



Professional educators agreement on criterion for measuring teacher effectiveness  
by Francis Allen Olson

A thesis submitted in partial fulfillment of the requirements for the degree of DOCTOR OF  
EDUCATION

Montana State University

© Copyright by Francis Allen Olson (1978)

Abstract:

This study investigated the criteria by which administrators and teachers, employed by Montana school districts, judged teacher effectiveness.

A similar study, which this study replicated, was carried out in the State of Delaware by Jenkins and Bausell.

The data for this study was gathered by a survey questionnaire which listed criteria for judging teacher effectiveness typed under the Mitzel Scheme. Administrators were sampled first and with their return an answer was given as to whether or not they gave permission to sample teachers on their respective staffs. Samples of administrators (N=665) was followed by sampling of teachers (N=9,428) and these results were compared by analysis of variance statistic at .05 level of significance. Comparison of the results of the study with the Delaware study was done by using Spearman's Coefficient of Rank Correlation. The Multiple Regression model was used to determine whether or not a significant relationship existed among the differences between the administrators and teachers' rating of each criterion measure of effectiveness and the teachers' ratings of administrators.

Conclusions reached are: (1) The highest rated criterion by teachers and administrators in Montana for measuring effectiveness was "classroom control" followed by "knowledge of subject matter" and "rapport with students". (2) The "amount students learn" a criterion uppermost in the minds of accountability proponents was considerably less significant in the minds of teachers. (3) Product criteria (measure of student learning and behavior) and process criteria (measure of teacher behavior) were rated significantly higher than presage criteria (measure of a teacher's personal or intellectual attributes). (4) Teachers' rating of effectiveness criteria was not significantly related to how teachers viewed their administrators' effectiveness. (5) Montana and Delaware teachers were in agreement that process and product criteria for measuring effectiveness are considerably more important than presage criteria. (6) Both Montana and Delaware teachers do not consider that "what students learn", is as important a criterion by which to measure effectiveness as other criteria. This view differs considerably from that of accountability proponents.

One of the more important recommendations coming out of this study is to determine whether or not parents and other constituents of Montana served by administrators and teachers are in agreement among themselves and with educators on what types of criteria are most important by which to judge effectiveness. It would be important to know if discipline (effectiveness in controlling his class) is the number one rated criterion.

© 1979

FRANCIS ALLEN OLSON

ALL RIGHTS RESERVED

PROFESSIONAL EDUCATORS AGREEMENT ON CRITERION  
FOR MEASURING TEACHER EFFECTIVENESS

by

FRANCIS ALLEN OLSON

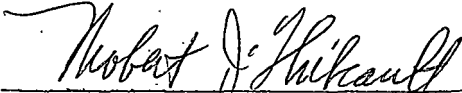
A thesis submitted in partial fulfillment  
of the requirements for the degree

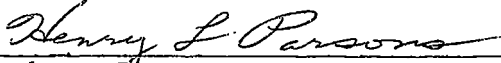
of

DOCTOR OF EDUCATION

Approved:

  
Chairperson, Graduate Committee

  
Head, Major Department

  
Graduate Dean

MONTANA STATE UNIVERSITY  
Bozeman, Montana

December 1978

## ACKNOWLEDGMENTS

Special acknowledgment goes to Dr. Robert Thibeault, my advisor and committee chairman. He has graciously provided many hours of guidance and encouragement throughout this study.

Acknowledgment is made to my reading committee, Dr. Albert Suvak, and Dr. Robert Van Woert and to other members of my committee, Dr. Douglas Herbster, Dr. Del Samson, and Dr. Alvin Fiscus.

Acknowledgment is also made to Dr. Eric Strohmeyer for his help and support.

Special thanks is given to Mrs. Louise Greene who spent many hours typing this dissertation.

Most of all I wish to thank my wife, Alyce, for the many hours that she spent helping with recording of data, running errands, and providing needed patience and encouragement.

TABLE OF CONTENTS

|  | Page |
|--|------|
| LIST OF TABLES . . . . .   |      |
| CHAPTER  |      |
| I. INTRODUCTION . . . . .  | 1    |
| STATEMENT OF THE PROBLEM . . . . .   | 15   |
| NEED FOR THE STUDY . . . . .   | 15   |
| PURPOSE OF THE STUDY . . . . .   | 16   |
| QUESTIONS TO BE ANSWERED . . . . .   | 17   |
| LIMITATIONS OF THE STUDY . . . . .   | 17   |
| DEFINITIONS OF TERMS . . . . .   | 18   |
| SUMMARY . . . . .  | 19   |
| II. REVIEW OF RELATED LITERATURE . . . . .                                   | 22   |
| INTRODUCTION . . . . .   | 22   |
| NEED FOR EVALUATING TEACHER EFFECTIVENESS . . . . .                          | 24   |
| FORCES WHICH CREATED THE NEED TO EVALUATE<br>TEACHER EFFECTIVENESS . . . . . | 25   |
| RENEWED EMPHASIS PLACED UPON MEASURING<br>TEACHER EFFECTIVENESS . . . . .    | 38   |
| TEACHER PARTICIPATION IN THE EVALUATION<br>OF THEIR SERVICES . . . . .       | 48   |

| CHAPTER  | Page |
|--|------|
| STUDIES OF TEACHER EFFECTIVENESS . . . . .                               | 51   |
| Product criteria . . . . .   | 69   |
| Process criteria . . . . .   | 69   |
| Presage criteria . . . . .   | 70   |
| STATUS OF PRESENT METHODS OF APPRAISING<br>TEACHER PERFORMANCE . . . . . | 70   |
| SUMMARY . . . . .  | 73   |
| III. PROCEDURES . . . . .  | 77   |
| DESCRIPTION OF THE POPULATION . . . . .                                  | 77   |
| SAMPLING PROCEDURE . . . . .   | 81   |
| METHOD OF COLLECTING DATA . . . . .                                      | 84   |
| METHOD OF ORGANIZING DATA . . . . .                                      | 86   |
| HYPOTHESES TESTED . . . . .  | 92   |
| HYPOTHESIS I . . . . .   | 92   |
| HYPOTHESIS II . . . . .  | 93   |
| HYPOTHESIS III . . . . .   | 93   |
| HYPOTHESIS IV . . . . .  | 93   |
| HYPOTHESIS V . . . . .   | 93   |
| HYPOTHESIS VI . . . . .  | 94   |
| HYPOTHESIS VII . . . . .   | 94   |
| METHOD OF ANALYZING DATA . . . . .                                       | 94   |
| PRECAUTIONS TAKEN FOR ACCURACY . . . . .                                 | 96   |
| SUMMARY . . . . .  | 96   |

| CHAPTER   | Page |
|---|------|
| IV. ANALYSIS OF DATA . . . . .                        | 98   |
| INTRODUCTION . . . . .                                | 98   |
| MEAN RATINGS OF CRITERIA . . . . .                    | 99   |
| THE TESTING OF HYPOTHESES . . . . .                   | 103  |
| SUMMARY . . . . .                                     | 138  |
| V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS . . . . . | 143  |
| SUMMARY . . . . .                                     | 143  |
| CONCLUSIONS . . . . .                                 | 147  |
| RECOMMENDATIONS FOR FURTHER STUDY . . . . .           | 149  |
| LITERATURE CITED . . . . .                            | 151  |
| APPENDIX A . . . . .                                  | 160  |
| APPENDIX B . . . . .                                  | 166  |

