



The determination of the criteria of effective teaching as perceived by teachers and students in Bozeman Senior High School in Bozeman, Montana
by Samorn Sucharit

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

Montana State University

© Copyright by Samorn Sucharit (1976)

Abstract:

The public has expressed an interest and concern in connection with the increased costs of education. Consequently, the question has arisen to whether the input of public money into educational institutions is obtaining the desired results or outcomes. This question centers around the concept of educational accountability which has a direct relationship to teacher effectiveness.

There have been many proposals forthcoming concerning evaluating teaching effectiveness. Among these proposals are the following methods which have been suggested by various educators. For example, the critical incident technique, interaction analysis in the classroom, the rating method, the observation method, the standardized test of student achievement. Recently, supervision of objectives and performance objectives to evaluate teaching effectiveness were recommended.

Some educators stated that the components of teaching ability were knowledge of subject matter; and technique of teaching and personality. Some educators proposed using student gains as a criterion to measure teaching ability. However, other educators argued against using student gains as a basis for evaluating teaching effectiveness. Teaching method, using multiple evaluators, using multiple bases for evaluation, using product, process and presage criteria to evaluate teaching effectiveness also have been proposed.

This research was conducted to determine the perceptions of teachers and students concerning the product, process and presage as the criteria of effective teaching. A 34-item questionnaire was developed as an instrument to obtain the data needed by the researcher. The 34 items in the questionnaire included 9 product, 12 process and 13 presage criteria. The senior high school in Bozeman, Montana was selected as a main resource for this investigation. Forty-nine teachers and two hundred fifty-three students responded to the instrument.

The results of the investigation revealed that product, process and presage criteria were not perceived differently by the teachers and students. It was also found that the process criteria received the highest ranking as it related to the degree of effective teaching. In addition, it was found that there was a positive relationship between the perceptions of teachers and students concerning the ranking of the degree of importance of the 34 criteria. Finally, it was found that the teachers perceived the product criteria as the most difficult to measure, while the second and third most difficult criteria to measure were the process and the presage criteria respectively.

THE DETERMINATION OF THE CRITERIA OF EFFECTIVE TEACHING
AS PERCEIVED BY TEACHERS AND STUDENTS IN BOZEMAN,
SENIOR HIGH SCHOOL IN BOZEMAN, MONTANA

by

SAMORN SUCHARIT

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

of

DOCTOR OF EDUCATION

Approved:


Chairperson, Examining Committee


Head, Major Department


Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

September, 1976

ACKNOWLEDGEMENT

This research would not have appeared without the assistance of the following professors. The first was my major advisor, Dr. Robert J. Thibeault who worked with me and who had personal involvement in my research from the beginning through the end of my program. Dr. Earl N. Ringo, Dean of the College of Education, encouraged and contributed his help in solving problems since I first came to this University. Dr. Eric Strohmeier assisted in research design and methods. Dr. Gerald D. Sullivan contributed his help in reading and correcting my English as did Dr. Carol Parker. Dr. Robert Nickelson, the graduate representative helped in necessary corrections. All of the above participated in my examination committee. To all of them, I would like to express my appreciation, and thanks for their assistance in this very special occasion.

Other professors whom I wish to thank at this time are Dr. Henry L. Parsons, the Assistant Dean of the College of Graduate Studies, Dr. Roger W. Snow, Dr. Richard Landis, Dr. Albert Hopkins and Dr. Al Suvak, who contributed their friendly help and sympathetic attitudes while I studied at Montana State University and at Eastern Montana College. Dr. Alfred Wilson helped in proper writing in research. Mr. Milton Negus, the superintendent, Mr. Gappmayer, the principal, Mrs. Kenny, the teachers and the students who contributed valuable help to me, and I am greatly appreciative.

Finally, I would like to express my appreciation to my wife, sons, daughters, relatives and friends, especially to Tom Everett who helped to read and correct English in the original copy, Vanita Singh, Cynthia Ware, and Glenda to correct spelling and English, for all their encouragements, help and patience given to make my education a success.

Samorn Sucharit

September, 1976

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Statement of the Problem	4
Need for the Study	5
General Questions to be Answered	7
General Procedures	8
Limitations.	8
Definition of Terms.	9
Summary.	11
II. REVIEW OF RELATED RESEARCH AND LITERATURE.	13
Demands for Educational Accountability	13
Development of Teacher Evaluation from 1912 to 1956.	14
Development of Teacher Evaluation from 1957 to 1969.	26
Development of Teacher Evaluation from 1970 to 1976.	37
Weaknesses in Evaluating the Outcomes.	39
Argument for Student Gains	40
New Approaches to Evaluating Teaching Effectiveness.	41
Summary.	43
III. PROCEDURES	44
Population Description and Sampling Procedures	44
Description of Investigation Categories.	45
Methods of Collecting Data	45
Methods of Organizing Data	46
Statistical Hypotheses	46
Analysis of Data	47
Precaution Taken for Accuracy	48
Summary.	48
IV. ANALYSIS OF DATA	50
Methods of Sampling and Number Sampled	50
Characteristics of the Sample.	51
Treatment of Data.	56
Presentation of Data	56
χ^2 Test of Independence Between Student and Teacher Population on Product, Process and Presage Criteria.	57
χ^2 Test of the Teachers' and Students' Perceptions on Items Dealing with Product Criteria.	60

