1989

Relationships between supervisor ratings of teacher performance and student achievement

Bruce E. Daniels
Iowa State University

Order Number 8920122

Relationships between supervisor ratings of teacher performance and student achievement

Daniels, Bruce E., Ph.D.

Iowa State University, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Relationships between supervisor ratings of teacher performance and student achievement

by

Bruce E. Daniels

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Professional Studies in Education
Major: Education (Educational Administration)

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1989

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION

The National Commission on Excellence in Education (1983), in their report A Nation at Risk, pointed out the decline in student achievement and urged reform measures. There has been a continuous demand from the American public, especially since the release of this famous report, for more accountability overall in our schools and, in particular, on the part of our teachers.

Popham (1975) vividly described the difficulty educators have faced through the years in attempting to develop adequate methods for valid and reliable teacher evaluation.

> Since the early 1900s educational researchers have spent endless hours trying to devise a satisfactory measure of teacher competence. In view of the enormous energy expended in trying to identify this elusive prey, some disgruntled researchers have opined that a defensible index of teaching skill ranks third behind two other hard-to-locate targets:
>
> Among Mankind's Perennial Quests
> 1. The Holy Grail
> 2. The Fountain of Youth
> 3. A Valid Index of Teaching Skill (p. 283).

Rosenshine (1976) also cited the difficulty and inconsistency involved in research on effective teaching and implied a need for continued study in the field.

> . . . research on observed teaching behavior is new, sparse, and not always consistent in results. What we have learned to date is offered more as hypotheses for future study than as validated variables for the training and evaluation of teachers. Although practitioners can easily amass a large number of questions on teaching methods for which they would like clear answers, at the rate we are going it will be years before many of these questions are even studied (p. 61).

Most important decisions about teachers--certification, competence,

hiring, retention, dismissal, tenure, deserving of merit--are dependent upon a judgment by a supervisor or principal of how well that teacher performs in the classroom. Ideally, these judgments result in effective teachers being certified, determined competent, hired, retained, granted tenure, and recognized as meritorious more than those teachers determined to be less effective. The accuracy of these judgments, unfortunately, has not been consistently established and is merely assumed rather than proven (Medley & Coker, 1987).

Recent research has identified certain teacher behaviors that consistently appear to be related to teacher effectiveness. Even though these relationships are correlational rather than causal, they do seem to refute opinion voiced in years past that variations in teaching make little difference in student learning and that educational research in teacher behavior-student achievement relationships is too inconsistent to have practical implications for our schools.

A consortium of five school organizations, the Minneapolis Public Schools, Edina Public Schools, Northfield Public Schools, and Breck School, all in Minnesota, and the Spirit Lake (Iowa) Community Schools, and a team of investigators from the Research Institute for Studies in Education (RISE) at Iowa State University was formed to develop a school improvement model that linked administrator performance, teacher performance, student performance, and staff development interventions. This total-systems approach to improving performance in classrooms and schools came to be known as the School Improvement Model (SIM) project. The SIM project was to demonstrate the effect of an articulated system of

administrative and teacher performance appraisal upon student achievement in fourth and eighth grade mathematics and fourth grade reading as measured by both norm- and criterion-referenced tests. Teachers and administrators were taught how to do their jobs better and coached for improved performance by appraisers using state-of-the-art tools and skills. The achievement of the learners was examined for improvement (Manatt & Stow, 1986).

For the purposes of the SIM project, school improvement was equated with improved student learner outcomes and started with changing teacher behavior. Therefore, it was essential that changes in teacher performance be accurately measured throughout the SIM project. Teacher performance was measured following specifications drawn by locally appointed steering committees. In this way, procedures and performance criteria were selected which were relevant to all stakeholder groups represented by members of the steering committee and which also could be replicated in subsequent years by other school organizations following the SIM design. The five consortium school organizations each selected some teacher performance criteria that were similar and some that were very different (Manatt & Stow, 1986).

Statement of the Problem

It seems logical that if teachers are to be held accountable for student achievement, then teacher behaviors addressed in staff development interventions and teacher evaluation programs should be those that hold some likelihood for improving student achievement. It follows

also that principals and other supervisors who evaluate and coach teachers should be able to identify those teacher behaviors that are related to better pupil learning.

The problem for the present investigation was to determine if a relationship existed between supervisor ratings of teacher performance and student achievement. Performance criteria selected were those identified by recent research as being related to teacher effectiveness. Student achievement was measured in fourth and eighth grade mathematics and fourth grade reading by using pretests and posttests, both norm-referenced and criterion-referenced.

The following questions more succinctly defined the problem which was addressed:

a. Is there a relationship between the supervisor ratings of the individual teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading?
b. Is there a relationship between the mean supervisor ratings of the teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading?
c. Is there a relationship between the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas and student achievement in fourth and eighth grade mathematics and fourth grade reading?
d. Is there a difference in the supervisor ratings of the individual teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower

third of those tested?

e. Is there a difference in the mean supervisor ratings of the teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested?

f. Is there a difference in the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested?

Purposes of the Study

The importance of principals and others who supervise and coach teachers being able to identify teacher behaviors which are significantly related to student achievement has been well-established. Therefore, the purposes of this study were:

a. To determine if supervisor ratings of teachers can be used to predict student achievement.

b. To determine which teacher performance criteria are related to higher student achievement in fourth grade reading and mathematics and eighth grade mathematics.

c. To determine which logical groupings of teacher performance criteria clustered into performance areas are related to higher student achievement in fourth grade reading and mathematics and eighth grade mathematics.

d. To improve future instruments to be used for teacher performance

evaluation.

e. To improve staff development interventions and teacher education programs by identifying effective teacher behaviors to be addressed by such training.

Objectives of the Study

In order to accomplish the purposes of this study, it was necessary to develop a teacher evaluation instrument for the teachers participating in the study. A data base which includes performance evaluation data and student achievement data for each teacher participating in the study was established. Mean student achievement scores were calculated for each teacher's class, and appropriate statistical tests for each of the hypotheses were selected and used.

Hypotheses to be Tested

The research questions posed by this investigation were tested by the following null hypotheses:

a. There will be no relationship between the supervisor ratings of the individual teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading.

b. There will be no relationship between the mean supervisor ratings of the teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading.

c. There will be no relationship between the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas and student achievement in fourth and eighth grade mathematics

and fourth grade reading.

d. There will be no significant difference in the supervisor ratings of the individual teacher performance criteria among the teachers whose class achievement means were in the upper third, and lower third of those tested.

e. There will be no difference in the mean supervisor ratings of the teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested.

f. There will be no difference in the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested.

Basic Assumptions

This study was based upon the following assumptions:

a. That the teachers evaluated in this study are a representative sample of the fourth grade reading and mathematics teachers and eighth grade mathematics teachers in the school districts which participated in the SIM project.

b. That the supervisors based their evaluation responses on their knowledge of the teachers' demonstrated ability related to the selected individual teacher performance criteria and not on other teacher behaviors.

c. That the supervisors completed each teacher evaluation instrument

independently.

d. That the student achievement testing done by the SIM team was conducted in a manner that provided valid and reliable results.