## LIST OF TABLES

| TABLE   | Page |
|---|------|
| I. CATEGORIES OF ADMINISTRATOR POPULATION . . . . .   | 80   |
| II. TEACHER SAMPLE CHARACTERISTICS . . . . .  | 81   |
| III. CATEGORIES OF THE ADMINISTRATOR SAMPLE . . . . .   | 83   |
| IV. NUMBER AND PER CENT OF TEACHER RESPONDENTS<br>BY DISTRICT AND SEX CATEGORIES . . . . .  | 84   |
| V. YEARS OF EXPERIENCE OF ADMINISTRATORS . . . . .  | 88   |
| VI. THE YEARS OF EXPERIENCE OF TEACHER RESPONDENTS<br>BY DISTRICT CLASSIFICATION AND SEX . . . . .  | 89   |
| VII. PERCENTAGE OF TEACHERS BY DISTRICT CLASSIFICATION .  | 90   |
| VIII. NUMBER TEACHER RESPONDENTS BY GRADE LEVEL<br>CLASSIFICATION . . . . .   | 91   |
| IX. RANK ORDER OF MEANS FOR SIXTEEN CRITERION MEASURE<br>OF TEACHER EFFECTIVENESS FOR ADMINISTRATORS OF<br>COMBINED CLASSES OF SCHOOL DISTRICTS . . . . . | 101  |
| X. RANK ORDER OF MEANS FOR SIXTEEN CRITERION MEASURES<br>OF TEACHER EFFECTIVENESS FOR TEACHERS OF COMBINED<br>CLASSES OF SCHOOL DISTRICTS . . . . .       | 102  |
| XI. RANK ORDER OF MEANS FOR SIXTEEN CRITERION MEASURES<br>OF TEACHER EFFECTIVENESS FOR TEACHERS OF CLASS I<br>SCHOOL DISTRICTS . . . . .                  | 104  |
| XII. RANK ORDER OF MEANS FOR SIXTEEN CRITERION MEASURES<br>OF TEACHER EFFECTIVENESS FOR TEACHERS OF CLASS II<br>SCHOOL DISTRICTS . . . . .                | 105  |
| XIII. RANK ORDER OF MEANS FOR SIXTEEN CRITERION MEASURES<br>OF TEACHER EFFECTIVENESS FOR TEACHERS OF CLASS III<br>SCHOOL DISTRICTS . . . . .              | 106  |



| TABLE   | Page |
|---|------|
| XIV. COMBINED MEANS OF RATINGS OF ADMINISTRATORS<br>AND TEACHERS OF MONTANA . . . . .   | 107  |
| XV. ADMINISTRATORS VERSUS TEACHERS LEAST SQUARE<br>MEANS AMONG THE THREE SUB-TESTS BETWEEN<br>ADMINISTRATORS AND TEACHERS . . . . .   | 109  |
| XVI. LEAST SQUARE MEANS OF COMPARISON OF TEACHERS OF<br>THREE CLASSES OF DISTRICTS: TEACHERS OF 1st,<br>IIInd, AND IIIrd CLASS SCHOOL DISTRICTS . . . . .   | 111  |
| XVII. A COMPARISON OF LEAST SQUARE MEANS FOR MALE<br>AND FEMALE TEACHERS . . . . .  | 114  |
| XVIII. LEAST SQUARES MEANS FOR ELEMENTARY VERSUS<br>SECONDARY TEACHERS FOR THREE SUB-GROUPS OF<br>TEACHER EFFECTIVENESS CRITERIA . . . . .  | 116  |
| XIX. LEAST SQUARE MEANS OF SIX EXPERIENCE CLASSES<br>FOR TEACHERS . . . . .   | 118  |
| XX. MULTIPLE REGRESSIONS AMONG THE 16 INDEPENDENT<br>VARIABLES OF TEACHER AND ADMINISTRATOR DIFFER-<br>ENCES AND THE DEPENDENT VARIABLE, TEACHER<br>RATING OF ADMINISTRATOR EFFECTIVENESS . . . . . | 122  |
| XXI. MEAN RATINGS AND RANK ORDER OF THE 16 CRITERIA<br>DELAWARE STUDY . . . . .   | 125  |
| XXII. CALCULATION OF SPEARMAN'S COEFFICIENT OF RANK<br>CORRELATION FOR MONTANA TEACHERS AND ADMINISTRATORS  | 129  |
| XXIII. CALCULATION OF SPEARMAN'S COEFFICIENT OF RANK<br>CORRELATION FOR MONTANA AND DELAWARE TEACHERS . .   | 130  |
| XXIV. CALCULATION OF SPEARMAN'S COEFFICIENT OF RANK<br>CORRELATION FOR MONTANA ADMINISTRATORS AND<br>DELAWARE TEACHERS . . . . .  | 131  |
| XXV. CALCULATION OF SPEARMAN'S COEFFICIENT OF RANK<br>CORRELATION FOR MONTANA TEACHERS OF FIRST AND<br>SECOND CLASS SCHOOL DISTRICTS . . . . .  | 132  |

| TABLE  | Page |
|--|------|
| XXVI. CALCULATION OF SPEARMAN'S COEFFICIENT OF RANK<br>CORRELATION FOR MONTANA TEACHERS OF FIRST AND<br>THIRD CLASS SCHOOL DISTRICTS . . . . .   | 133  |
| XXVII. CALCULATION OF SPEARMAN'S COEFFICIENT OF RANK<br>CORRELATION FOR MONTANA TEACHERS OF SECOND AND<br>THIRD CLASS SCHOOL DISTRICTS . . . . . | 134  |
| XXVIII. T-TEST OF SIGNIFICANCE OF THE MEAN DIFFERENCE<br>FOR TEACHER AND ADMINISTRATOR . . . . .   | 137  |

## ABSTRACT

This study investigated the criteria by which administrators and teachers, employed by Montana school districts, judged teacher effectiveness. A similar study, which this study replicated, was carried out in the State of Delaware by Jenkins and Bausell.

The data for this study was gathered by a survey questionnaire which listed criteria for judging teacher effectiveness typed under the Mitzel Scheme. Administrators were sampled first and with their return an answer was given as to whether or not they gave permission to sample teachers on their respective staffs. Samples of administrators (N=665) was followed by sampling of teachers (N=9,428) and these results were compared by analysis of variance statistic at .05 level of significance. Comparison of the results of the study with the Delaware study was done by using Spearman's Coefficient of Rank Correlation. The Multiple Regression model was used to determine whether or not a significant relationship existed among the differences between the administrators and teachers' rating of each criterion measure of effectiveness and the teachers' ratings of administrators.

