2. The teacher is able to explain concepts in a way I can understand  |       |   |   |   |        |
|   | _____ |   |   |   |        |
| 3. The teacher can demonstrate for me applications of these concepts.   |       |   |   |   |        |
|   | _____ |   |   |   |        |
| 4. The instructor is aware of what stage I am at in the learning process.   |       |   |   |   |        |
|   | _____ |   |   |   |        |
| 5. The instructor gives me prompt feedback and constructive criticism.  |       |   |   |   |        |
|   | _____ |   |   |   |        |
| 6. The instructor helps me move on to the next higher step in my learning process in a way that makes good sense. |       |   |   |   |        |
|   | _____ |   |   |   |        |

Never					Always
1	2	3	4	5	

7. The instructor allows me to make a try at the material to be learned with a minimum fear of penalty for making an error.

---

8. If while learning I should make a mistake, I feel the instructor would support me and help me learn from the mistake.

---

9. The instructor takes some personal and/or professional risk in allowing me to make mistakes.

---

10. The instructor involves himself--his skill, his knowledge, his feelings, in the learning process with his group.

---

11. The instructor deals honestly with me and with what is taking place at the moment in the group.

---

12. The instructor has a good knowledge of his subject.

---

13. The instructor seems not to care how I learn the material as much as that I learn the material.

---

Low					High
1	2	3	4	5	

14. The instructor seems to fit naturally into the teaching role.

---

15. Rate the overall effectiveness of this teacher for you.

---

COMMENTS: Any additional observations that could be useful to the instructor in improving his instructional competence are welcomed.

## Results

Sixty-four questionnaires were answered and returned to the departmental secretary. Thus, one-third of the students did not complete the questionnaire. Of the sixty-four completed questionnaires, twenty were not used in the correlations carried out on the data because they were incomplete. This resulted in a total N of 44.

For the purpose of this analysis the fifteen items were collapsed into a six item scale for two reasons: 1) the items seemed to naturally be subsumed into six basic content areas, and 2) the computer system available has a maximum leading of a 6 X 6 array. All numerical data from the 15 item scale were used in analyzing the collapsed scale. The numerical rating of question thirteen was inverted to agree with the other questions on the form. The combined questions resulted in the following six groups of items:

Group A combined questions 2, 3, and 12, and fell into the major category of personality characteristics, and techniques.

Group B combined questions 1, 4, 5, and 6 and fell into the major category of techniques.

Group C combined questions 7 and 8 and fell into the major category of interaction.

Group D combined questions 9 and 10 and fell into the major category of risk taking.

Group E combined questions 11 and 13 and fell into the major category

of interaction.

Group F combined questions 14 and 15 and were considered to be evaluation of instructor as a teacher.

The general categories of personal characteristics, techniques, interaction, and risk taking were correlated with the category of evaluation of teacher. The correlations between the grouped items and evaluation of the teacher are as follows. Y is the dependent variable, in this case it is the rating of effectiveness of this teacher (items 14 and 15 from the preliminary scale).

Group A - X vs Y = .61293 (techniques) (personal characteristics)

Group B - X vs Y = .53723 (techniques)

Group C - X vs Y = .50491 (interaction)

Group D - X vs Y = .51391 (risk)

Group E - X vs Y = .19754 (interaction)

## APPENDIX E

## TEACHER CHARACTERISTICS

Introduction

In the introduction to the section on "The Teacher's Personality and Characteristics" in Handbook of Research on Teaching, Getzels and Jackson (1963) state:

The personality of the teacher is a significant variable in the classroom. Indeed, some would argue it is the most significant variable. The educational impact of an Ichabod Crane or a Mark Hopkins, of a Mr. Chips or a Socrates, is surely not due solely to what he knows, or even to what he does, but in a very real sense what he is. There has always been concern with the personal qualities of teachers, and recently this concern has become the basis for a growing body of research.

Most of the studies done in this area have involved questionnaire techniques and some have used instruments such as the Minnesota Teacher Attitude Inventory, the Thematic Apperception Test, Kuder Preference Record, Strong Vocational Interest Blank, the Authoritarianism (f) Scale, the Minnesota Multiphasic Personality Inventory, etc. These are aimed at measuring attitudes. The personality factors of the teachers have been measured by instruments such as the Guilford-Zimmerman Temperament Survey, Cantell's Sixteen Personality Factor Questionnaire, and the Rorschach has been used, though infrequently.

The results of these kinds of studies have resulted in such global results (friendly teachers when compared to unfriendly ones would have a higher need for affiliation and a lower need for aggression), as to render them essentially useless as far as enabling people involved in education to use them either as predictors of who might be good teachers, or of evaluating those teachers already in the profession.

In 1960 Ryans did an extensive study of teacher characteristics. The reader is referred to this study as being valuable in terms of giving some information about the aims, methods, and findings of work on teacher personality and behavior of this period (Getzels and Jackson, 1963).

In this area as in most other areas of teacher evaluation, the literature indicates that there is very little known for certain about the nature and measurement of teacher personality, or about the relation between teacher personality and teacher effectiveness.

In an effort to add new information to this body of literature the following study was done.