Delimitations

This study is a small component of the School Improvement Model (SIM) project conducted in the school organizations of Minneapolis, Northfield, Edina, and Breck, all in Minnesota, and Spirit Lake, Iowa, sponsored by the Northwest Area Foundation. As such, the following delimitations apply:

a. All supervisors who evaluated fourth grade reading and mathematics teachers and eighth grade mathematics teachers were asked to complete a summative evaluation of those teachers based on selected performance criteria.

b. Summative evaluation instruments were completed on approximately eighty teachers in the five school districts. Of these, complete data sets (student achievement scores and teacher evaluation ratings) exist for fifty-three teachers. Only these teachers were selected and analyzed for the present study.

c. Teacher performance criteria that research on teacher effectiveness has identified as effective teacher behaviors and which were utilized by all five school organizations in the project were used in the study.

d. Reports of student achievement scores on both norm- and criterion-referenced tests, by teacher, in fourth grade reading and mathematics

and eighth grade mathematics were used.

e. Due to administrative changes at the Breck School, student achievement data were not available, and, therefore, Breck was not represented in this study.

f. Student achievement data used in this study were not disaggregated by gender, race, or socioeconomic status.

g. All principals in Northfield, Edina, Breck, and Spirit Lake were provided with thorough training in teacher performance evaluation along with one-third of the principals in Minneapolis.

h. Only the principals trained in teacher performance evaluation techniques were utilized in this study; all others were excluded.

i. A common problem of most teacher performance evaluations is that ratings are too lenient because evaluators fear the negative feedback from teachers caused by such ratings. The present investigation avoided this problem because the ratings were covert in the sense that they were not shared with the teachers.

The Iowa State University Committee on the Use of Human Subjects in Research reviewed this project and concluded that the rights and welfare of the human subjects were adequately protected, that risks were outweighed by the potential benefits and expected value of the knowledge sought, that confidentiality of data was assured, and that informed consent was obtained by appropriate procedures.

CHAPTER II. REVIEW OF LITERATURE

The review of the literature for this study focuses on employee performance appraisal in general and performance evaluation of teachers in particular. A second focus relates to measures of student achievement which can be attributed to instruction received in the classroom. The employee performance appraisal review relates the accuracy of performance ratings to the purposes for which the ratings are used and also to the consequences of the appraisal for both the rater and the ratee. The review of teacher performance evaluation concentrates on supervisor ratings of teacher performance along with the criteria reported by various studies to be related to teacher effectiveness. The review of student achievement measures examines concerns related to student gain scores, norm-referenced tests, and criterion-referenced tests.

> Most thinking people have always praised evaluation, at least in the abstract, as an intellectually defensible activity. Only the charlatan or the incompetent, it would seem, has reason to fear the effects of evaluation. Through the centuries, therefore, our most capable scholars have recommended that human beings engage in evaluative operations, that is, the evaluation of their own actions, the evaluation of other people's acts, the evaluation of myriad aspects of their environment. Evaluation has historically been viewed, and quite properly so, as an integral activity of a rational approach to life (Popham, 1975, p. 1).

Employee Performance Evaluation

Employee performance appraisals are typically used in organizations for administrative decisions such as promotion, transfer, and financial incentives; employee development interventions; and personnel research. Such appraisals tend to lack accuracy when they are obtained for

administrative purposes rather than for employee development or personnel research. The differences in accuracy can be explained in terms of the relationship between the purposes of appraisal and the appropriateness of confidentiality, rater role conflict, and the possible negative impact of appraisal results on future rater-ratee relationships. Raters may use one set of standards for judging another's performance if the rating is to be used for research purposes and another, more lenient, standard if the results are to be used for administrative decisions such as promotion or demotion (Barrett, 1966; Mohrman & Lawler, 1983; Murphey, Balzer, Kellam, & Armstrong, 1984).

Performance appraisal for personnel research

Performance appraisals for personnel research purposes usually are not available to the ratee and, therefore, are not likely to affect future interactions between rater and ratee. They are separate from the customary role of the evaluator and have no effect on the ratee's receiving valuable organizational rewards. Ratings obtained under such conditions pose no threat to the status quo of either supervisor or employee, and the supervisor likely will be motivated to assign the performance ratings he or she feels most accurately reflect true job behavior on the part of the employee (Decotiis & Petit, 1978).

Performance appraisal for employee development

Performance evaluation conducted for the purposes of employee development decisions are usually more accurate than administrative appraisals while being less accurate than appraisals conducted for

personnel research. Employee development may be viewed by the rater as being in the best interest of the ratee. Under conditions of performance goal setting or mutual problem solving, a rater may be motivated to give accurate performance appraisal ratings. However, if the rater is required to not only identify employee development needs but also to be responsible for corrective changes in ratee behavior, inaccurate (lenient) ratings may result as a means of avoiding such responsibilities (Decotiis & Petit, 1978).

Performance appraisal for personnel decision

Appraisals to collect employee data to help in making administrative decisions are likely to be the least accurate. If an organization attempts to link the administration of employee rewards to job behavior, the results of appraisal will have to be shared with the employee. The rater is often required to justify the ratings, is not comfortable in doing so, and is reluctant to have a negative impact on future rater-ratee interactions. Therefore, he or she will be motivated to assign lenient and inaccurate performance ratings. Regardless of purpose, the motivation of the evaluator to assign accurate performance ratings is higher when the results of the appraisal are confidential from the employee being evaluated (Decotiis & Petit, 1978).

Leniency in performance appraisals

There is a common tendency toward leniency in military ratings. Typically, 95 percent of a unit's personnel are rated in a category identified as including only the top 5 percent. Evidence points to the

purpose of the ratings as having an effect on leniency. Officers are rarely promoted unless they are ranked in the top 5 percent of those being evaluated. This leads researchers to conclude that appraisers inflate their ratings if they are to be used for promotional purposes (Henderson, 1984).

In the teacher performance evaluation system in the Dallas Independent School District, the elementary principals rated their own teachers. Each teacher was also rated by the principal of another elementary building in the district. The teacher's own principal continued to work and interact with the teacher before, during, and after appraisal ratings were made. The second principal appraiser could observe and rate the teacher, report the ratings to the teacher's building principal, and exit the evaluation process without any further face-to-face contact with the teacher. The mean score for these outside appraisers was lower than the mean rating assigned by the teachers' own building principals. This shows the tendency for leniency in the ratings of evaluators having ongoing relationships with appraisees (Peterson, 1988).

It is necessary for the appraiser of employee performance to be motivated to provide an accurate report of performance or at least not be motivated to give an inaccurate report. "Ultimately, it seems that the best way to get accurate reporting is to do nothing with the data, because any use is likely to result in problems for the appraiser" (Mohrman & Lawler, 1983, p. 186). The motivation of the rater of employee performance to be objective is important to the accuracy of such

performance ratings. ". . . [I]f the rater is motivated to make accurate judgments about his ratees and if he feels free to be objective, then he has a good chance of achieving his aim. . ." (Taft, 1971, p. 176). Most of the attention has focused on the technical aspects of performance appraisal and little on human factors which affect the quality of appraisal ratings. Those acknowledging such human factors see the appraisal of human performance as an activity based upon a certain amount of subjectivity (Ilgen, 1983).

The evolution of teacher evaluation may well depend on the ability of researchers and practitioners to deal successfully with the technical and human barriers that have prevented the achievement of performance appraisal systems in the past. Blumberg (1974) sees those responsible for evaluating teachers as being in a difficult position. They are required to perform the seemingly conflicting functions of helping teachers teach and then of appraising that teaching. He calls the situation that results "a private cold war."