Conclusions reached are: (1) The highest rated criterion by teachers and administrators in Montana for measuring effectiveness was "classroom control" followed by "knowledge of subject matter" and "rapport with students". (2) The "amount students learn" a criterion uppermost in the minds of accountability proponents was considerably less significant in the minds of teachers. (3) Product criteria (measure of student learning and behavior) and process criteria (measure of teacher behavior) were rated significantly higher than presage criteria (measure of a teacher's personal or intellectual attributes). (4) Teachers' rating of effectiveness criteria was not significantly related to how teachers viewed their administrators' effectiveness. (5) Montana and Delaware teachers were in agreement that process and product criteria for measuring effectiveness are considerably more important than presage criteria. (6) Both Montana and Delaware teachers do not consider that "what students learn", is as important a criterion by which to measure effectiveness as other criteria. This view differs considerably from that of accountability proponents.

One of the more important recommendations coming out of this study is to determine whether or not parents and other constituents of Montana served by administrators and teachers are in agreement among themselves and with educators on what types of criteria are most important by which to judge effectiveness. It would be important to know if discipline (effectiveness in controlling his class) is the number one rated criterion.

## CHAPTER I

### INTRODUCTION

Defining the effective teacher has been a continuing process carried out by researchers for many years. What research has said about teacher effectiveness is that it is not the clearly defined trait that many would have us believe. Research has indicated that teacher performance is one of the most complex human phenomenon that researchers have been privileged to study (Ellena, 1964).

Teacher effectiveness has continued to be a subject of much interest to educators and more recently to the public whom they serve. Interest in teacher effectiveness by the public has arisen from the current viewpoint of accountability in education which has focussed on the people's right to know what is taking place in the schools (House, 1973).

Addressing the February, 1975 meeting of the American Association of School Administrators which was held in Dallas, Texas, Frank Gary stated:

Measuring the effectiveness of teaching . . . is still a topic of interest to the public and school administrators. There has not been too much progress in the area of measuring practitioner effectiveness because of the educational stance that it is impossible to make valid judgments about anything as complex and personal as teaching ability (Gary, 1975).

Biddle and Ellena point out in the preface to their review of research that:

Probably no aspect of education has been discussed with greater frequency, with as much deep concern, or by more educators and citizens than has that of teacher effectiveness--how to define it, how to identify it, how to measure it, how to evaluate it, and how to detect and remove obstacles to its achievement (Biddle and Ellena, 1964).

Research has said that teacher effectiveness cannot be summarized in a few words. However, many people who have had contact with schools--whether as students, parents, or interested citizens--feel qualified to make dogmatic pronouncements about teacher effectiveness. Teacher effectiveness has been a matter of long concern in all efforts to improve education. Before the turn of the century, studies were conducted in this country which attempted to isolate the factors which contributed significantly to teaching effectiveness. One bibliography alone lists 1,006 researches made in this area from 1890 to 1949 (Domas and Tiedeman, 1950).

It has been pointed out that teaching must be defined before it can be evaluated and effectiveness predicted. From Ellena's summary of teacher effectiveness studies, it was evident that part of the difficulty associated with the prediction of teacher effectiveness had arisen from the fact that teaching was described differently by different people, and the teaching act varied from person to person, and from situation to situation.

One of the most difficult problems in teacher effectiveness studies has been that researchers had to assume that effectiveness was either a

statement about an attribute of the teacher, a statement about an attribute of a teacher in a particular teaching situation, or a statement about the results which come out of a teaching situation (Ellena, 1961). Gage, who searched for a scientific basis to describe teacher effectiveness pointed out that during most of the history of education, "What knowledge, understanding, and ways of behaving should teachers possess," has been found through raw experience, tradition, common sense, and authority (Gage, 1972). He defined research on teacher effectiveness as the relationship between teacher behaviors and characteristics and their effects on students. In their relationship teacher behavior was considered an independent variable (Gage, 1972).

Flanders has stated that, "Knowledge about teaching effectiveness consists of relationships between what a teacher does while teaching and the effect of these actions on the growth and development of his pupils." From his point of view an effective teacher interacted skillfully with pupils in such a way that they learned more and liked learning better compared with the ineffective teacher. He described teaching effectiveness as being concerned with those aspects of teaching in which the teacher "has direct control and current options" (Flanders, 1970).

A number of researchers as well as some professional associations supported the position that the ultimate criterion by which to judge a teacher's competence was the impact that the teacher exerted upon the learner to bring about behavioral change in the learner. Reluctance in

accepting pupil change as the chief criterion of teacher effectiveness has arisen both from the technical problems in assessing learner growth and from philosophical considerations (Travers, 1973). The chief concern among the technical problems in assessing learner growth has centered on the adequacy of measures for assessing a wide range of pupil attitudes and achievement at different educational levels, and in diverse subject-matter areas. Philosophical differences have centered upon the selection of desirable changes to be sought in learners and value differences observed in the preferred methodologies of teacher competence researchers (Travers, 1973).

Research, which has tried to determine that teaching effectiveness has something to do with what a teacher is, assumed that teacher success can be predicted in terms of individual teacher personality traits. Both laymen and the majority of professional educators have clung to the idea that ability to teach is correlated in some way with such personality factors as a sense of humor, empathy, industriousness, willingness to cooperate, physical attractiveness and health, love of knowledge, creativity, and so forth. To some extent, nearly every teacher evaluation program in existence has taken these factors into consideration. Yet numerous research studies have failed to find a significant cause-and-effect relationship between traits and teaching effectiveness. Research has shown that we tend to place the highest values on those traits we ourselves possess or think we possess (Brighton, 1965).

Barr cautioned those who pursued the traits approach to describing teacher effectiveness that personal qualities such as considerateness, cooperativeness, ethicality, which are used to assess teacher effectiveness are not directly observable. These qualities are inferences drawn from data. He described this concern as follows:

These data may be of many sorts arising from the observation of behavior, interviews, questionnaires, inventories, or tests. Whatever the source of information, judgments, about the qualities are inferences, and subject to all the limitations, associated with inference making including the accuracy of the original data upon which the inferences are based, and the processes of inference making (Barr, 1961).

Barr related that if one has considered the qualities of the individual in terms of characteristics of performance, he has utilized a behavioral approach to assess teacher effectiveness. He noted further that those who have interpreted personality in behavioral terms in assessing teacher effectiveness have attempted to integrate the concept of personality with that of methods. Historically, this concept has been considered an important aspect of teacher effectiveness. The problem which has been encountered in the behavioral approach to assessing teacher effectiveness is that of choosing and defining the personal qualities that have appeared to be pertinent to teacher effectiveness. The literature has given one the impression that the choice of personal qualities used to assess teacher effectiveness has been based very much upon personal preference (Barr, 1961). Barr stated,

If judgments about teachers are based upon observations of teachers' behaviors, how do we know what to look for and what to



ignore? Whether a behavior, or aspect of behavior, is pertinent to some particular quality depends on how the quality is defined. Many subtle shades of meanings will probably need to be considered (Barr, 1961).