Chapter	Page
χ^2 Test of Independence Between Student and Teacher Population on Items Dealing with Process Criteria. . .	66
χ^2 Test of Independence of Teacher and Student Populations on Items Dealing with Presage Criteria. . .	72
Comparison of Product, Process and Presage Criteria by Numbers of Significant and Not Significant Difference. . .	79
Comparison of Product, Process and Presage Criteria of Level of Significance.	80
The Significance of Product, Process and Presage Criteria by Degree of Importance.	81
Test of Difficulty	84
Test of Relationship	86
Summary.	87
 V. FINDINGS, CONCLUSIONS, RECOMMENDATIONS	 90
Findings	90
Some Indices of a Demand for Educational Accountability. . .	90
Some of the Milestones in the Development of Teacher Evaluation	91
New Approaches to Evaluating Teaching Effectiveness. . .	93
Demographic Findings	94
The Chi Square Test of Independence Between the Teacher and Student Populations on Items Dealing with Product, Process and Presage.	95
Test of Relationship	96
Test of Difficulty	96
Conclusions.	97
Recommendations.	98
 APPENDICES	 100
APPENDIX A: Questionnaire for Teachers.	101
APPENDIX B: Questionnaire for Students.	108
APPENDIX C: Comments of Teachers and Students Individually by Sex	113
APPENDIX D: Form Letter Sent to Teachers.	129
APPENDIX E: Form Letter Sent to Students.	130
APPENDIX F: Follow-up Notes From Dr. Robert J. Thibeault	131
 BIBLIOGRAPHY	 132

LIST OF TABLES

Table		Page
1	Distribution of Total Sample Population of the Students and Percentage of Returns of the Questionnaires by Grade and Sex.	51
2	Distribution of Total Sample Population of Teachers and Percentage of Returns of the Questionnaires by Sex.	53
3	Distribution of Teachers Who Returned the Questionnaires by Age.	53
4	Distribution of Years of Teaching Experience of Teachers.	55
5	χ^2 Test of Independence Between Student and Teacher Populations on Items Dealing with Product Criteria	59
6	χ^2 Test of Independence Between Student and Teacher Population on Item Dealing with Process Criteria	67
7	χ^2 Test of Independence Between Student and Teacher Populations on Item Dealing with Presage Criteria	73
8	Comparison of Product, Process and Presage Criteria by Numbers of Significant Difference and Not Significant Difference	79
9	The Comparison of the Product, Process and Presage Criteria by Ranking of the Level of Significance and the Degree of Freedom.	81
10	The Comparison of Product, Process and Presage Criteria on the Degree of Importance that were Highly Perceived by the Teachers and the Students.	82
11	The Grouping of Means with Respect to the Product, Process and Presage Criteria	85

Table	Page
12	The Ranking of Means of the Teachers' and the Students' Concerning the Degree of Effective Teaching 88

ABSTRACT

The public has expressed an interest and concern in connection with the increased costs of education. Consequently, the question has arisen to whether the input of public money into educational institutions is obtaining the desired results or outcomes. This question centers around the concept of educational accountability which has a direct relationship to teacher effectiveness.

There have been many proposals forthcoming concerning evaluating teaching effectiveness. Among these proposals are the following methods which have been suggested by various educators. For example, the critical incident technique, interaction analysis in the classroom, the rating method, the observation method, the standardized test of student achievement. Recently, supervision of objectives and performance objectives to evaluate teaching effectiveness were recommended. Some educators stated that the components of teaching ability were knowledge of subject matter; and technique of teaching and personality. Some educators proposed using student gains as a criterion to measure teaching ability. However, other educators argued against using student gains as a basis for evaluating teaching effectiveness. Teaching method, using multiple evaluators, using multiple bases for evaluation, using product, process and presage criteria to evaluate teaching effectiveness also have been proposed.

This research was conducted to determine the perceptions of teachers and students concerning the product, process and presage as the criteria of effective teaching. A 34-item questionnaire was developed as an instrument to obtain the data needed by the researcher. The 34 items in the questionnaire included 9 product, 12 process and 13 presage criteria. The senior high school in Bozeman, Montana was selected as a main resource for this investigation. Forty-nine teachers and two hundred fifty-three students responded to the instrument.

The results of the investigation revealed that product, process and presage criteria were not perceived differently by the teachers and students. It was also found that the process criteria received the highest ranking as it related to the degree of effective teaching. In addition, it was found that there was a positive relationship between the perceptions of teachers and students concerning the ranking of the degree of importance of the 34 criteria. Finally, it was found that the teachers perceived the product criteria as the most difficult to measure, while the second and third most difficult criteria to measure were the process and the presage criteria respectively.

Chapter 1

INTRODUCTION

Because of the increasing costs of education and demand for educational accountability, the public has questioned whether the input of money into educational institutions has obtained the desired outcomes. In any discussion of the topic of educational accountability, the issue of teacher effectiveness maintains a prominent position. (Coleman 1973:3)

Concurrently, educational administrators have sought more viable evaluation procedures to objectively judge matters of promotion, release, or renewal of teachers' contracts, while teachers have asked to provide input into these judgmental procedures based on their perception of what constitutes effective teaching. Another aspect of educational accountability involves student assessment of teaching effectiveness. Jenkins and Bausell (1974:572) noted that:

For the teacher, the notion of accountability quickly translates into an assessment of the quality of his instruction and the concomitant selection of criteria by which one will judge his effort. Since the accountability movement centers on teacher effects, it seems only just to consult teachers regarding their views on teacher effectiveness, in particular, on criteria they use to evaluate their own effectiveness. We suspect that discrepancies in conceptions of teacher effectiveness may be at the root of the strong feelings raised by the accountability issue. Uncovering these conceptions may serve to modify the approach taken by accountability advocates.