Teacher Evaluation

The emergence of the problem for this study can be traced to the early 1900s when multifactor rating scales first caught the attention of educators and came under their close scrutiny. Teacher evaluation instruments of this type confront the supervisor with a list of teacher performance criteria thought to be related to teacher effectiveness. The supervisor rates the teacher on each criterion by recording a number representing his/her opinion of the status of the teacher (Boyce, 1915;

Medley & Coker, 1987).

Early teacher evaluation criteria

Data were gathered from an analysis of 209 teacher rating scales which were collected from cities, state departments of public instruction, and departments of education in universities. A total of forty-six states were represented in data analyzed. The most frequently mentioned teacher effectiveness characteristics were: results of instruction; scholarship and professional preparation; ability to cooperate with others; personal appearance; appearance of classroom; interest in work, pupils, patrons, subject matter, etc.; discipline; considerateness; leadership; and daily preparation (Barr & Emans, 1930).

The published results of the first test of the teacher performance rating instruments reported one significant correlation between teacher performance ratings and teacher effectiveness, measured in terms of students' achievement test gain of .45, although the median correlation was only .24. These findings certainly cast doubt on the accuracy of supervisors' judgments of teacher performance. The report cited ten main divisions of rating schemes: (1) results of instruction; (2) classroom management; (3) technique of instruction; (4) personal equipment; (5) cooperation; (6) academic training; (7) professional training; (8) loyalty; (9) experience; and (10) general intelligence. However slight, the report indicated a trend to place more and more emphasis upon the tangible results of instruction in the rating of teachers (Hill, 1921).

Early attention to teachers' personal attributes led to presage-

product instead of process-product studies. Presage variables included those of teacher appearance, intelligence, leadership, and enthusiasm. Product variables were usually ratings by supervisors and principals. This approach produced some degree of consensus about desired traits of teachers, but no correlation between specific teacher behaviors and student achievement resulted (Brophy & Good, 1986).

Many studies were conducted over the next forty years, and most of them reached the same conclusion. While advocating the use of teacher effectiveness criteria based on measured pupil gains rather than on expert opinion, they found no correlation appreciably different from zero between evaluation of teachers on various rating scales and the achievement of their pupils (Anderson, 1954; Barr, Torgerson, Johnson, Lyon, & Walvoord, 1935; Brookover, 1945; Gotham, 1945; Hellfritzsch, 1945; Jayne, 1945; Jones, 1946; LaDuke, 1945; Lins, 1946; Medley & Mitzel, 1959).

Teacher evaluation concerns

It has been suggested that supervisor ratings do not correlate with pupil achievement because they reflect the emotional climate of the class rather than how much pupils are learning. It is not clear whether teacher performance ratings fail to relate to student achievement because it is impossible to assess learning in progress or because supervisors do not know how or do not know what to observe. To expect supervisors to tell how much a class is learning by looking at it assumes two things: (1) there is a pattern of behavior exhibited whenever optimum learning

takes place; and (2) the supervisor can recognize this kind of behavior when he or she sees it. The hurdle imposed by these criteria is significant (Medley & Mitzel, 1959).

In concluding that no single, specific, observable teacher act was found which was related to pupil learning, some possible explanations are submitted:

1. The teaching act is so complex with its many nebulous factors that a single observable teacher behavior divorced from others may likely not produce a measurable effect.
2. The various observed teacher behaviors are neither positive nor negative when removed from the situations giving rise to them.
3. The relationship which exists between certain teacher behaviors and pupil gain may be curvilinear rather than linear.
4. The primary difference in pupil gain scores may depend on factors inherent in the pupil rather than on teaching procedures (Jayne, 1945).

These same concerns confront researchers today in the study of teacher effectiveness.

The Coleman investigations concluded that students' family background, peer group, and social class are the most important variables related to pupil learning. Other variables such as student IQ motivation, prior achievement, and ethnic group are also important. They found that all other variables have minimal impact on achievement. A small fraction of the variation in pupil learning is explained by school-related variables, and only a minute part of this variation is explained

by teacher effects (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966).

The results of these studies have had no obvious effect on teacher evaluation practices. Most educational decisions today which involve teachers and teacher effectiveness are based on supervisor judgments which, according to Medley and Coker (1987), "are only slightly more accurate than they would be if they were based on pure chance" (p. 243).

Teacher effectiveness

The following statement about teaching expressed the sentiments of most students of teacher effectiveness in contemplating its complexities:

> . . . I wish to admit that I do not know what constitutes the 'essence of teaching'. I have come across hints, glints, and glimpses of this complexity, but it is much more diverse than my capability to describe it. Like the ocean, it is an awesome idea--and worth reaching out for (Glick, 1982, p. 44).

Typically, a teacher is considered effective if class means on multiple-choice standardized achievement tests are higher than expected based on pretest performance and if the usual process-product assumptions are met. These assumptions include: a common curriculum; that summative evidence of student knowledge is adequate; that no distinction is made among knowledge, effort, or test-taking skill; and effectiveness is aggregative across students within a classroom; i.e., regardless of performance distribution, the class mean is used to represent class performance. A similar definition of teacher effectiveness as used in process-product research refers to the ability of a classroom teacher to produce such higher-than-expected gains. This definition, referring to

standardized achievement test gains, with the added condition of similar gains on locally-constructed, district-specific criterion-referenced achievement measures, will be used to describe teacher effectiveness in the present study. Relatively effective teachers get their achievement gains from all levels of student ability, and relatively ineffective teachers do not depress the achievement for any particular level of student aptitude. The suggestion is that teacher effectiveness and student aptitude do not interact in a systematic way (Good, 1979; Good & Beckerman, 1978; Shavelson, Webb, & Burstein, 1986).

Difficulties in identifying effective teachers

Several reasons for not knowing if teachers make a difference in pupil achievement have been noted:

1. There is disagreement over what effects a teacher is called upon to produce.
2. Those using student gain as criterion of teacher effectiveness have paid little attention to individual students; class achievement has not been disaggregated.
3. There has been confusion over the use of a variety of terms and their definitions.
4. Teacher effects are nearly indistinguishable from effects of other teachers, parents, peer groups, etc.; the teacher does not operate in isolation from other teachers and other people.
5. Every behavioral act is the result of many antecedent and concurrent factors which are interrelated in a complex way.

It has been observed that student achievement is influenced by many factors that teachers do not control. Five influences on teacher effectiveness and, thus, on student achievement which are beyond the teacher's control are often mentioned: (1) the comparability of responsibility; (2) comparable materials, facilities, and time; (3) schools with similar learning climates; (4) similar levels of instructional support; and (5) adequate training. The effects of these influences must be removed before consideration of student achievement scores for evaluation of teachers (Cangelosi, 1984; Haertel, 1986; Ornstein, 1982).

If teacher competence is to be assessed by measuring teachers' effects on pupil learning, all other factors affecting pupil outcomes must be regarded as sources of errors of measurement. Five such sources are worthy of note: (1) teacher effort; (2) teacher performance; (3) inside situational influences--situational factors affecting teacher performance; (4) outside situational influences--ones that affect pupil learning without affecting teacher performance; and (5) pupil effort. Considering these five sources of error, it is easy to see why measures of teacher effectiveness are so unstable (Medley & Crook, 1980; Rosenshine, 1970; Soar, Medley, & Coker, 1983; Veldman & Brophy, 1974).