The personality of the teacher has been a significant variable in the classroom. Many have argued that the educational impact of a teacher is not due solely to what he knows or does, but to what he is as well. After an in-depth study of this problem, Getzels and Jackson concluded that despite the critical importance of the problem and a half-century of prodigious research effort, little is known for certain about the nature and measurement of teacher personality, or about the relation between teacher personality and teacher effectiveness (Averch and others, 1971) (Lewis, 1973) (Gage, 1972).

In his summary on teacher effectiveness studies, Ellena pointed out that teachers differ widely with respect to maturity, intellectuality, personality, and other characteristics. The demands of the subjects they teach, the scope and the structure of the objectives to be achieved--all contribute to diversity. In addition, Ellena noted that local control has exerted its influence toward diversity. For almost any goal one might choose, it was possible to find a continuous spectrum of values, opinions, and goals. The notion of the "good teacher" described by Ellena as basic to the study of teacher effectiveness has turned out to be almost as vague and diffused as the range of human experiences relative to teaching (Ellena, 1961). Rabinowitz and Travers summarized their problem, and repeated by Ellena, as follows:

There is no way to discover the characteristics which distinguish effective and ineffective teachers unless one has made or is prepared to make a value judgment. The effective teacher does not exist pure and serene, available for scientific scrutiny, but instead a fiction in the mind of men. No teacher is more effective than another except as someone so decides and designates . . . (Ellena, 1961).

Most likely the reason that the effective teacher has not existed pure and serene has been due to the fact that under local control of schools, the teaching act is free to vary from school system to school system. The job of the teacher thus varies according to the location of the job. The particular job a teacher has been expected to perform has varied from grade to grade. Because definitions of teacher functions has varied by grade level and school system little headway has been made in solving the problem of successfully measuring teacher effectiveness in spite of the immense number of studies that have been conducted (Ellena, 1961).

To make sense of the diverse inquiries that have been undertaken in the name of teacher effectiveness, Travers related that it has become necessary for one to make "distinctions in purposes". He pointed out that the administrator has been looking for knowledge of teacher effectiveness in order to make a better decision in situations such as hiring or firing a teacher. The instructional supervisor or teacher wanted to know what instructional procedures are most likely to prove useful in achieving certain instructional ends with given students. Researchers' purposes, according to Travers, included:

satisfying a desire to describe accurately what teachers do, searching for associations between theoretically or empirically derived variables and learning, and demonstrating the power of a given factor or instructional operation to make a practical difference upon the outcome sought (Travers, 1973).

Those who have been interested in teacher effectiveness have had different purposes and consequently have varied their interpretations of the problem. Some who have investigated the problem of teacher effectiveness would have been satisfied to know whether or not a teacher was getting desired results with the results indicating effectiveness, not the process used. Others wanted to know how to increase the probability of attaining desired results. Researchers who were interested in process were searching for lawful teaching behavior, *i.e.*, validated procedures for achieving instructional ends. Their assumption was that effective teaching would have been recognized when lawful relationships were established between instructional variables and learner outcomes--that certain procedures in teaching would have, within certain probability limits, been labeled as effective or ineffective (Travers, 1973).

To date there are no such laws, only a few leads or practices that are more likely than others to maximize the attainment of selected instructional ends. Researchers such as Gage (1968) had hoped to establish scientific laws for teaching; other researchers agreed with Dewey (1929), who held that it was an error to believe that scientific findings and conclusions from laboratory experiments to such activities as

helping the teacher make his practice more intelligent, flexible and better adapted to dealing with individual situations (Travers, 1973).

Bolton suggested that the purposes of teacher evaluation vary somewhat from school district to school district. Included were many of the following:

- (a) to improve teaching . . . by determining what actions can be taken to improve teaching systems, the teaching environment, or teacher behavior,
- (b) to supply information for modification of assignments,
- (c) to protect individuals and the school system from incompetence,
- (d) to reward superior performance,
- (e) to validate the selection process, and
- (f) to provide a basis for the teacher's career planning and growth and development.

Bolton summarized by stating, "All of these purposes might be expressed by saying: The purpose of teacher evaluation is to safeguard and improve the quality of instruction received by students" (Bolton, 1973).

Wilson, in presenting a paper to the annual convention of the National School Boards Association, April 17-22, 1975, described the purposes of evaluation as follows:

Before we come to grips with the methods to be used in evaluating teachers, there must be a clear understanding of the purpose for the evaluation in the first place. As a superintendent of schools it is clear to me that teachers are evaluated for two major reasons. First, teacher evaluation takes place for the

specific purpose of improving the quality of instruction. The focal point of all education is the learner . . . . The second major reason for teacher evaluation is to identify those staff members who are perpetrating such crimes against youngsters that their removal from the classroom and from the profession is the major objective. In other words the evaluation process is used to document teacher ineffectiveness so that termination can be accomplished (Wilson, 1974).

From the standpoint of the local school official, Ellena pointed out that the extent to which any procedure has been used in teacher evaluation depended on how much and what kind of evidence was desired in making decisions about local school personnel. These concerns were for immediate and self-terminating information. There was no concern from the local standpoint about adding to the fund of knowledge about teacher effectiveness to the extent that it could be predicted and explained accurately (Ellena, 1961). Travers, in his review on teacher effectiveness, related that decisions require judgments about teachers have been made by many--teacher educators, school personnel officers, administrators, supervisors; and teachers. Wise choices about teachers have been made when adequate data was at hand for judging. He added:

Complete data have typically not been available; possibly because those who have been making decisions have not given enough thought to what is required for making warranted decisions about a teacher and, accordingly have not arranged for the collection of data.

A second reason that data are not available, according to Travers, is that researchers have not pursued their investigations with awareness of the practical decisions that must be made by those working with teachers (Travers, 1973).

The school official, as suggested by Ellena, has sought "to determine how well a teacher performed his job in terms of certain specified and more often unspecified criteria." He has not been concerned with whether or not the job he asked the teacher to perform was representative of the class of such jobs or if the teacher performed the class of jobs well. On the other hand, the researcher, according to Ellena, has been concerned with how well a teacher could perform, "in any of a class of jobs which share many common characteristics, as well as with identifying these common characteristics" (Ellena, 1961).

The difference between a school official's concern and a researcher's concern, as pointed out by Ellena, has several implications. For example, the overall or intuitive ratings may be used by a school official to help make a general assessment of how well a teacher has performed and the general assessments thus gained has provided relevant and useful information for the immediate school situation. From Ellena's point of view, overall ratings have not generally been useful or relevant because such ratings have low reliabilities, and have not been consistent with the purposes of researchers who wish to predict and to describe. Ellena has evaluated ratings as follows:

An overall rating for research purposes implicitly assumes that when a teacher receives a rating of 80 per cent effective, it means that a teacher is as effective in doing the same things as every other teacher who is rated as 80 per cent effective by other school officials. An overall scale does not show that teachers are effective in doing the same things. It only shows that raters thought teachers were effective in doing whatever it

was they did. An overall rating is simply a means of letting the criterion which the researcher wants to predict, vary in its meaning, without showing that it so varies. It is impossible to predict consistently a criterion whose meaning constantly shifts (Ellena, 1961).