According to Jenkins (1974:572), Harold Mitzel in his contribution to the 1960 edition of the Encyclopedia of Education Research,

categorized the measures utilized to evaluate teaching effectiveness into three criteria labeled product, process and presage. (1) Product criteria, when teachers are judged by their effectiveness in changing student behavior, the judge is employing product criteria. The teacher is judged on the basis of a measurable change, in what is viewed as his product, student behavior. It can be assumed that measures of growth in skills, knowledge of subject matter, and attitude which can be logically or empirically attributed to the teacher's influence constitute acceptable data in the product category. For gains in knowledge of subject matter one might use such measures as standardized achievement test, end-of-lesson, or unit quizzes. Student performances which could be taken as indicators of self-acceptance, of attitude toward school subjects or toward learning in general, and of respect for others and their opinions might qualify as effective goals and thus also fall within the product category. If there is some confusion about the product category, it probably arises not so much from the notion of using student change as a criterion as from the difficulty in gaining consensus of what products are the appropriate domain of the school.

(2) Process criteria, when teacher evaluation is based upon classroom behavior, either the teacher's behavior, his students' behavior or the interplay of teacher/student behavior, the judge is using process criteria. The process behaviors chosen to measure are

believed to be worthwhile in their own right and thus are not necessarily related to product criteria. Some variables upon which teachers could be rated are their verbal behavior, methods, classroom control and individualization of instruction. Students might be rated for their verbal behavior, attentiveness, and conformity to classroom routine. Teacher/student interaction might be judged for rapport and climate. (3) Presage criteria, when teacher evaluation is based upon one's personality or intellectual attributes such as industry, adaptability, intelligence, character, and his performance in training, his knowledge of achievement such as marks in education courses, success in student teaching, national teacher examination, knowledge of educational facts or his inservice status characteristics such as tenure, years of experience or participation in professional organizations, the judge is employing presage criteria.

According to Ellena (1961:foreword), it is difficult to determine who is a good teacher and upon what basis judgments are formed. These are familiar questions in every part of the country. Probably no aspect of public education has been discussed with greater frequency and with deeper concern by both lay citizens and educators than teacher competence such as how to define it, how to measure it, how to reward it, how to detect and remove obstacles to its achievement. Ellena (1961:foreword) further stated that:

From New York to Hawaii, from Texas to Alaska, people are seriously asking that classroom teachers, school

administrators, and members of boards of education begin to study teacher competence. Responding to this insistent request, three associations, the American Association of School Administrators, the Department of Classroom Teachers of the National Education Association, and the National School Boards Association, have joined together to study the cluster of complex questions about teacher competence -- questions which are at the very heart of educational policy, administrative leadership, and the day-to-day instruction in the more than a million classrooms in this country.

The present study attempted to determine teachers' and students' perceptions of the criteria which should be used to assess teaching effectiveness.

Statement of the Problem

Due to the demands for educational accountability and the lack of definite studies in this area and in an effort to consider teachers' and students' perceptions in matters concerning the evaluation of effective teaching, the problem of this study was to survey the attitudes of teachers and students in a senior high school and thus provide school administrators with indices of effective teaching that coincide with teacher and student perceptions. There were three sub-categories of the main problem in this study. First, what was the degree of importance of the criteria as indices of effective teaching; second, what was the degree of difficulty in measuring the characteristics of effective teaching; and third, what relative value should be placed on characteristics of product, process, and presage criteria?

Need for the Study

In addition to reporting the result of this study to school officials, and adding another dimension to the body of information about assessing teacher effectiveness, the researcher felt a personal need for information on the perceived characteristics of effective teaching to improve teacher training programs in his native country, Thailand. Prior to beginning graduate studies in the United States of America, the researcher taught in a teacher training college in Bangkok, Thailand. It was hoped that this study would provide insights and information to help him and other Thailand college instructors, to institute changes in teacher training programs. McNeil (1971:35) indicated the need for further study as follows:

More than one-half of the nation's teachers report no confidence in their school system's program of teacher evaluation. Student teachers are unhappy with the vague criteria for assessing their qualifications for entry into the profession.

Thomas (1971:1) supported the need for the study of effective instruction. He stated:

The task of identifying effective teachers, or effective teaching is crucial to teacher education, to teacher selection, to teacher performance, and ultimately to the survival of the human society. Yet today no general agreement exists as to what constitutes effective teaching, and no standards of teacher effectiveness are commonly agreed upon.

According to Mitzel (1960:1485), upon the assumption that the

most appropriate criteria of teacher effectiveness are those which have relevance to significant education outcomes, it is critical as to which outcomes are selected. Critics of present-day education like Pestor, Hutchins and Riesman have exhorted educators to pay more attention to the development of students' intellectual powers. It is clear that leaning toward one or the other philosophical position regarding the appropriate goals of education will undoubtedly influence the selection of criterion measures of teacher effectiveness. We need much precise, painstaking research in teacher effectiveness oriented toward a variety of educational situations. We need research in field situations (functioning classrooms) with massive samples of teachers and students. We need research in laboratory situations, as Robinowitz and Travers suggest, with small samples and careful control over experimental learning conditions. Perhaps, most of all, we need a comprehensive theory of teacher behavior and learning to channel the research efforts that undoubtedly will be undertaken. A contemporary research effort relative to teacher effectiveness compared with that engaged in forty years ago suggests that little progress has been made toward theory formulation. Although some encouraging beginnings have been made in conducting teacher-competence research within the framework of an explicitly developed theory, none of the research in this area has made maximum use of unifying theoretical conceptions from

learning, group dynamics, psychotherapy or cultural anthropology. Ryans' discussion of the importance of making continued attempts to develop a research guided theory seems well taken.