In pointing out potential problems in teacher effectiveness research, it should be noted that the problem of teacher evaluation calls attention to the complexity of human behavior and the difficulty of doing productive research on it. Every behavioral act is the result of many antecedent and concurrent factors that are complexly interrelated. To

identify and separate these factors and their relations is a formidable task. The behaviors and outcomes we are attempting to study cannot be isolated as a stable construct without distortion. Controlled experiments on people are difficult not only because of the great variety of factors that must be controlled but also because people are not always easy to manipulate and place into nice and neat categories (Capie & Cronin, 1986; Ornstein, 1986c; Sherman & Giles, 1983).

Several explanations of problems in teacher effectiveness research have been offered:

> . . . when the phenomena being studied contain several variables that interact in a complex way and when simple relationships combine to produce a host of combinations and an unmanageable number of other relationships, then the data are likely to be sketchy (Ornstein, 1982, p. 68).

Several years earlier, a similar conclusion was reached:

> Certainly until we are able to establish adequate criteria of teaching competency, our whole system of teacher training, appointment, promotion, and tenure fundamentally is on shaky ground (Ryans, 1949, p. 696).

Popham (1971) reported results revealing that experienced teachers may not be significantly more proficient than "people off the street" with respect to accomplishing intended behavior changes in learners. In three separate replications, groups of experienced teachers were unable to outperform nonteachers in bringing about specified changes in learners. He observed:

> The ability to bring about specified behavior changes in learners is by no means the only dimension to consider in evaluating a teacher. . . . Even so, however, it may not be an overstatement to assert that the skill necessary to bring about intentional changes in learners should be considered a necessary but not sufficient attribute of the high quality

teacher (p. 602).

The difficulties in assessing effective teaching have been noted by referring to the interrelationships within the context of teaching,

> . . . researchers need to take into account the fact that teaching skills do not exist in a vacuum. They must be studied in the context of a specific teaching method, the method must be linked to specific curriculum objectives and perhaps to specific curriculum materials. Furthermore, the method and materials must be used by specific teachers and students in a specific instructional setting. Finally, tests must be created which are responsive to the idiosyncratic outcomes of the teaching skills and curriculum objectives (Gall, 1977, p. 43).

Concerning these interrelationships, it has been observed that theories which attempt to explain teacher effects on student outcomes increasingly refer to coherent teaching strategies instead of isolated teaching skills. We cannot now (and maybe ever) conduct the classical experiment that holds everything else constant in order to manipulate a single variable, but we can conduct useful experiments involving treatment packages involving several elements. The importance of recognizing that numerous teacher behaviors combine to affect student learning is uniformly recognized (Brophy, 1979; Doyle, 1985; Good & Mulryan, in press).

Manatt, Palmer, and Hidlebaugh (1976) found ninety-four teacher performance criteria which proved to be sufficiently powerful to separate high teacher performance from mediocre performance. They reported thirty of the items which were representative of the total number. Each item appeared to fall into one of five rubrics descriptive of teacher behavior. The five categories were productive teaching techniques, positive interpersonal relations, organized/structured class management,

intellectual stimulation, and desirable out of class behavior.

Despite the difficulties which present themselves when attempts are made to assess teacher effectiveness, it is essential for educators not to assume that classroom events are too difficult to assess with sufficient accuracy, but rather, that they try to better understand the behaviors and methods for determining and improving teacher behavior and teacher effectiveness (Ornstein, 1986a).

Indicators of teacher effectiveness

Eight categories in assessing teacher effectiveness have been suggested: (1) teacher characteristics; (2) teacher experiences; (3) teacher properties; (4) student characteristics; (5) student experiences; (6) student properties; (7) classroom or process variables--teacher behavior, student behavior, and teacher-student interaction; and (8) student outcomes--immediate effects, long-term effects, and unplanned effects. The first six categories are background, independent teacher-student variables. They result in classroom or process variables which, in turn, result in student outcomes called product or dependent variables. There are a number of issues involved in the validation of measures of teacher effectiveness (teacher's ability to promote student learning). Teacher characteristics may be assessed at five different points: (1) pre-existing characteristics (knowledge, skills, attitudes) necessary for success in teacher education curricula and as a teacher; (2) teacher competence; (3) teacher performance within a particular classroom setting; (4) student learning experiences; and (5) student

learning outcomes. The validity of measures of teacher competence depends on the content validity of measures of student learning outcomes used and how well outcomes attributable to teachers are separated from outcomes due to context factors (Medley, 1985; Ornstein, 1983).

Teacher Performance and Student Learning

Using student performance on achievement tests as a measure of teacher competence derives from four propositions:

1. An important function of teaching is to enhance student learning.
2. Many factors affect student achievement, including teacher performance, achievement measures used, and student characteristics.
3. Students differ in level of past achievement, ability, and willingness and opportunity to learn. These individual student differences must be taken into account.
4. Teacher evaluation can be formative or summative. Knowledge about student achievement can be useful for either purpose.

These propositions are summarized in the following statement:

> If one believes that one goal of teaching is to enhance learning, that is, to help students acquire facts, understand concepts, and achieve skills, and if one also believes that learning occurs, in part, as the result of a teacher's effort, then the improvement of teaching is most apt to occur when connections can be made between teaching and learning (Millman, 1981, p. 147).

Despite the difficulties identified with evaluating the effectiveness of teachers and the various reports questioning whether or

not teachers make a difference in pupil learning, research has continued to attempt to link teacher performance ratings to student learning. Teaching can be defined as the art of selecting and organizing stimuli to achieve desired responses of the learner, and it can be observed and measured in the behaviors of the teacher and the students, in interactions between the teacher and the students, and in the effectiveness of the teacher in bringing about changes in student behavior (Coleman et al., 1966; Ornstein, 1986b).

Teacher effectiveness criteria

Teachers most effective in improving student achievement are clear in the expression of their ideas, variable and flexible in their approaches to teaching, enthusiastic, and task-oriented.

> Relationships are not strong, and one would not expect them to be. The relationships between what a teacher does and the products of teaching is much less strong than the relationship between the level of skill of a craftsman and the excellence of the product. Craftsmen have full control over their tools and materials, but teachers do not (Travers, 1981, pp. 19-20).

Medley (1979) noted, "The effect of schooling on the individual pupil depends to a considerable extent on who his teacher is" (p. 11). He identified five characteristics of an effective teacher: (1) possessor of desirable personal traits; (2) user of effective methods; (3) creator of a good classroom atmosphere; (4) master of a repertoire of competencies; and (5) professional decision maker who has learned when and how to utilize the mastered competencies.

A survey of school districts in northern California following the passage of the Stull Act, which mandated teacher evaluation, determined

that there was agreement among school personnel and board members that all certified personnel should be evaluated, and that student performance should be the main criterion for evaluating teacher effectiveness. How much students learn is related to their opportunity to learn. When teachers stress achievement in academics, expect their students to master the curriculum, and allocate most of their available instruction time to curriculum-related activities, student achievement is maximized (Berliner, 1984; Ferrera, 1974).

Berliner (1980) discusses factors related to effective classroom teaching such as allocated time, engaged time, success rate, academic learning time, opportunity to learn, content coverage, curriculum-test congruence, and direct instruction. These are seen as important variables which are able to discriminate between more and less effective teachers.

In their search for effective classroom practices, Cooley and Leinhardt (1978) extensively studied the variable "opportunity to learn." This variable overlapped with allocated and engaged time measures noted by Berliner (1980). They concluded:

> . . . the major generalization with respect to classroom processes must be that the most useful construct in explaining achievement gain is the opportunity that the children had to learn the skills assessed in the achievement test (p. 32).