As pointed out by Ellena, officials of local school districts and researchers have had different purposes for describing teacher effectiveness. As a result of this difference in purpose researchers experienced the problem of predicting a criterion of teacher effectiveness that was relevant to local school board needs. Ellena described this problem by stating:

If a researcher uses any procedure that permits the definition of teacher effectiveness to vary, he will not be successful in predicting. If he does not let the definition vary, he places himself in the position of having to specify what the function of teachers should be. If he so specifies, he either usurps the function of local school districts in deciding what the functions of a local teacher should be, or else runs the risk of predicting a criterion that some local school boards considered inconsequential or irrelevant to how they define teacher performance (Ellena, 1961).

Travers has stated in his Second Handbook of Research on Teaching that, "Professionals and laymen alike are unhappy with what is loosely called the evaluation of teachers" (Travers, 1973). He summarized the results of national surveys which indicated the reasons for dissatisfaction with most evaluations are: (1) lack of confidence in the school system's evaluation program, (2) infrequent observation of tenured teachers, (3) inaccurate evaluation, (4) administrative staff have little time to effectively evaluate and make judgments of staff, and (5) evaluations are poorly communicated to others (Travers, 1973).

Many considerations beside teacher effectiveness entered into decisions such as whether to hire, to grant tenure, to fire teachers. Travers pointed out that the practice of assessing a teacher without having had valid data regarding his ability to effect changes in pupils seemed wanting. In contrast, information about the teacher's personal characteristics, relations with other adults, appearance, political attitudes, etc., has been plentiful and easily acquired. Appraisals of a teacher on the basis of factors unrelated to the progress of pupils has allowed the value preference of individuals and local communities to operate (Travers, 1973).

Bolton expressed the view that judgments regarding teachers are made inevitably and if the criteria were appropriate and the data were sound, resulting judgments would be useful. He stressed the fact that in evaluating teachers, judgments should be made in relation to objectives rather than the personal worth of people. He pointed out that evaluation should establish whether the teacher reached various standards, not whether the teacher did better or worse than other teachers. He emphasized the idea that teachers should be helped to improve their contribution to the learning of school children (Bolton, 1973).

A review of research supported the position that the more widely used criteria for assessing teacher competency included student ratings, self ratings, administrator ratings, and peer ratings. Assessments of classroom environment, personal attributes, performance tests, alterna-



tive criteria (contract plans using student gain) and systematic observations provided additional criteria for judging teacher effectiveness. Travers reviewed the work of McNeil and Popham who have cautioned in their assessment of teacher competence,

Any single criterion of effectiveness is confounded by a number of factors. One factor stems from who is doing the measuring; a second is the kind and quality of instrument used; a third is faithfulness in applying the instrument as its designer intended; and a fourth is the purpose for applying the criteria--how the data are used (Travers, 1973).

Research has generally supported the conclusion that effectiveness in teaching is best evidenced by criterion measures which detect pupil growth as a result of the teacher's instruction. However, as pointed out by Wolf, teachers are not fond of evaluation. He stated their concern as follows:

. . . They suspect any measure designed to assess the quality of their teaching, and any appraisal usually arouses anxiety.

If teachers are to submit to an assessment of their performance, they would probably like reassurance that the criteria and method of evaluation that are used would produce credible results (House, 1973).

According to Wolf, teachers have believed that the standards for evaluating what is effective teaching are too vague and ambiguous to be worth anything. They have felt that current appraisal techniques fall short of collecting information that accurately characterize their performance. They received the ultimate rating as depending more on the idiosyncrasies of the rater than on their own behavior in the classroom. As a result, teachers saw nothing to be gained from evaluation.

### Statement of the Problem

The emphasis placed upon "accountability" by the public during the decade of the 1960's had intensified the search by school districts in the 1970's to find improved ways to evaluate teacher effectiveness. The problem inherent in the search for improved ways to evaluate teacher effectiveness was that of selecting suitable criteria upon which both administrators and teachers agreed which truly measured teacher effectiveness. School districts, facing this problem needed to know how much agreement existed between teachers and administrators on effective criteria and to determine what kinds of criteria were appropriate for judging the effectiveness of teachers. If it was determined that administrators and teachers varied greatly in their views of the perceived importance of criteria for measuring teacher effectiveness, then continuation of the evaluation process would have resulted in increased sensitivity and mistrust on the part of teachers toward administrators who judged teaching effectiveness. It was necessary for school district administrators to include the teacher in determining the criteria used to evaluate their own effectiveness.

### Need for the Study

It is evident from a review of the literature that the task of identifying effective teachers and effective teaching is crucial to teacher education, teacher selection, teacher performance, and ulti-

mately, to the survival of society. Crucial as this need is and in view of the enormous amount of research directed at identifying effective teaching, it is disturbing to note that there has been no general agreement upon what constitutes effective teaching, or standards of teaching effectiveness. A substantial amount of pressure has been placed upon school districts to evaluate teaching effectiveness because of the accountability impact. This practice is a sensitive issue to teachers. Research seems to bear out the fact that teachers should indeed be concerned. Very evident in the research reviewed is the need to involve more than just the administrator or supervisor in evaluation of teaching effectiveness. The literature pointed out the fact that teachers should be involved in the evaluation process.

#### Purpose of the Study

The purpose of this study was to determine what factors were important from the teacher's viewpoint in identifying effective teaching, compare the findings with the administrator's viewpoint and determine whether or not there was agreement on the criteria for judging effective teaching. By first determining whether or not teachers and administrators agreed upon the criteria for judging effective teaching, this study provided a means whereby some conclusions could be reached by school districts concerning their efficiency in meeting the public demand of the best possible education for the tax dollar.

Questions to be Answered

The questions to be answered by this study were:

1. Is there agreement among teachers in Montana on the criteria that describes the effective teacher?
2. Is there agreement among school administrators in Montana on the criteria that describes the effective teacher?
3. What is the degree of agreement between administrators and teachers in Montana schools on the criteria that describes the effective teacher?
4. What is the degree of agreement between elementary and secondary teachers on the criteria that describes the effective teacher?
5. Is there a relationship between the criteria differences as perceived by teachers and their administrators and the rated effectiveness of the administrator in helping the teacher to improve his effectiveness?

Limitations of the Study

The limitations of this study were:

1. The study was limited to the geographic area of the State of Montana.
2. The elementary and secondary teacher population of Montana schools comprised the teacher population from which the sample was drawn.
3. The district superintendents and principals of Montana schools comprised the administrator population from which a sample was drawn.