General Questions to be Answered

General questions to be answered in this study:

1. What were some of the indices of a demand for educational accountability?
2. What were some of the milestones in the development of teacher evaluation during the last sixty years?
3. What were some of the newer approaches and concepts of teacher evaluation?
4. What were the major strengths and weaknesses of evaluation by outcomes?
5. According to teacher perceptions, to what degree were these characteristics effectively measured?
6. What relative value should be placed on characteristics of product, process and presage criteria?
7. Did student perceptions of effective teaching significantly differ from teacher perceptions?
8. What was the degree of importance of each criteria as an index of effective teaching?

General Procedures

The general procedures to be followed in this study were:

1. To conduct a thorough review of literature and related research as it pertained to evaluating teacher effectiveness.
2. To develop a survey questionnaire to be administered to the entire teaching staff of Bozeman Senior High School and to a random sampling of the student body of that school in order to determine teacher and student perceptions of the characteristics of effective teaching.
3. To report the results of this survey in tabular and narrative form.
4. To statistically compare the perceived relative values of utilizing these criteria.
5. To statistically compare teacher and student perceptions of the characteristics of effective teaching.
6. To determine degree of difficulty of measuring the criteria.

Limitations

This study was limited in the following ways:

1. The majority of the sources considered in the review of related literature was from the Montana State University Library and ERIC inter-library loan resources.

2. The survey was limited to the teaching staff and student body of one large secondary school and therefore generalities derived from the results of the survey had limited applications to other school systems. However, it was assumed that the data gained from this study might be added to other studies on evaluating teaching effectiveness.

Definition of Terms

For the purpose of this thesis, certain key concepts are used in the following contextual settings:

Accountability. The theory that teachers and school systems may be held responsible for actual improvement in pupil achievement and that such improvement is measurable through tests of teacher effectiveness constructed by outside agencies. (Good 1973:5-6)

Behavioral objectives. The aims or objectives of education stated as actual performance criteria or as observable descriptions of measurable behavior. (Good 1973:393)

Critical thinking. Thinking that proceeds on the basis of careful evaluation of premises and evidences and comes to conclusions as objectively as possible through the consideration of all pertinent factors and the use of valid procedures from logic. (Good 1973:608)

Effective teaching. Use of a plan for instruction or presentation which causes a desired change in the learners' behavior. (Good 1973:589)

Evaluation criteria. The standards against which a person, a group, a procedure, or an instrument may be checked. (Good 1973:220)

Evaluation instrument. Any of the means by which one obtains information on the progress of the learner and the effectiveness of instruction; quantitative and qualitative data, objective measures, subjective impressions, tests, observation, anecdotal records, case studies and sociometric methods may all serve as instruments for deciding whether instrumental objectives have been attained. (Good 1973:221)

Hierarchy. Any graded organization, whether mental, physical or social, in which each rank (except the highest) is subordinate to the ranks above. (Good 1973:280)

Product criteria. When teachers are judged by their effectiveness in changing student behavior, the judge is employing product criteria. The teacher is judged on the basis of a measurable change in what is viewed as his product, student behavior. What constitutes acceptable products, or changes, has never been made altogether clear. But it would seem that measures of growth in skills, knowledge of subject matter and attitude which could be logically or empirically attributed to the teacher's influence constitute acceptable data in the product category. (Jenkins and Bausell 1974:572)

Process criteria. When teacher evaluation is based upon classroom behavior, either the teacher's behavior, his student's behavior,

or the interplay of teacher/student behavior, the judge is using process criteria. The process behaviors chosen to measure are believed to be worthwhile in their own right and thus are not necessarily related to product criteria. Some variables upon which teachers could be rated are their verbal behavior, methods, classroom control, and individualization of instruction. (Jenkins and Bausell 1974:572)

Presage criteria. When teacher evaluation is based upon one's personality or intellectual attributes (industry, adaptability, intelligence, character), his performance in training, his knowledge of achievement (e.g. marks in education courses, success in student teaching, national teacher examination, knowledge of education facts) or his inservice status characteristics (e.g. tenure, years of experience, or participation in professional organizations), the judge is employing presage criteria. (Jenkins and Bausell 1974:572)

Teacher merit system. A plan by which promotion, increase in pay and general advancement within a school system are determined by the degree of efficiency with which the teachers perform their duties; may be combined with other plans, such as experiences or training evaluation, in arriving at salary increases or promotion. (Good 1973:363)

Summary

Due to the increased costs of education, the public has expressed

a concern for and increased effort in educational accountability. In any discussion of the topic of educational accountability, the issue of teacher effectiveness maintains a prominent position. School administrators have sought more viable evaluation procedures while students and teachers have asked to provide input into evaluation procedures based on their perceptions of effective teaching. The matter of assessing teacher effectiveness is a complex issue and necessitates considering the criteria relative to effective teaching.

Due to the demands for educational accountability and in an effort to consider teachers' and students' perceptions in matters concerning evaluation, the problem of this study was to survey teacher and student attitudes in a senior high school and thus provide school administrators with indices of effective teaching.

In addition to adding another dimension to the bank of information concerning teaching effectiveness, the researcher hoped to utilize understandings gained from the results of this study to improve teacher training in his native country, Thailand.

The first procedure in this study was to conduct a review of literature and related research concerning teacher evaluation. This will be found in chapter two.

Chapter 2

REVIEW OF RELATED RESEARCH AND LITERATURE

This chapter will review the methods of evaluation of instruction that have been employed. The researcher believed that the findings would provide suggestions as to the best method or instrument to measure and evaluate effective teaching. The materials are arranged chronologically to make it convenient for the reader.

Demands for Educational Accountability

Because of increased costs of education, serious questions concerning public education have been raised. It has been asked if the increased educational cost has resulted in increased productivity. Burrup (1974:55) asked two questions involving public expenditures in education as follows:

The spiraling cost of education and the changing social climate of the country have combined to raise serious questions concerning public education. Have the increased costs resulted in proportionately increased productivity? Has education, with an increase of 1,000 percent in expenditures while the gross national product was increasing 400 percent (1947 to 1969) really justified such vast expenditure?