Teaching is a cognitive activity requiring a high level of competence for making decisions in complex and dynamic environments. Four factors which can be controlled or influenced by teachers are known to affect student behavior, attitude, and achievement: (1) Pre-

instructional factors; (2) During-instruction factors; (3) Climate factors; and (4) Post-instructional factors (Berliner, 1984).

Pre-instructional factors

Many complex decisions must be made before instruction takes place. Certain decisions facilitate or retard achievement, affect student attitude, and impact student behavior. Among these are content decisions, time allocation decisions, pacing decisions, grouping decisions, and decisions about activity structures. The mean number of hours of schooling per year is significantly related to achievement in mathematics and reading, and content covered is positively related to achievement. Students taught with structured curricula do better than those taught with individualized or discovery approaches, and students receiving direct instruction achieve more than those expected to learn new material or skills on their own or from each other (Berliner, 1984; Rosenshine, 1976; 1982).

The most consistently replicated findings in the study of relationships between teacher performance and pupil learning link pupil achievement to the quantity and pacing of instruction. There is a balance of consideration between maximizing coverage of content by pacing students through the curriculum as quickly as possible and the needs to progress in small or moderate steps; monitor student practice of a new learning until they reach mastery; and see that students mesh new learning with previously learned concepts and skills and apply it to new situations. The pacing of a class will depend upon pupil ability and

developmental levels, the nature of the subject matter, student-teacher ratio, and the teacher's instructional and classroom management skills. Teachers should maximize student success by choosing tasks students are capable of mastering and by clearly explaining the tasks before allowing independent practice. The more challenging the task, the more teachers must monitor and provide help to students needing it (Brophy & Good, 1986).

Content relevant activity was studied in first year high school algebra courses in the Austin (Texas) Independent School District. The California Achievement Test was used to adjust raw posttest scores through analysis of covariance. The covariance-adjusted scores were used as the criterion of teacher effectiveness. The study concluded that a high degree of classroom activity relevant to lesson content positively influenced achievement (Smith, 1979).

Walberg (1988) lists nine educational productivity factors. Two of these relate to the amount of time (quantity) students engage in learning and the instructional experience (quality) including method and curricular aspects or content. He cites two ancient adages about learning which illustrate these two factors: "Time matters; practice makes perfect. Content or subject matter also matters; students learn what they do while learning" (p. 26).

<u>During-instruction factors</u>

When teachers are interacting with students, many factors affect whether or not learning will take place. A few of these factors are

engaged time, time management, monitoring success rate, academic learning time, monitoring, structuring, questioning, wait time, and closure. Studies were examined where teachers implemented training to increase student academic achievement and where students actually had higher achievement. Several teaching functions were found to promote learning: (1) daily review, checking previous day's work, reteaching if necessary, and checking homework; (2) providing overviews in new content/skills, giving detailed or redundant explanations, proceeding in small steps, and phasing in new skills while old skills were being mastered; (3) high frequency of questions and overt student practice, prompting, and feedback with student response; (4) giving feedback/correctives, simplifying questions, giving clues, explaining, and reviewing; (5) providing time for independent practice and seatwork until mastery is achieved; and (6) providing weekly and monthly reviews and reteaching if necessary (Berliner, 1984; Rosenshine, 1982).

A study was conducted wherein the classes of sixty-eight teachers were observed (thirty-nine English, twenty-nine mathematics). The mathematics classes observed supported relationships among teaching behaviors and student outcomes, while the English class results were less clear. In mathematics, it was noted that more effective teachers were active, well organized, and strongly academically oriented; emphasized whole-class instruction while devoting some time to seatwork; managed classrooms efficiently; asked many questions; were more energetic, enthusiastic, nurturant, and affectionate than their less successful colleagues; used praise often; and encouraged and accepted student

contributions (Evertson, Anderson, Anderson, & Brophy, 1980).

Two studies investigated the effects of variation in teacher wait time on science achievement. Achievement was found to be significantly related to teacher variables and pupil engagement when teacher wait time (TWT) and question quality were manipulated. Pupil attending and generalizing were enhanced by a mean TWT of three seconds when questions were relevant, clear, and at varied cognitive levels. The general conclusion of the studies was that TWT produces increases in student science achievement (Tobin, 1980; Tobin & Capie, 1982).

Climate factors

A winning school climate is a common characteristic of effective schools. Climate is a combination of beliefs, values, and attitudes shared by all who play an important role in the school.

> When the climate is right, people are inspired to do their best. Teachers and students--supported by parents and others--do what needs to be done to stimulate learning. Achievement generally rises. Individuals succeed. The school succeeds. The community thrives . . . and so does our nation (Sweeney, 1988, p. 1).

Climate describes the characteristics of the classroom that appear to lead to achievement. Those which seem to be particularly important are communicating academic expectations for achievement, developing a safe, orderly, and academically focused environment for work; sensible management of deviance; and developing cooperative learning environments. Ten characteristics of excellent teachers that would appear especially to contribute to a positive classroom climate have been noted: supportive family background, strong personal faith, enthusiasm for teaching, self-

confidence, communicative skills, Socratic in approach, warmth, concern for students, avoidance of failure in students, and professionalism (Berliner, 1984; Van Schaack, 1982).

In a study comparing the teaching behaviors and characteristics of groups of more effective and less effective teachers of seventh and eighth grade mathematics, the more effective teachers expressed higher expectations for pupils, gave more academic encouragement, exhibited stronger management skills, had more efficient transitions, and were more receptive to student input. They manifested less anxiety, more confidence, greater task orientation, and more enthusiasm (Evertson, Emmer, & Brophy, 1980).

Another study found that teacher efficacy is a critical variable in teacher and school effectiveness. Personal teaching efficacy or level of confidence in personal teaching ability and teaching efficacy or general expectation of student success both correlated significantly with language and mathematics learning outcomes (Tracz & Gibson, 1986).

Time provided for academic activities is not always spent engaged in those activities. Since engaged time is related to student achievement, factors affecting this variable would also indirectly affect achievement. Rates of engaged time depend upon the teacher's ability as a classroom manager. Key effective classroom management indicators are listed as good classroom preparation, installing classroom rules and procedures at the beginning of the year, "with-it-ness" and overlapping in student interaction, smoothness and momentum in pacing of the lessons, appropriate challenge level and variety in assignments, consistent

procedures of accountability along with feedback especially concerning seatwork, clarity about how and when students can seek and receive help, and available options when students finish their work (Brophy, 1986; Brophy & Good, 1986).

Rosenshine (1976) reported that student inattention was negatively related to achievement while results for student attention or on-task behavior were positive but not as high or consistent as those for inattention. He also noted positive correlations for students working in groups as long as they were being supervised.

In discussing the difference that schools and teachers can make in student achievement, Brookover, Beady, Flood, Schweitzer, and Wisenbaker (1979) stated:

> . . . the more the teachers and administrators believe that their students, regardless of race and family background, are capable of higher achievement, and the more this belief is translated into real and observable classroom and school behavior, the higher the resulting mean achievement is likely to be (p. 134).

Post-instructional factors

Many teaching practices typically occur after an instructional episode is completed. Four of those which relate to student achievement are tests, grades, feedback, and evaluation. There is substantial and statistically significant agreement that positive reinforcement and feedback is one of five broad teaching constructs positively associated with higher student learning outcomes. Fair and consistent feedback on student performance appears to relate to student achievement. Classroom observations also reveal that students gain in reading skills when

teachers spend more time discussing homework, providing supportive feedback, checking for understanding, and evaluating and summarizing (Barnes, 1981; Berliner, 1984; Stallings, 1981; 1985; Waxman & Walberg, 1982).