### Definition of Terms

Terms defined for the purpose of this study were:

1. Teacher. A person who is certificated to teach in Montana and who will be under contract to teach during the 1976-1977 school year in any school district in Montana.

2. Administrator. A person who is certificated by the State of Montana for the purpose of administrating a school or school district and who is employed either as a principal or superintendent in any school or school district in Montana during the 1976-1977 school year.

3. Teacher Effectiveness. This term is the degree of success a teacher achieves in attaining the desired outcomes that a school district wishes to obtain the teaching-learning environment.

4. Evaluative Criteria. Evaluative criteria are measures of teacher effectiveness.

5. Types of Criteria. Criteria are typed in accordance with Mitzel's Scheme of process criteria, product criteria, and presage criteria.

Process criteria are measures of teacher effectiveness based upon classroom behavior either the teacher behavior, his students' behavior or the interplay of both.

Product criteria are measures of teacher effectiveness in terms of measurable change in student behavior as a product of teaching.

Presage criteria are evident measures of teacher effective-

ness based upon a teacher's personality or intellectual attributes, performance in training, years of experience, tenure, etc.

### Summary

Defining the effective teacher has been a continuing process carried out over many decades by researchers and investigators. A review of the literature indicated that the purpose of most completed research has been to improve teaching performance in order to provide better education for children.

The process of identifying the effective teacher in earlier times depended upon a subjective evaluation of the teacher's personality traits and behavior in light of some particular authority's judgment as to what was acceptable or unacceptable. This process was usually accomplished by the use of some type of rating instrument.

In more recent years the use of the rating instrument received severe criticism by both teachers and administrators because both believed its primary use by evaluators was for the purpose of dismissing teachers. As a result of the criticism directed at the use of the rating instrument, the emphasis in evaluation shifted from the subjective approach to a more objective approach which resulted in the positive practice of identifying the strengths and weaknesses of teachers. The purpose of evaluation became that of correcting weaknesses and reinforcing strengths of the teacher. The emphasis became one of

measuring teacher effectiveness centering on product measurement through previously agreed upon objectives of instruction. To evaluate a teacher's performance, it became necessary for school administrators and teachers to determine the characteristics of the effective teacher and the ingredients of effective instruction. The problem for administrators and teachers was that of agreeing upon the criterion measures of effective teaching.

Traditionally the evaluation of a teacher's effectiveness was conducted by the teacher's immediate supervisor, usually the principal. Other forms of teacher evaluation which emerged more recently included peer evaluation, self-evaluation, pupil evaluation or combinations of these.

There was, by no means, total agreement among school districts of the nation that any or all of the newer trends in the evaluation process contained total answers to the teacher effectiveness problem. One of the biggest problems encountered, regardless of the approach a school district followed in evaluation, was that of defining the criteria by which teaching and teachers were to be assessed.

## CHAPTER II

### REVIEW OF RELATED LITERATURE

#### Introduction

This review was organized into four elements of studies and research relating to development and change that has taken place in determining teacher effectiveness. The initial section concerns the forces at work which created the need to evaluate teacher effectiveness. The second portion relates the status of present appraisal methods used in teacher evaluation. The third section reviews experimental studies of teacher effectiveness that illustrate the problems inherent in a study of this nature. Some specific trends, criterion and design models were reviewed and described. Areas of emphasis in the review of literature includes:

1. Status of present methods of evaluating teacher performance.
2. Studies of teacher performance.

For the purpose of statistical comparison, this study followed closely a study reported in 1974 by Jenkins and Bausell on how teachers view the effective teacher (Jenkins and Bausell, 1974). The purpose of their study was to consult teachers and administrators regarding their views on teacher effectiveness, in particular, on criteria they used to evaluate their own effectiveness. To provide some structure for such an inquiry, Jenkins and Bausell developed a survey instrument which was based on the category labels of product, process and presage employed by



Harold Mitzel in his contribution to the 1960 edition of the Encyclopedia of Educational Research. A more elaborate description of Mitzel's categories of teacher effectiveness criteria appear later in this chapter and a copy of the instrument used by Jenkins and Bausell appears in Appendix. A.

Briefly described product criteria in Mitzel's scheme are employed where a teacher is judged on the basis of a measurable change in what is viewed as his product, student behavior. Process criteria is used when a teacher's evaluation is judged by either his behavior in the classroom or that of his pupils or the interplay of both teacher/student behavior. Presage criteria is used if a teacher's evaluation is judged in terms of the teacher's personal or intellectual attributes, his performance in training, his knowledge or achievement or other pre-service characteristics (Mitzel, 1960).

Jenkins and Bausell administered a survey instrument which included an assortment of product, process and presage criteria to a random sample of all public school teachers and administrators in the State of Delaware. Respondents who numbered two hundred sixty-four (N = 264) were instructed to assume that adequate measures were available to measure each of the criteria listed. The instructions listed were replicated for this study as well as the continuum used for responses. This information appears in the instrument which is located in the Appendix. A.

The criteria and the ratings given them by Delaware teachers and administrators appear in Table XXI. When the responses of elementary teachers, middle school teachers, secondary teachers and principals were compared, the results indicated that although these groups might be expected to have different biases, their ratings were remarkably similar. The average correlation between these groups was .93 (Jenkins and Bausell, 1974).

Perhaps the most revealing aspect of the survey according to Jenkins and Bausell was the rating given to the criterion, Amount Students Learn. This criterion in the Delaware study was not seen as particularly important in judging teacher effectiveness relative to the other criteria rated. The implication of the rating received by Amount Students Learn for accountability proponents should be obvious. While those in the accountability movement stressed student learning as the primary basis for educational decision making, educational practitioners, at the same time, affirmed their preference for other criteria as indicated by the Delaware study. Results of their study and comparison tables with the Delaware study are listed in Chapter IV (Jenkins and Bausell, 1974).

#### Need for Evaluating Teacher Effectiveness

The study of the effectiveness of a teacher and particularly how teachers themselves viewed the effective teacher gained increased momentum in recent years with the emphasis placed upon "accountability" by the soaring cost of education in the 1960's. The cry for accountability intensified the search in the 1970's for improved ways to evaluate and to

standardize these procedures. Impetus for this search arose from the needs of two groups: teachers, on the one hand, who sought the security of fair objective standards of evaluation; and the public, on the other hand, who sought assurance that its tax dollar was well spent (Oldham, 1974). Because teacher accountability remained the center of debate, any discussion of the topic turned sooner or later to the issue of teacher effectiveness. For the teacher the idea of accountability quickly translated into an assessment of the quality of his instruction and the related necessity of selecting a criteria by which one would judge his effort. Because the accountability movement centered on teacher effects, it seemed imperative that teachers be consulted regarding their views on teacher effectiveness, and particularly upon the criteria they used to evaluate their own effectiveness (Jenkins and Bausell, 1974).