These unanswered questions have resulted in less public confidence in schools. Moreover, the public has refused to accept professional explanations for increasing the cost of education. In addition, public dissatisfaction has resulted in taxpayer dissatisfaction, teacher and student militancy, and social unrest in schools.

The traditional confidence that the average citizen had in his schools has decreased. (Burrup 1974:55)

Development of Teacher Evaluation from 1912 to 1956

According to Ayres (1912:307), the final citadel in which the old guard is now making its last stand consists of the objection that the most important elements of true teaching can never be measured.

Ayres stated further that:

As school executives make practical application of the newer scientific tests, no fact stands out with more impressive distinctness than that the teachers whose classes make the best records are the teachers who are the most truly successful in the shaping of character.

To answer the problem in question, Ayres (1912:308) referred to the scientific method as follows:

Simple as it sounds, this change from asking "What results should we get?" to asking "What results we are getting?" is the keynote of the whole scientific method in education. To answer the question in its new form means the development of units of measurement, and when these are secured the standards of attainment will work themselves out automatically.

In discussing educational efficiency, Ayres (1912:310) stated:

Education today wishes to be efficient, as industry is efficient; education wishes to know what the product is, and to gauge the time, quantity, and value elements.

The National Education Association (NEA) (1932:5) stated the demand for skilled teaching as follows:

Teaching, if it is to be skilled, and no other kind should be acceptable, demands a trained, experienced, and competent teaching personnel. The product of the public

school will in many cases continue to be disappointing until our children are instructed by teachers properly qualified for this important service. Making good citizens is a skilled service and cannot be performed by unskilled workers.

In order to measure teaching and to get qualified personnel, the NEA (1932:41) said that:

Authorities are generally agreed, however, that although the rating of teachers may be a useful supervisory device, it has not yet been sufficiently refined to permit any but rather rough distinctions between different grades of teaching ability. For this reason the use of rating scales in the administration of the regular salary schedule cannot be undertaken with great confidence in the outcome. Teachers are particularly resentful of anything akin to the capricious in a matter so important to them as their salaries.

Hartmann (1933:7) discussed the concern for human efficiency by saying:

The ideal of human efficiency would be the production of the maximum output of the highest quality in the shortest time, with the least expenditure of energy and with the maximum satisfaction.

In applying the above principle to efficient teachers he stated:

The ideally efficient teacher is the one who can accomplish the largest number of important and socially desirable changes in the greatest number of pupils in the shortest possible time, with the least expenditure of energy and with the maximum satisfaction in the learning process and its outcome by all concerned.

Hartmann (1933:14) noted that measuring the product of a teacher's effort was the greatest difficulty the researcher faced for the result was not seen in a week, not in a month, nor in a year, but some ten to twenty years after the initial experience. Hartmann

(1933:16) stated further that:

A sober, critical review and appraisal of the impasse facing research in this field has been made by Symonds. He notes that the high correlation which Boyce found between general rating for teaching efficiency and such items as "development of pupils" (.88), "growth of pupils in subject-matter" (.87), and "attention and response of class" (.86) suggested to other workers the use of pupil achievement as a measure of teaching efficiency. This implied the adoption of the Jesuit principle that faith without works is dead. "After all, the final criterion of any activity is the results produced. Traits in the teacher are valuable as measures of teaching efficiency only when they are effective in producing desirable changes of learning in pupils." Such doctrine is both true "Westernism" and good behaviorism. Consequently, the proposal that the accomplishment ratio be employed as a means of estimating efficiency was inevitable. As developed by Franzen, the formula read very simply:

Teaching efficiency - Final AR - Initial AR

In 1935, Lancelot (1935:14-16) investigated teaching efficiency by comparing the grades of engineering students obtained from sixteen teachers at Iowa State College in the years 1920-1928. According to Lancelot, the number of superior and average students taking courses under a specific teacher were known. The average of all the college grades of the groups was computed; including both the estimated grade and the actual grade of students taking courses from each professor. The means of the estimated grades were subtracted from the means of the actual grades of each group and the probability that other similar groups would exceed departmental norms were calculated. According to Lancelot, fifty-two superior students taking courses under teacher A obtained a probability of .996, while 28 superior

students taking courses under teacher J obtained a probability of .002. This research showed that of a large number of superior groups taking Mathematics I under teacher A, 996 out of each 1,000 could be expected to receive a subsequent higher mathematic average than those normally received by comparable student groups in this course. With other groups taking the course under teacher J, for example, only 2 out of 1,000 would probably do so. Probabilities as to better-than normal results obtained under other teachers lie between these extremes.

Fifteen years later Domas (1950:109-110) stated the problem of recognizing and rewarding merit in teaching as follows:

The efficiency of teachers may be gauged either by the Accomplishment Quotients of the pupils or by the judgment of supervisors. These proposals were opposed by teachers.

According to Domas (1950:111), a group of supervising principals in a class in supervision at the John Hopkins University prepared a list of factors which evidenced skills in teaching; the list consisted of only observable, improvable factors. The major division into which the factors were classified and listed were the following: (1) changes brought about in the pupils; (2) changes noted in the teachers; (3) visible changes in the classroom. Producing changes in pupils, measured in terms of educational objectives, was the ultimate criterion of teaching success.

Another point noted by Domas (1950:106) concerning components of

good teaching ability was that:

An analysis of the opinions of a large group of employers of teachers concerning the factors involved in good teaching indicated that the three components of teaching ability are: (1) knowledge of subject matter; (2) technique of teaching; and (3) personality.