Interrelated factors of instruction

The most consistently replicated findings in effective teaching research demonstrate a link between quantity and quality of time available for instruction and student learning. The major intervening variable between teacher behavior and student achievement is academic learning time (ALT)--the amount of time students spend engaged in academic tasks they can perform with relatively high success rates. Research reports that large increases in ALT has been found to be a powerful correlate of student learning gains. Studies of ALT point to certain teacher behaviors and practices which increase time available for instruction and student engagement in academic work. These practices and processes include classroom management and organization strategies, instructional strategies, and strategies that communicate high expectations for student performance (Brophy, 1987; Evertson, 1986; Fisher, Filby, Marliave, Cohen, Dishow, Moore, & Berliner, 1978; Seifert & Beck, 1984; Stallings, 1980).

Rosenshine and Furst (1971) listed eleven teacher behavior variables which showed the strongest evidence of a relationship to student achievement. The variables were teacher clarity, variability, enthusiasm, task-oriented behaviors, student opportunity to learn

criterion material, use of student ideas and general indirectness, criticism, use of structuring comments, types of questions, probing, and level of difficulty of instruction. They found the best results were obtained on the first five variables. Several other research efforts have studied these same teacher variables and reached similar conclusions (Borich & Kash, 1979; Borg, 1979; Dangel & Hopkins, 1979; Good & Grouws, 1977; Martin, 1979; Mintzes, 1982; Otto & Schuck, 1983; Ouzts, 1986; Smith, 1977). In a study by Coker, Medley, and Soar (1980), teachers and expert consultants developed a list of teacher performance competencies and measured the competencies in one hundred classrooms in a school system over a two-year period. They found that basically the same eleven teacher behaviors were related to student achievement after the results were analyzed.

Relationships between teacher clarity, the variable at the top of Rosenshine and Furst's list, and outcome measures of student achievement were examined. Measures of teacher clarity were positively related to postinstructional measures of student achievement. A number of specific clarity behaviors were identified that appeared to be strongly linked to desirable student outcomes. Some of note were the use of relevant examples during explanation, questioning for understanding, repeating things students misunderstood, providing time for practice, explanation and stopping for student thinking, and teaching at an appropriate pace (Hines, Cruickshank, & Kennedy, 1985).

An important finding concerning effective schools and classrooms is that students' classroom behavior is the most direct link to student

achievement. Another finding is that teachers' behavior can affect students' behavior in ways leading to improved student learning. Teachers have impact on student behavior and achievement by planning, managing, and instructing in ways that keep students involved and successfully covering appropriate content (Squires, Huitt, & Segars, 1984).

In studies which employed the Teacher Performance Assessment Instrument, positive relationships were found between student achievement and ratings on the following performance indicators: (1) uses teaching methods appropriate for objectives, learners, and environment; (2) uses instructional materials that provide learners with appropriate practice on objectives; (3) demonstrates ability to work with individuals, small groups, and large groups; (4) maintains appropriate classroom behavior; and (5) manages disruptive behavior among learners (Gage & Giaconia, 1981; Tobin, 1986).

A synthesis of the research on teaching would include characteristics of effective teachers along with characteristics of effective instruction. The characteristics of effective teachers are (1) clear instructional goals; (2) knowledge of subject matter and teaching strategies; (3) effective communication of expectations to students; (4) adapt instruction to student needs; (5) teach higher order thinking skills as well as knowledge level cognitive objectives; (6) incorporate other subject areas into their instruction; (7) accept responsibility for student outcomes; and (8) are reflective practitioners. The characteristics of effective instruction include: (1) Teachers promote

learning by communicating to their students what is expected and why; (2) Teachers promote learning by providing students with strategies for monitoring and improving their own learning efforts and with structured opportunities for independent learning activities; and (3) Effective teachers know the misconceptions students bring to the classroom that might interfere with their learning of the proper subject matter, as well as know the subject matter they intend students to learn (Porter & Brophy, 1988).

Most of the research on teaching can be summarized by four broad characteristics of effective teachers:

1. Effective teachers accept responsibility for student learning. They are more effective than teachers who believe students' family backgrounds are responsible for what students learn.
2. Effective teachers motivate student learning by communicating their expectations for student learning, by monitoring student understanding, by encouraging independent learning, by modeling, and by providing group and individual learning experiences.
3. Effective teachers have a strong knowledge of subject matter, adapt their instruction according to student needs and situation, and see teaching as involving the inducement of change in an existing body of knowledge and beliefs.
4. Effective teachers choose, adapt, and use materials and resources effectively (Brophy, 1987; Brophy & Good, 1986; Duttweiler, 1988).

Student Achievement Measures

The inclusion of student achievement in appraising teacher performance has been mandated in several states. These states and many others are conducting studies to identify promising ways to examine student achievement. In making decisions about teachers advancing on a career ladder or receiving performance-based pay, many states and districts are progressing toward using student achievement as a major factor in determining awards. To accomplish this goal, attempts are under way to document student achievement that exceeds expected growth (Career Ladder Clearinghouse, 1988).

Outcomes-based teacher evaluation

Product models for teacher evaluation based on results or outcomes of instruction have been proposed. Emphasis of such models is not on the methods, styles, or processes, but on the results of achievement tests. It has been suggested that since we cannot prove any one method, style, or process of teaching superior to any other, then all we can do is go by results. Since the business of the teacher is the promotion of student learning, the assessment of teacher effectiveness by measuring changes in student achievement over a previously specified period of instruction would appear feasible. This period could be as short as a single lesson or as long as a school year. Under the assumption that the most important function of teaching is to enhance student learning, there is an inherent logic in using student performance data to assess teachers, even though there are also significant problems (Borich, 1977; Feldvebel,

1980; McGreal, 1983).

Measure of student gain

In considering the measurement of change, there are several interesting psychometric attributes of change scores. The ceiling effect limits the range of difficulty of test items. Students scoring 95 percent on a pretest cannot improve as much on a posttest as students scoring 45 percent on the pretest. Change or gain scores assume equal intervals at all points of the test. Using the above example, gain scores equate student pretest-posttest progress of 45 to 48 percent with student progress of 95 to 98 percent when these gains are not likely to reflect equal student learning. Other attributes of gain scores worthy of mention are regression toward the mean and unreliability due to high pretest-posttest correlation. There are, however, statistical procedures such as partial correlation, analysis of covariance, and multiple regression which are able to overcome some of the limitations mentioned. When parallel forms of a test are used as pretest and posttest or when it is necessary to compare performances within and across grade levels, scaled scores, which possess the property of equal intervals, should be used (Anderson, 1954; Berk, in press; Borg & Gall, 1983).

There are several possible difficulties in developing a system of teacher evaluation based on pretest-posttest class gain scores. The gain score model has some sources of invalidity and measurement error. Between-class, between-grade, and between-subject variability of objectives, instruction, resources, and student characteristics; teaching

to the examinations; and regression effects preclude the trouble-free selection of appropriate achievement tests, precise estimation of gain, the setting of a meaningful criterion for superior teacher productivity, and the inference that estimated gain is attributable solely to teacher effort (Berk, 1984; Schein, 1985; Soar, Medley, & Coker, 1983).