#### Forces Which Created the Need to Evaluate Teacher Effectiveness

In addressing the National Association of Secondary School Principals' annual convention in Anaheim, California in March of 1972, Governor Ronald Reagan who was then the Governor of California referred to the growing need of public education to become more accountable in the decade of the seventies. To "re-establish the public's confidence in education and our school system," Governor Reagan described as yet another responsibility the public had given to its education system.

Governor Reagan described the public's eroding confidence in education although education traditionally had been America's major public priority. Of the reasons described by him as contributing to the eroding public confidence, crisis stemming from financial problems and the feeling that people had reached the limit of their ability to pay higher taxes seemed most paramount. How this mood affected education was described by Governor Reagan in his statement,

However unjustified educators feel the attitude may be, there is a feeling among our people that our schools are not doing all that they should, or doing it as efficiently and as economically as they could (Reagan, May 1972).

The implications that Governor Reagan's address held for measuring teacher effectiveness as one way to meet the public's demand for accountability is summarized in part of his address as follows:

We must develop ways to evaluate objectively the performance of teachers, to find the best, and to reward them for superior performance. In California last year, we passed legislation to require evaluation of teacher performance.

You can probably guess the result. The deadline for conforming to this new law had to be postponed. Because we have promoted by seniority alone for so long, we have had to start from the basics to determine just what should be measured in evaluating teachers and how to measure it.

. . . However difficult it may be, we are determined to develop fair, realistic and reasonably flexible methods of measuring teacher performance (Reagan, May 1972).

Herman supported the primary reason expressed by Governor Reagan that the education institutions of this nation were besieged by internal

and external forces demanding that these institutions be held accountable and show evidence of having used the taxpayers money wisely before asking for additional money. Herman took the position that one of the most basic elements in accountability was staff evaluation. This element, according to Herman, dealt with definitions of what we were doing, who was responsible for doing it, and how did we measure the effectiveness of the work assigned each individual within the program. Herman described two basic ideas that needed to be included within each district's plan of evaluation regardless of the ultimate number of personnel involved in the evaluation. These were described as: (1) a self-evaluation must be done by the employee and (2) the employee's immediate supervisor has to arrive at judgments based upon his evaluations when administrative decisions, such as whether or not to grant a teacher tenure, needed to be made (Herman, 1973).

Ornstein and Talmage have stated that the concept of accountability was borrowed from management. They described the concept of accountability applied to education as

. . . "holding some people (teachers or administrators), some agency (board of education or state department of education), or some organization (professional organization or private company) responsible for performing according to agreed-upon terms" (Ornstein and Talmage, 1974).

In the past, so stated Ornstein and Talmage, students alone were held accountable for specific objectives in terms of student changes in achievement and behavior. According to Ornstein and Talmage most people

believe that everyone, including teachers and administrators, should be held accountable for their work. What many educators objected to, and even feared, was the oversimplified idea of accountability as the sole responsibility of the teacher or principal. Accountability should have included not only teachers and administrators but also parents and community residents, school board members and taxpayers, government officials and business representatives, and most importantly the students. Ornstein and Talmage summarized their concern about the concept of accountability as an idea which was spreading throughout the country regardless of the fact that there was no evidence that it would reform the schools. One of the major difficulties which seemed to plague the accountability movement was that of measuring learning (Ornstein and Talmage, March 1974).

One process which the call for accountability in the seventies forced upon some school districts was termed management by objectives (MBO). A number of school districts turned to this new concept of management, alternately referred to as management by mission, goals management, and results management in the hope that because the concept had been used successfully in business and industry for more than a decade, it would likewise prove successful for school districts.

Although MBO and accountability have been frequently teamed in the literature and in school district improvement efforts, they have not been considered as generic teammates. MBO preceded the accountability-

in-education movement by at least a decade. The term management by objectives was first used by Drucker in his book Practice of Management in 1954. McGregor of M.I.T. and Likert of the University of Michigan had used it to justify the application of findings in behavioral research to the business situation. Since then, results management has been widely installed throughout the United States and other countries, notably Great Britain, where business, industry, and government have found it a productive way of managing their enterprises (Read, March 1974).

Read listed several administrative practices which he felt MBO would strengthen. Read stated that successful implementation of MBO would . . . "eliminate the tendency to evaluate personnel in terms of their personality traits; substituting instead, their performance in terms of results" (Read, 1974). For the purpose of this paper the practice of determining teacher effectiveness seemed most appropriate.

As pointed out by Howard one should not be led to believe that

the idea of accountability, the adoption of business practices in education, is new, that it has just been discovered by some of our brighter, abler, and more responsible people in education (Howard, 1974).

He noted that in the early 1900's we were blessed in having a number of educators who, in response to pressures from business, industry, and the general public, were able to devise methods for determining efficiency and educational output in the schools. "Educational efficiency

experts" and "educational engineers" were names given to the responding educators of those days (Howard, 1974).

Miller described the influence of business practice on accountability by noting that developments in the field of management techniques required sharper expertise in goal setting, planning, and establishing of cost effectiveness measures. Management also increased its skill in evaluation and assessment which in turn fostered the move toward accountability, Miller concluded (Miller, 1972).

Miller perceived accountability as a means of holding an individual or group responsible for a level of performance or accomplishment for specific pupils. He emphasized that program goals would be developed for each activity, thus clarifying the purposes and goals of all programs and making it easier to assess results. He believed that educators would have to develop greater skill in goal setting, diagnosing needs, and analyzing learning problems. He also noted that increased emphasis on improved communication and involvement of pupils and parents would be a necessity and would result in better understanding and support of the school program (Miller, 1972).

Many persons were threatened by the idea of accountability and even more were disturbed by the apparent way in which the concept was being implemented. A major cry from the teachers was that standards for them and for the pupils were likely to be set by central office administrators. They feared that the required levels of performance would be



unrealistic and unobtainable, thus triggering punitive actions toward pupils and teachers. Teachers did not want to become the scapegoats when the school systems did not produce what the parents, the boards, or the administrators demanded. Teachers pointed out that while they were likely to be the ones held accountable, they often did not have the resources or power to alter policies or practices which must be changed if improvement were to come about.

Many worried that implementation of accountability would cause education to focus on that which could be easily identified and measured. The area of academic achievement would most likely get the most attention at the expense of the affective domain. What was certain in the minds of many was that accountability would surely increase the educational bureaucracy which, to some, already constituted a serious impediment to improving instruction.

As described by Kibler, the use of instructional objectives was consistent with the concept of accountability which was described as the balancing of money spent for education with the amount students learned. Accountability in education as described by some writers was rapidly gaining acceptance from both the public and the federal government. Unfortunately, some educators who had negative attitudes about accountability also had become negative about instructional objectives (Kibler, 1974). Apparently, the negative attitudes about instructional objectives were based on the misconception that using instructional objectives lead

to accountability in education. Kibler felt that few comforting words could be said to those teachers who viewed accountability-based educational systems as a threat. If accountability-based educational systems did become the norm, experience in the use of instructional objectives would enable teachers to adapt to the system more easily (Kibler, 1974).