He further stated that the rating scale proposed by a committee of Chicago teachers included and covered the following areas: (1) instructive qualifications, (2) professional aspirations and opinions, (3) general practice, (4) personal qualifications, and (5) achievement of pupil growth.

On the topic of "For what is the teacher paid?" Domas (1950:109) stated as follows:

Relationship between salary, and experience, training and teaching load were studied in groups of 7159 women teachers and 1520 men teachers throughout the state of Ohio. The data were tabulated according to sex of teachers, type of school, i.e., elementary and secondary, and rural and urban. (1) In all types of school positions and for both sexes, except for men in the one room rural school, there is a relatively high correlation between salary and experience. (2) In all types of schools and for both sexes, there is a relatively low correlation between salary and training. (3) The correlation between salary and teaching load is low in nearly every case. (4) From one type of position to another, salary increases are accompanied by positive changes in the amount of training.

Further Comas (1950:109) noted that several studies in the evaluation of faculty services at the college level were examined to determine qualifications which were of primary importance for ranking the faculty of teachers' colleges. Qualifications selected were the following: (1) teaching ability; (2) scholarship and scholarly

ability; (3) experience; (4) personal qualities; (5) membership in learned societies; and (6) capacity for departmental and college administration. Domas noted that in order to measure teaching ability, researchers discussed teaching methods, processes, personality, and results.

In 1951 Jensen (1951:84) studied the critical requirements for teachers which included: (1) The area of Personal Qualities referring to the emotional stability of the teacher as revealed by the interaction of teachers with pupils and associates, and also behavior that reflected honesty, fairness and objectivity. (2) The area of Professional Qualities included the classroom practices of teachers as they related specifically to the learning process. Included therein are the instructional skills of the teachers, knowledge of subject matter taught, organization of instructional materials, ability to diagnose individual and group weaknesses, and ability to provide necessary remedial instruction. (3) The area of Social Qualities referred to the teacher's face-to-face relationship with students and associates, and included ability to understand and appreciate the feelings of others, employment of the democratic approach in human relationships, and friendliness.

Remmers (1952:242-248) identified the definition of teacher effectiveness by noting that criteria of teacher effectiveness contained two parts. They were criterion and effectiveness. Criterion

was defined as a standard against which a measurement is made in estimating the validity of the measurement. A criterion is always concerned with one or more specified dimensions of whatever is being measured; one criterion dimension may be the length of the table and another may be its weight. In identifying effectiveness as a criterion dimension of teachers, he implied that the purpose of the measurement or appraisal of teachers was to estimate whether they will produce desired amounts and types of changes in pupils' behavior.

Remmers explained that effectiveness is the degree to which an agent produces effects. There are three categories of effect, in terms of the object affected, (a) the pupil, (b) school operations, and (c) the school-community relationship. The effects on pupils that are relevant as criterion dimension of teacher effectiveness are the extent to which educational objectives are attained.

Flanagan (1954:354) described the development of a method of studying activity requirements which he called the critical incident technique. The technique grew out of the studies carried out in the Aviation Psychology Program of the Army Air Force in World War II. The success of the method in analysing such activities as combat leadership and disorientation in pilots resulted in its extension and further development after the war. This developmental work was carried out primarily at the American Institute for Research and the University of Pittsburgh. The five steps included in the critical

incident procedure as most commonly used at the present time are:

(a) Determination of the general aim of the activity. This general aim should be a brief statement obtained from the authorities in the field which expresses in simple terms those objectives to which most people would agree. (b) Development of plans and specifications for collecting factual incidents regarding the activity. The instruction to the persons who are to report their observations need to be as specific as possible with respect to the standards to be used in evaluating and classifying the behavior observed. (c) Collection of the data. The incident may be reported in an interview or written up by the observer himself. In either case it is essential that the reporting be objective and include all relevant details. (d) Analysis of the data. The purpose of this analysis is to summarize and describe the data in an efficient manner so that it can be effectively used for various practical purposes. It is not usually possible to obtain as much objectivity in this step as in the preceding one. (e) Interpretation and reporting of the statement of the requirements of the activity. The possible biases and implication of decisions and procedures made in each of the four previous steps should be clearly reported. The research worker is responsible for pointing out not only the limitations but also the degree of credibility and the value of the final results obtained. (Flanagan 1954:354-355)

It should be emphasized that critical incidents represent only raw data and do not automatically provide solutions to problems. However, a procedure which assists in collecting representative samples of data that are directly relevant to important problems such as establishing standards, determining requirements, or evaluating results should have wide applicability.

The application of the critical incident techniques which have been made to date are discussed under the following nine headings:

(a) measures of typical performance (criteria); (b) measure of proficiency (standard samples); (c) training; (d) selection and

classification; (e) job design and classification; (f) operating procedures; (g) equipment design; (h) motivation and leadership (attitudes); and (i) counseling and psychotherapy.

Flanagan (1954:355) summarized the critical incident technique as follows:

The critical incident technique, rather than collecting opinions, hunches and estimates, obtains a record of specific behaviors from those in the best position to make the necessary observations and evaluations. The collection and tabulation of these observations make it possible to formulate the critical requirements of an activity. A list of critical behavior provides a sound basis for making inferences as to requirements in terms of aptitudes, training, and other characteristics. It is believed that progress has been made in the development of procedures for determining activity requirements with objectivity and precision in terms of well-defined and general psychological categories. Much remains to be done. It is hoped that the critical incident techniques and related developments will provide a stable foundation for procedures in many areas of psychology.

According to Barr (1955:262), he stated Ryans and Wandt identified the following factors for teacher behavior: tendency to be sociable, businesslike, reactive, tolerant, and pleasing. In addition they identified one factor of pupil behavior, tendency to participate, which was linked with certain teacher behavior.