The pretest-posttest gain score use as a measure of student achievement and, thus, teacher effectiveness does have potential difficulties. However, by using statistical methods to minimize these problems and by keeping the caveats in mind, student gain scores can provide valuable insight into teacher effectiveness. In using pretest-posttest gain scores to measure pupil learning, pretest scores have been found to account for most of the variance in the posttest scores. Therefore, when analyses of achievement used pretest scores as covariates, statements about relationships between teacher performance criteria and student achievement actually refer to the degree to which posttest scores are predictable from the teacher performance criteria beyond what is predictable from the pretest scores (Evertson, Anderson, Anderson, & Brophy, 1980; Rose & Medway, 1981).

Popham (1975) reported the most widely employed index of teaching skill to be the rating. Systematic observations were second where judgments from good to bad were not necessary, only an assertion whether a specific behavior occurred. The third most widely used index of teaching skill was found to be pupil performance on tests, typically achievement tests. Until recent years, norm-referenced tests were used in such investigations. However, with the emergence of criterion-

referenced measures, we may see substantial differences in the way pupil test performance is used as an index of teaching performance.

Several problems with norm-referenced tests render them unsuitable for purposes of teacher evaluation. They tend to be (1) off-target from local curricular emphases; (2) inadequately described for purposes of instructional amelioration; and (3) unlikely to contain a sufficient number of items measuring important concepts, such items having been eliminated due to student tendencies to respond correctly yielding insufficient variance. Therefore, the bulk of studies where norm-referenced measures have been employed as dependent measures have failed to show they are serviceable. "Nationally normed they may be, sensitive to instruction they are not" (Popham, 1975, p. 291).

Criterion-referenced tests represent substantial improvement as indicators of teacher effectiveness. Logically, such tests appear to be more sensitive to instruction than their norm-referenced counterparts. Thus, they would seem to hold more promise for tapping a teacher's effects on students (Popham, 1975).

Summary

Teaching is a complex and contextual activity, and the evaluation of teacher performance is subjective under the best conditions. There are difficulties with the measurement of student achievement. However, this review of literature has noted some steps, adjustments, and techniques that can be used to minimize or at least to lessen the concerns associated with teacher evaluation and student achievement.

When evaluators of teacher performance are required to share ratings with teachers, they are reluctant to create potential problems for future interactions with the teachers and often assign lenient ratings due to this reluctance. Keeping the results of the ratings confidential has been shown by research to motivate them to assign ratings that are more accurate.

A multitude of teacher effectiveness criteria have been suggested by the research literature. The SIM Teacher Performance Evaluation instrument addresses twenty-five of these in the following format:

I. Productive Teaching Techniques

(1) Demonstrates ability to inspire and to motivate students.
(2) Communicates effectively with students.
(3) Uses a variety of appropriate evaluation methods which provide students with specific feedback.
(4) Prepares appropriate evaluation feedback.
(5) Provides opportunities for individual differences.

II. Organized, Structured Classroom Management

(6) Effectively uses available materials and resources.
(7) Demonstrates evidence of effective planning and organization.
(8) Manages student behavior in a constructive manner.
(9) Organizes students for effective instruction.

III. Positive Interpersonal Relations

(10) Demonstrates effective interpersonal relationships.
(11) Promotes positive self-concept.
(12) Demonstrates sensitivity in relating to students.
(13) Promotes self-discipline and responsibility.

IV. Professional Responsibilities

(14) Is involved with the accomplishment of the district and building goals.

V. High Gain Teacher Descriptors

(15) Spends time at the beginning of the learning demonstrating

processes to the students (cueing).
(16) Uses more controlled (guided) practice before assigning homework (independent practice).
(17) Monitors seatwork closely.
(18) Plays a key role in modeling and giving concrete examples.
(19) Has higher expectations.
(20) Is pleasant, but not affectively extreme.
(21) Has more energy.
(22) Plans better.
(23) Wants more feedback from students.
(24) Wants more feedback from supervisors and principals.
(25) Moves more quickly through the curriculum.

Despite the difficulties involved with measures of student gain, several techniques exist to minimize the detrimental effects. Pretests and posttests measured with standard scaled scores produce equal intervals on the measurement scale and allow comparisons within and across classes and school organizations. Measuring achievement on a posttest with the pretest as a covariate provides a measure that avoids the noted shortcomings of using raw gain scores. Also, procedures utilizing the full range of scores on both independent and dependent variables maintain the variation in both measures and, thus, provide more accurate results than the grouping of one or both variables into ordinal groups of data. The multiple regression procedure or the Pearson product-moment correlation procedure using adjusted posttest scores provide these characteristics.

Concerning the type of achievement tests used, criterion-referenced tests are typically district specific, and they are often locally constructed. As such, they are more sensitive to the objectives being taught than are the external norm-referenced tests and are more likely to identify teacher effects on students. They allow teachers to test what

they teach and teach what they test.

The patterns followed in the most recent research studies on teacher effectiveness have concentrated on assessing student achievement in reading and mathematics. These two subject areas are usually measured at one elementary grade level and one secondary grade level. The conventions and suggestions covered in this summary of the review of literature were incorporated into the present investigation of the relationships between teacher performance ratings and pupil learning.

CHAPTER III. METHODS AND PROCEDURES

The School Improvement Model (SIM) project was a total-systems approach for improving classroom performance, schools, and entire school organizations. The four year project was centered in a consortium of five school organizations, the Minneapolis Public Schools, Edina Public Schools, Northfield Public Schools, and the Breck School (private), all in Minnesota, and the Spirit Lake (Iowa) Community Schools. The Northwest Area foundation of St. Paul, Minnesota, provided a planning grant and a subsequent, four-year research award. The consortium schools devoted the time and staff development activities of their teachers and administrators and the learning measures of their students from July 1, 1980, through August 1, 1984.

The SIM project was designed to demonstrate the effect of an articulated system of administrator and teacher performance appraisal, coupled with staff development interventions to encourage productive behaviors, on student learning in reading and mathematics at both elementary and secondary school levels as measured by both norm-referenced and criterion-referenced tests. Teachers and administrators were taught how to do their jobs better. They were coached for better performance by appraisers using state-of-the-art appraisal tools and skills. Then, pupil learning was measured for any improvement.

Teacher Performance Evaluation

The objectives of the SIM project were to measure and improve the performance of both teachers and administrators. The primary purpose of

schools is instruction, and student-learner outcomes are the major determinant of the effectiveness of such instruction. School improvement and also teacher effectiveness, in the SIM project, were identified by improved student-learner outcomes. For this reason, it was important that changes in teacher performance be documented throughout the final years of the project.

Teacher performance was measured following procedures and criteria developed and adopted by locally appointed stakeholder groups. Each of the five school organizations selected some teacher performance criteria that were the same, or similar, and some that were quite different. Procedures also differed somewhat according to school organization size and local needs and objectives. Although SIM only measured student achievement in fourth grade reading and mathematics and eighth grade mathematics, all teachers were eventually to be evaluated using the newly developed system.

Student Achievement Testing

SIM, as did most recent school reform movements, defined an effective school as a school in which student achievement could not be predicted on the basis of gender, race, or socioeconomic status. Another commonality with recent school reform studies was that the program assessing student learning was established at the fourth grade in reading and mathematics and the eighth grade in mathematics. Teachers and administrators at all grade levels were involved in various school improvement activities. However, for economic reasons, only two grades

and subject areas were tested. Students were administered both pretests and posttests since student achievement gain was of primary interest. External, standardized norm-referenced tests and internally developed criterion-referenced tests were both used for each subject at each grade level.