#### Method

Each teacher of each sampled class at Portland State University (N=30, plus the exemplar teacher, N=31) was asked to fill out a questionnaire (adapted from one designed by Stuart).



The purposes of this section were:

1. To investigate emotional, inter-personal, intr-personal, and environmental aspects of the sampled teachers.

2. To compare the results of this investigation with regard to those teachers rated high on the scales administered to the classes (both the Davis and the investigator's), against those rated low on these scales. The high group contained the questionnaires answered by the teachers who had scored 6 or 7 on the scales answered by the class they were teaching as sampled for the dissertation.

At the time the students were filling out the rating scale the teachers filled out the questionnaire. Both were returned at the same time. Both were totally anonymous.

The questionnaire is as follows:

#### BIOGRAPHICAL INFORMATION ON TEACHER

1. Circle the number which represents the highest grade of schooling which you have completed.

1 2 3 4  
college

1 2 3 4  
Post-graduate

2. Circle the number which represents your age group  
19 and under      20-24      25-30      30-35      35-40      40 and over

3. My father and mother (circle) both approved of my career choice; both disapproved;      father disapproved      mother disapproved.

4. My childhood and adolescence, for the most part, were spent in (circle) open country; a town of 2,500; a city of 2,500 to 10,000; 10,000 to 50,000; 50,000 and over.

5. Did you ever attend Sunday school or other religious school for children and young people: (circle) Yes; No;

6. Religious activities (circle) never attend; attend less than once per month, more than four times per month.
7. Indicate the number of friends of the same sex you now have: (circle) almost none a few several many.
8. How much conflict was there between you and your father? (circle) None; very little; moderate; a good deal; almost continuous.
9. How much attachment was there between you and your father? None; very little; moderate; a good deal; very close.
10. How much conflict was there between you and your mother? (circle) None; very little; moderate; a good deal; almost continuous.
11. How much attachment was there between you and your mother? (circle) none; very little; moderate; a good deal; very close.
12. Give your appraisal of the happiness of your parents: (circle) very happy; happy; about averagely happy; unhappy; very unhappy.
13. My childhood on the whole was: (circle) very happy; happy; about averagely happy; unhappy; very unhappy.
14. In my childhood I was: (circle) punished severely for every little thing; was punished frequently; was occasionally punished; rarely; never.
15. In my childhood the type of training in my home was: (circle) exceedingly strict; firm but not harsh; usually allowed to have my own way; had my own way about everything; irregular (sometimes strict, sometimes lax).
16. Do you often feel lonesome? Yes No (circle)
17. Are you usually even tempered and happy in your outlook on life? (circle) Yes No
18. Do you usually avoid asking advice: (circle) Yes No
19. Do your feelings alternate between happiness and sadness without apparent reason (circle) Yes No
20. Are you considered critical of other people (circle) Yes No

21. Does discipline make you discontented? (circle) Yes No
22. Do you try carefully to avoid saying anything that may hurt anyone's feelings (circle) Yes No
23. Sex of person responding is: (circle) Female Male

### Results

A total tally was done on all the questions. A factor analysis was done on the following questions:

8. How much conflict was there between you and your father? (circle)  
None; very little; moderate; a good deal; almost continuous.
9. How much attachment was there between you and your father? None;  
very little; moderate; a good deal; very close.
10. How much conflict was there between you and your mother? (circle)  
None; very little; moderate; a good deal; almost continuous.
11. How much attachment was there between you and your mother? (circle)  
none; very little; moderate; a good deal; very close.
12. Give your appraisal of the happiness of your parents: (circle)  
very happy; happy; about averagely happy; unhappy; very unhappy.
13. My childhood on the whole was: (circle) very happy; happy;  
about averagely happy; unhappy; very unhappy.
14. In my childhood I was: (circle) punished severely for every little  
thing; was punished frequently; was occasionally punished; rarely;  
never.
15. In my childhood the type of training in my home was: (circle)  
exceedingly strict; firm but not harsh; usually allowed to have my  
own way; had my own way about everything; irregular (sometimes strict,  
sometimes lax).

Three factors emerged as shown in the following table:

FACTOR ANALYSIS ON BIOGRAPHICAL INFO ON TEACHERS

NORMALIZED UNROTATED FACTOR LOADINGS

Variable	1	2	3
Q 8	0.57121E 00	-0.69334E-01	0.67953E 00
Q 9	-0.70619E 00	0.14800E 00	0.10206E-01
Q 10	0.23488E 00	0.60371E 00	-0.52895E 00
Q 11	-0.74756E 00	0.14886E 00	0.31926E 00
Q 12	0.62532E 00	0.24773E 00	0.20123E-01
Q 13	0.50837E 00	0.55517E 00	-0.46356E-01
Q 14	-0.68366E 00	0.34524E 00	-0.65757E-01
Q 15	-0.11230E 00	0.77923E 00	0.46310E 00

The three factors which emerged could be interpreted as having to do with:

1. Inter-family relationships.
2. Attachment to mother.
3. Is a bi-polar factor having to do with parental attachment and general happiness.

A t-test was done to ascertain whether or not there was any difference between the high and low rated teachers on these factors.