Barr further stated that Hearn, in case studies of seventy-seven teachers, found skill in human relationships to be important; Cook found that teacher attitudes toward children correlated significantly with teacher-pupil relationships. Reed found a relationship beyond chance expectancy between the teacher's effectiveness in the classroom

as evaluated by the students and that aspect of the teacher personality which permits him to be a person who accepts other people, Anderson found high school teacher morale related to high pupil achievement; and Erickson, from a factor analysis of teaching ability considered the following to be related to teacher efficiency:

(a) positive character, (b) bohemian attitude, and (c) well controlled character stability.

Barr further stated that Montross found positive correlation between certain objective measures of temperament, such as speed of tapping, reaction time, fluency, and right- and left-hand coordination, and teacher success; Page and Travers found a triad of Rorschach patterns associated with patterns of behavior considered desirable by supervisors; Simpson, Gaier, and Jones concluded that resourcefulness is a function of attitudes and habits of applying existing knowledge and skills in practical situations rather than a function of teacher knowledge and information; Schultz and Ohlsen found that outstanding student teachers were creative and enthusiastic, had genuine interest in students, and organized their work well; Adaval found that India's teachers with advanced degree showed knowledge superior to that of teachers with lesser degree, and Allman found prospective teachers superior to other students in such basic competencies as mental ability, reading comprehension, and achievement in elementary-school subjects.

According to Barr (1955:263) rating scales of one form or another receive extensive attention. He stated that Kessler and Hosley constructed five-point rating scales for evaluating nursery school teachers. They found these, under the conditions in which they were used, to correlate .71 with ranks assigned by supervisors.

Barr (1955:263) indicated that Berkshine and Highland, Harding and Long, and Ryans recommended the forced-choice rating scale. He stated that Berkshine and Highland concluded from a study of such a scale that it is probably better to combine with the forced-choice rating procedure some of the more conventional rating forms. Further Barr stated that Daval and Chatterjee found, with carefully constructed rating scales, a high agreement among those who assess the abilities of trainees in a teacher education institution, and Guelso, and Hobson and Schlenk reported plans for evaluating military instructor training.

A number of tests of qualities or characteristics of teachers were proposed. Among these was the Minnesota Teacher Attitude Inventory which has been subjected to extensive study. Barr indicated that Leeds, using "Expert" ratings as the criterion in rating principals and students, found that the ratings of principals and "Experts" were more closely in agreement than either of these when compared with pupil reactions. He concluded that pupil ratings make a unique contribution to teacher evaluation.

Further Barr (1955:263) indicated that Wrightstone and others used sociometric techniques to study and improve intra-staff acceptability of teacher isolates. They found a close association of the data that occurred with teachers popularity; and Jerecke explained the construction of a teaching judgment test to evaluate teaching success. From a study of forty-one and fifty-five teachers graduated from the University of West Virginia with one to fifteen years of experience he concluded: (a) that teaching experiences seemed to have a connection with teaching success, (b) that some unnamed factors as measured by the Bernreuter Personality Inventory affected teaching success, (c) that scholastic ability as measured by a master's examination and the teaching judgement test were related to teaching success. (Barr 1955:263-264)

In 1956 Nelson (1956:24) stated as follows:

It is possible that effectiveness in teaching is made up of several aspects, namely, disciplinary control, teacher-pupil relationship, and instructional excellence. It is also possible that teachers may display different degrees of these aspects in different school situations. If this is true, then a teacher who might be considered to be an excellent one in one school might not be as effective in another school in which a different aspect of teaching was emphasized.

According to Nelson (1956:57) there were rather high interrater consistency coefficients among the ratings of teachers by pupils, by supervisors and by observers, the interrater consistency coefficient obtained from his studies being .62, .54, and .79 on disciplinary control, teacher-pupil relationship and instructional excellence

respectively. These results, as viewed by Nelson, were sufficiently reliable as he stated:

In view of the interrater consistency of the scores obtained for the three measuring instruments it must be concluded that these instruments are sufficiently reliable to be used in the evaluation of effectiveness for both groups and individual teachers.

Development of Teacher Evaluation from 1957 to 1969

Beginning in 1957, the State of North Carolina began studying the use of merit systems. During the 1957 session of the General Assembly of the State of North Carolina, the problem of teacher salary was discussed. Following this, the Board of Directors of the North Carolina Education Association notified the State Board of Education of their concern with and interest in the merit rating system. During the year 1958, the State Board of Education of North Carolina was in session and raised the question concerning the merit plan for teachers. As a result of the session, a committee was authorized to study the problem in question.

In 1959, according to McPhail (1967), the General Assembly of the State of North Carolina discussed the problem of merit pay again and adopted two resolutions pertaining to teacher evaluation. The first resolution directed the State Board of Education to study teacher evaluation, rating and certification. The second resolution, the General Assembly empowered and authorized the governor to appoint a commission to study the merit pay system.

In the year 1960, Ryans (1960:1490), studying teacher effectiveness, proposed the probable correlates of teacher effectiveness in the contemporary United States as follows:

The following generalizations regarding the relationships between teacher characteristics, as predictors, and teacher effectiveness, as a criterion abstracted from various criteria measures reported in the literature, appear to be in order.

Measured intellectual abilities, achievement in college, general cultural and special subject matter knowledge, professional information, student teaching marks, emotional adjustment, attitudes favorable to students, generosity in appraisals of the behavior and motives of other persons, strong interest in reading a literary matters, interest in music and painting, participation in social and community affairs, early experiences in caring for children and teaching (such as reading to children, taking a class for the teacher), history of teaching in family, size of school and size of community in which teaching, cultural level of community, and participation in avocational activities, all appear to be characteristics of the teacher which are likely to be positively correlated or associated with teacher effectiveness in the abstract.