Hottleman, Director of Educational Services, of the Massachusetts Teachers Association described the negative impact that accountability had on public education in his statement that,

The accountability movement probably offers more potential for harm to public education than any other idea ever introduced, yet more and more highly placed education officials hop on the bandwagon daily. One common element among the major accountability movers is their backgrounds. They are mostly administrators, testing experts, or private businessmen. Teachers' organizations and individual teachers are notably absent (Hottleman, 1974).

Hottleman described the accountability movement in public education as first becoming visible in 1970 when President Nixon announced, "School administrators and school teachers are responsible for their performance and it is in their interest as well as in the interests of their pupils that they be held accountable." Hottleman explained that the President was probably influenced by Leon Lessinger, the Assistant Commissioner of Education, who openly stated his intention to make public education accountable (Hottleman, 1974).

In summarizing the accountability movement Hottleman noted that the accountability movement had not begun as a way to improve learning

opportunities for children but in response to problems which arose out of the increasing costs of public education. The proponents, in the main, were not public educators but were those who had an accounting mentality that viewed sorting, classifying and measuring as significant per se. Hottleman viewed the overemphasis on measurement as promising greater conformity, the diminishing of humaneness, individuality, and creativeness in public education, and if unchecked, threatened a concerted move toward educational mechanization. In the opinion of Hottleman teachers were viewed as the least important resource in seeking answers about the improvement of education. In his view what was needed was a reduction of funds spent by the measurement fanatics and an increase in funds spent in finding ways of surfacing and implementing the ideas of practicing teachers (Hottleman, January 1974).

Weiss described educational accountability as a threat to the privacy and security of educators who worked in greater privacy than almost any other professional group. Because educators worked in relative privacy compared to other professional groups they looked upon the concept of accountability as having strong implications of distrust for their effectiveness. As Weiss stated, "Educators, like all of us, know that 'accountability' does not enter into the discussions between persons or agencies with great confidence in each other," therefore, educational accountability carried with it an obvious presumption of guilt. (Weiss, April 1973).

Weiss's observation described in some degree the defensiveness that educators displayed toward the concept of accountability. While most people believed that everyone, including teachers and administrators, should be held accountable for their work, educators objected to, and feared, the oversimplified idea that accountability was the sole responsibility of the teacher or principal.

The response of teachers to a state demand for accountability and assessment was described in the research carried out by Bleecher. In the State of Michigan a demand for accountability and assessment was seen as a rational response to political pressures from taxpayers who felt heavily taxed. Because taxpayers wanted to know what they were getting for their annual two billion dollars spent on education, the legislature passed an act which ordered a program designed to assess pupil learning in the basic educational skills to take effect immediately. In order to comply with the educational assessment act, the State Department of Education advocated a six-step model which presumed would lead to educational accountability. The response of teachers was rejection in the form of minimal compliance and by pressure from the organized teacher groups (Bleecher, December 1975).

The accountability movement which received renewed emphasis in education in the early part of this decade still continues. This movement has been summarized by Popham, as a public challenge to education by his statement that:

The public is clearly subjecting educational institutions to increased scrutiny. Citizens are not elated with their perceptions of the quality of education. They want dramatic improvements in the schools, and unless they get them, there is real doubt as to whether we can expect much increased financial support for our educational endeavors. And the public is in no mood to be assuaged by promises. 'Deliver the results,' we are being told. No longer will lofty language suffice, and yesterday's assurances that 'only we professionals know what we're doing' must seem laughable to today's informed layman.

The distressing fact is that we haven't produced very impressive results for the nation's children. There are too many future voters who can't read satisfactorily, can't reason respectably, don't care for learning in general, and are pretty well alienated from the larger adult society (Popham, May 1972).

Many educators particularly administrators responded to the accountability challenge which Popham described as inevitable and accepted the premise that the schools must indeed be accountable. Thus, according to Popham, "the course was set to find the most expedient way to accomplish accountability". For teachers the challenge could be described as an admission of guilt for failure of students to learn. Popham suggested that educators accept the accountability challenge by increasing classroom teachers' skills in producing evidence that their instruction yielded worthwhile results for learners. One way, suggested by Popham, of showing results was to place appropriate measures of student performance in the hands of the teacher. The measures suggested by Popham were tests of instructional objectives described in the literature as criterion-referenced measures (Popham, May 1972).

Measuring student performance by testing has been referred to in

the literature as product measurement which in turn has been one method used to measure teacher effectiveness. This method of measuring teacher effectiveness has not been popular with teachers for a number of reasons. The common problem is that attempts to evaluate teachers on the basis of pupil's test performance tend to focus teaching too narrowly on the specifics measured by the test (Rosenshine, 1970) and (Veldman and Brophy, 1974). Grogman described the dangers inherent in accountability measures that focus on short-term goals which are the kind measured by tests:

As teachers are threatened with accountability measures that focus on short-term measurable goals, their only recourse is to stress what is stressed in the accountability measures, frequently to the detriment of more important learnings, which may be underemphasized or overlooked. If not measured (and they generally are not in accountability systems), such skills as socialization, cooperation, and communication undoubtedly will suffer (Grogman, May 1972).

In recent years many who are charged with the responsibility of evaluating teachers have begun to consider product evaluation methods. Thus, trying to imitate industry, evaluation centered on student achievement, which in part depended upon test scores (Thomas, December 1974). Yet, Medley and others concluded from their research that only short-term goals which were almost certainly the least important goals of education are validly measured by tests. The validity of "teacher tests" of ability to achieve short-term outcomes as predictors of overall teacher effectiveness is by no means self-evident. "Their validity, according to Medley, predictor of over all teacher effectiveness, must be empiri-

cally demonstrated before their use is justified" (Medley, June 1975).

In view of the limitations placed upon tests of student achievement as a criterion to measure teacher effectiveness, it was not surprising to find that teachers questioned product measurement as a measure of their effectiveness. This position was summarized in the Fleischmann Report which was made to the New York State Commission on the quality, cost, and finance of elementary and secondary education for the State of New York.

The Report stated,

Because of the many circumstances that influence learning, educators have traditionally been reluctant to submit to evaluation on the basis of student performance. They have argued that learning is in too many ways beyond their control and that it is therefore unfair to judge school effectiveness by measuring student achievement alone" (The Fleischmann Report, 1973).

Not only were educators reluctant to be evaluated on the basis of student performance, they questioned any measure designed to assess their teaching effectiveness. Wolf described this concern of teachers in the following statements:

Teachers are not fond of evaluation. They suspect any measure designed to assess the quality of their teaching, and any appraisal usually arouses anxiety . . . . If teachers are to submit to an assessment of their performance, they would probably like reassurance that the criteria and method of evaluation that are used would produce credible results. . . . Teachers probably believe that the standards for evaluating what is effective teaching are too vague and ambiguous to be worth anything. They feel that current appraisal techniques fall short of collecting information that accurately characterizes their performance (Wolf, 1973).

Supporting this point of view House noted that little demand existed



















































































































































































































































































