There was no statistical difference on any of the factors.

MEAN 1 REFERS TO THE LOW GROUP  
MEAN 2 REFERS TO THE HIGH GROUP

T = 0.806763	MEAN 1 = 0.095499
T = 0.277859	MEAN 1 = 0.031868
T = 0.141780	MEAN 1 = 0.016325

A brief description of the correlations shows that:

Question 8 correlates with 9 (-3864), 14 (-41935). This is

conflict with father, little attachment to father and frequent punishment.

Question 9 correlates with 11 (45570). This might mean that there is a relationship between attachment to father and an attachment to mother. This might indicate a good home environment.

Question 10 correlates with no other question in the matrix with the possible exception of a weak correlation with question 13 (happy childhood) (2996).

Question 11 correlates with 9 (41619), 14 (46768). This has to do with attachment to mother, attachment with father, rare punishment.

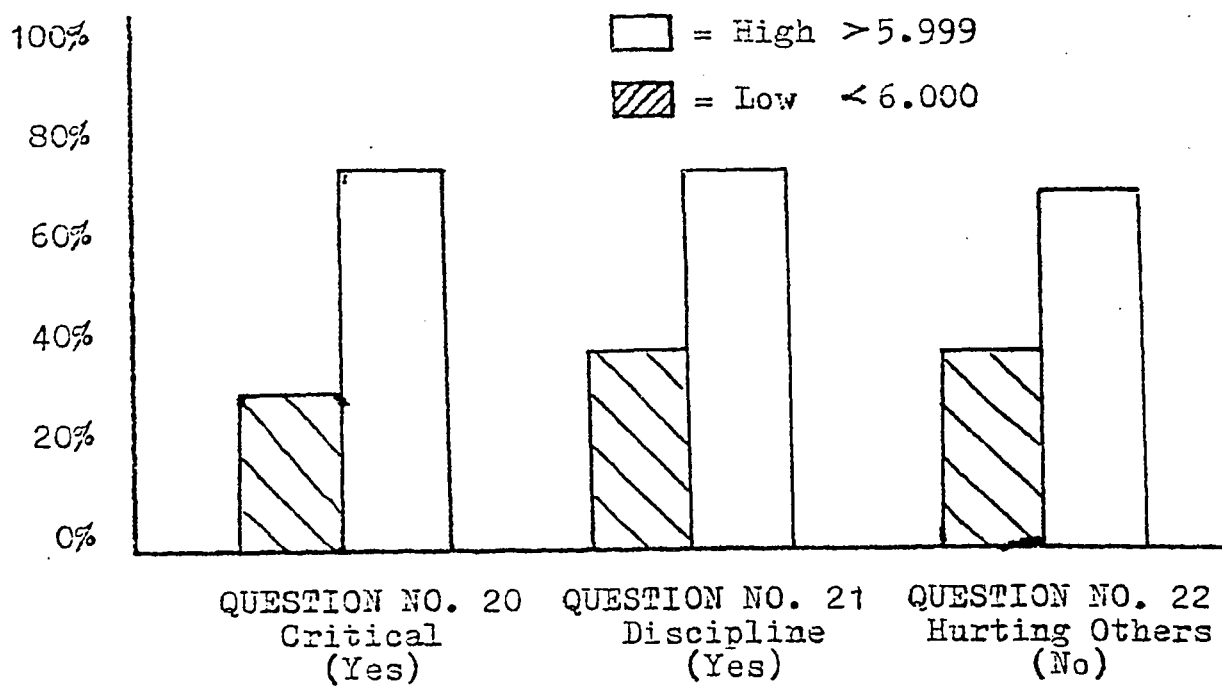
Question 12 correlates with 9 (-45570). This is happiness of parents with little attachment with the father.

Question 13 correlates with none of the other questions (childhood happiness).

Question 14 correlates with 8 (-41935), 11 (46768). Punishment is correlated negatively with conflict with father, and positively with attachment with mother.

Question 15 correlates with none of the others. A very slight correlation (all.2-3) with attachment to mother, happy childhood, rarely punished.

On the other questions on the questionnaires which were binominal and so required t-tests, there were no statistical difference between the high and low rated group except on questions 20, 21, and 22. These differences are presented in tabular form in the following table.



### Dimensions of an effective teacher

The method discussed in this study revealed 16 underlying traits of an effective teacher. Specifically, cued behaviors were grouped into the following sixteen, sub-categories:

1. creates positive learning climate
2. demonstrates knowledge
3. reaches closure
4. communicates well
5. gives immediate feedback
6. reinforces learning
7. lets student make own mistakes
8. experiential learning
9. egalitarianism
10. uses self-correcting in light of new knowledge
11. uses divergent approaches
12. encourages student participation and supports their efforts
13. enjoys teaching
14. is trusting
15. risk taking
16. knowledgeable

These sixteen sub-categories were then formed into four major categories: interaction, personal characteristics of the teacher, techniques and risk taking. It was these major categories which were reflected in the final scale of items.

The major categories might be divided into two classes: 1) the process, what a good teacher does, and 2) the content, what a good teacher is.