In the same year Mitzel (1960:1482) discussed the important attributes of criterion measures by referring to Steven's discussion of operationism. He pointed out that concepts are defined in terms of the operations that produced them. Thus, teacher effectiveness as a concept has no meaning apart from the criterion measures or operational definition of success as a teacher. These measures should possess four basic attributes: (1) relevance, (2) reliability, (3) freedom from bias, and (4) practicality.

Relevance as a criterion attribute is a product of a rational analysis of the job functions and the job objectives. In so far as a

criterion measure reflects the behaviors required in the achievement of job objectives, it is relevant. For some types of production work, such as operating a punch press, the job objectives are fairly easy to specify and are often self-defining. Indeed, a simply count of the number of units produced by an operator during a specified period will often suffice as the directly relevant job criterion. On the other hand, analysis of tasks such as teaching does not readily yield criteria which are directly relevant to the job functions. As a consequence, we have often had to depend upon direct and fragmentary definitions of teaching success. The difficulty stems, in part, from lack of agreement among educators on a hierarchy of goals and objectives for teaching. Teachers in elementary school, for example, have been asked to work for changes in such different types of student behavior as basic intellectual skills, attitudes, personal-social adjustment, cultural appreciations, and health habits. This proliferation of school goals without accompanying agreement as to which ones are more important and which ones are less important, has made the task of selecting relevant teacher effectiveness criteria almost impossible.

Mitzel (1960:1485) discussed the issues pertinent to criterion selection as follows:

The day-to-day demands of training, hiring, and promoting teachers have forced the educational community to accept, either implicitly or explicitly, certain criteria of effectiveness. Criterion decisions, whether they involve personality attributes of teachers, number of years of experience, number of credits in graduate studies, or growing in basic skills of

students taught, all involve some commitment on the following issues: (a) Is teacher effectiveness multi-dimensional or unidimensional? (b) Should teaching effectiveness be evaluated primarily against the intellectual cognitive goals of education or primarily against affective attitudinal goals?

For the past fifty years, most research and administrative practice seems to have been based on the assumption that teaching competence is a unitary trait. Although it has long been felt that there are degrees of effectiveness among teachers, many educators still act on the assumption that the teacher who stimulates the greatest student growth in one basic skill will stimulate the greatest growth in other skills, as well as problem-solving, social adjustment, and other educational objectives (23, 24, 25). It is also frequently assumed that the effective teacher, however defined, is equally effective with all children. The weight of the evidence, though fragmentary, preponderantly supports a multidimensional view of teaching effectiveness.

During the year 1961-1963, the study of the merit system was made by the commission of the state of North Carolina. There were about twelve thousand educators and leaders at the state level involved in this study. Approximately, four hundred thousand dollars was spent in an effort to determine the degree of superiority in teaching and to pay the teachers in terms of their effectiveness.

The results revealed in the report to the 1965 General Assembly indicated that (1) a uniform statewide program of merit pay was not feasible, (2) a merit pay program at the local level would be possible under certain provisions, and (3) the instructional improvement at the state level should be continued and strengthened. (McPhail 1967)

In the year 1961 the American Association of School Administrators (1961:15) stated the following about the measure of teacher

effectiveness:

If the purpose of teaching is to bring about desired changes in pupils, the obvious measure of teacher effectiveness is the extent to which the teacher actually produces such changes. Unfortunately many difficulties intrude upon this happy prospect: (1) it is difficult to measure pupil growth; (2) it is difficult to determine precisely how much of the change can be directly attributed to the teacher.

AASA (1961:15) discussed the variety of methods used in evaluating teaching effectiveness. Among the variety of methods that have been used are achievement quotients, raw gain, residual gain, and other measures which are usually variants of the residual gain procedure. The achievement quotient, patterned after the IQ test was designed to take into account both achievement and ability. Achievement age (such as in arithmetic) is divided by mental age to yield a quotient which is then expressed as a percent (i.e., multiplied by 100). Scores less than one hundred presumably indicate below-expected achievement. Achievement for a class is then expressed as an average of achievement quotients. The method has obvious arithmetical as well as logical limitations, so that it is now rarely used. Achievement quotients are intimately connected with limitations in the tests used; individuals with high achievement near the test ceiling usually tend to get lower IQ scores than those with low achievement scores.

In studies prior to 1940, raw gain was probably the most widely used measure of pupil change. This measure was obtained by subtracting a pretest score (given at the beginning of the evaluation period) from

a final test score (an equivalent form of the same test given at the end of the evaluation period). If the groups were matched initially, both on knowledge of the subject and intellectual ability, some of the contaminating factors were said to be reduced. Such matching may be possible under certain controlled study conditions, but these conditions were not usually present in typical teaching situations, where the rule was wide variability within classes and situations that faced teachers. The practice then was to ignore this problem.

In recent studies some form of residual gains are most frequently applied as procedures for studies of pupil gains. Residual gain provides for some equating among the factors that potentially influence results. A common type of residual gain is the estimate (usually by means of regression equations) of the progress that students in the class should make, and then to compare difference between expected and actual progress. This method has been formalized in a procedure known as the analysis of covariance. Other methods include linear response surfaces, contour analysis, and, at a much more direct level, the comparison of regression coefficients (usually the regression of initial score on final score) within various classrooms.

While most investigators agree that student gains are the ideal criterion measures, many difficulties intrude. The difficulty of constructing sufficiently accurate and comprehensive measure makes this a costly procedure. It is not surprising that the frequency of