Norm-referenced tests were selected by each school organization individually. Mostly, these tests were already being administered in either fall or spring. For their norm-referenced measures, Minneapolis chose the California Achievement Test; Northfield selected the Comprehensive Test of Basic Skills; Edina and Spirit Lake used the Iowa Test of Basic Skills; and Breck used the Comprehensive Testing Program along with norms for independent schools.

For the criterion-referenced measures, the SIM project consultants worked with the teachers of the selected grade levels and subject areas to create and refine the criterion-referenced tests to assure that the desired learning objectives were being tested. Teachers were asked to choose objectives for each test that they considered to be the most important of the year and for which a majority of the students would receive first-time instruction during that year. Test items were chosen from the Instructional Objectives Exchange (IOX) item bank developed at the Center for Evaluation at UCLA or were written by teachers in the school organization. Items for some of the mathematics tests were selected from the Project MEASURE item bank made available by the Northern Trails Area Education Agency in Clear Lake, Iowa. Each test measured approximately twenty objectives, and a general policy of

utilizing five test items per objective was followed.

The criterion-referenced tests were piloted during the 1982-83 school year, the third year of the project. Subsequent to this pilot, significant revisions of the tests in terms of objectives measured and items on the tests were accomplished. All test items and formats were critiqued and improved by SIM consultants and then pilot-tested a second time before their use in 1983-84.

Both norm-referenced tests and criterion-referenced tests were administered with pretests in September and posttests in May of the 1983-84 school year. Coordinators of the SIM project provided logistical assistance to and observed the test administration process.

Relating Teacher Performance and Student Achievement

The problem for the present study was to determine if there was a relationship between supervisor ratings of teacher performance and student achievement during the fourth year of the Northwest Area Foundation's School Improvement Model project. Performance criteria selected were those which have been identified by recent research as being related to teacher effectiveness. Student achievement was measured in fourth and eighth grade mathematics and fourth grade reading by using pretests and posttests, both norm-referenced and criterion-referenced. The following research questions were posed to accomplish the primary task of ascertaining whether such a relationship between teacher performance and student achievement actually existed:

a. Is there a relationship between the supervisor ratings of the

individual teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading?

b. Is there a relationship between the mean supervisor ratings of the teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading?

c. Is there a relationship between the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas and student achievement in fourth and eighth grade mathematics and fourth grade reading?

d. Is there a difference in the supervisor ratings of the individual teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested?

e. Is there a difference in the mean supervisor ratings of the teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested?

f. Is there a difference in the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested?

Identification of the Research Subjects

This investigation is a small component of the much larger SIM project which comprised a consortium to form a school improvement model

linking staff development and program development with improved student achievement. The five school organizations of Minneapolis, Edina, Northfield, Breck, and Spirit Lake were chosen to cover the broad spectrum of school environments: urban, suburban, college community, private, and rural.

During the 1983-84 school year, students in the Northwest Area Foundation SIM project were pretested and posttested in fourth grade reading and mathematics and eighth grade mathematics with both norm- and criterion-referenced tests. Pretests were administered in September and posttests in May. Approximately eighty classes and, therefore, eighty teachers were involved from the five school organizations. Of these eighty, complete data sets of student achievement scores and summative teacher evaluation ratings existed for fifty-three teachers. These fifty-three teachers were selected and analyzed for the present study. The remaining twenty-seven teacher data sets were lost due to teacher turnover, transfers, and/or retirements of evaluators, and school closings in Minneapolis. Furthermore, Breck school chose not to participate in this part of the study.

Collection of Data

In terms of student achievement results, class means on both pretests and posttests, norm- and criterion-referenced, in fourth grade reading and mathematics and eighth grade mathematics were available for all fifty-three teachers. Also available were the grand mean and standard deviation for each grade level and subject tested. In addition

to the student achievement results, the SIM Teacher Performance Evaluation instrument (Appendix A) was used for the present investigation. Summative evaluations were completed for all fifty-three teachers by their respective supervisors after the close of the 1983-84 school year. The SIM Teacher Performance Evaluation instrument is a research-based evaluation instrument which measures teacher performance on twenty-five criteria (Manatt & Stow, 1984). Each of the criteria of teacher performance is rated on a continuum from 1 (low performance) to 7 (high performance):

1.0 - 2.9 = Needs Improvement

3.0 - 4.9 = Meets Standards

5.0 - 7.0 = Exceeds Standards.

It should be noted that the teacher performance ratings were completed by the supervisors of the fifty-three teachers in confidence and retained by the SIM staff. The results of these summative ratings were not shared with the project teachers and, therefore, inaccuracies due to leniency on the part of the supervisors should have been minimized.

Methods of Statistical Treatment

Development of the data base

Before statistical tests could be conducted for each of the research hypotheses, an appropriate data base had to be developed from which the tests could be performed. The graphic responses for each of the criteria on the SIM Teacher Performance Evaluation rating instrument were first converted to a numeric mode which was on a continuum from 1.0 to 7.0 in

increments of 0.1. Each of the fifty-three teachers, therefore, had numeric ratings on all twenty-five performance criteria. A mean rating was also calculated for each teacher. The teachers were not grouped by district, gender, or by any other criteria. Each teacher was assigned an identification number from 1 to 53.

As mentioned previously, the five school organizations used four different norm-referenced tests, and the criterion-referenced tests used were locally-constructed and district-specific. In order to compare the results within and across teachers' classes and school organizations and to insure equal intervals on the measurement scale, the raw pretest and posttest class means, both norm- and criterion-referenced, were converted to standard scores (Z-scores) which had a mean of 50 and standard deviation of 10. Additionally, each posttest class mean was adjusted by removing the pretest effect according to the formula:

$$\bar{Y}_{j(adj)} = \bar{Y}_j - b(\bar{X}_j - \bar{X})$$

where $\bar{Y}_{j(adj)}$ is the adjusted posttest class mean, $\bar{Y}_j$ is the posttest class mean, b is the regression coefficient, $\bar{X}_j$ is the pretest class mean, and $\bar{X}$ is the pretest grand mean (which was 50).

After these conversions were completed, the teacher performance criteria ratings, the mean teacher performance rating, the standard pretest scores, the standard posttest scores, and adjusted posttest scores were entered into the WYLBUR system at Iowa State University. The Statistical Package for the Social Sciences, SPSS^x (Norusis, 1983) was used to test the research hypotheses.

Statistical analysis

Appropriate tests of statistical significance were selected in order to test the null hypotheses presented in this study. The specific tests chosen to address each of the hypotheses are the following:

Hypothesis a. There will be no relationship between the supervisor ratings of the individual teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading.

Hypothesis b. There will be no relationship between the mean supervisor ratings of the teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading.

Hypothesis c. There will be no relationship between the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas and student achievement in fourth and eighth grade mathematics and fourth grade reading.

Hypotheses a, b, and c were tested by using stepwise multiple regression. Separate tests were run for each of the twenty-five performance ratings, the mean performance criteria rating, and the mean ratings of the four clusters of performance criteria. These tests were run for fourth grade mathematics with both norm- and criterion-referenced tests, for eighth grade mathematics with both norm- and criterion-referenced tests, and for fourth grade reading with both norm- and criterion-referenced tests.

It has been well-established that pretest scores account for most of the variance in posttest scores, so the present study was concerned with relationships between teacher performance ratings and posttest scores after the effects of the pretest had been removed. Thus, each multiple regression test used the pretest score as one independent variable and

the individual performance criteria ratings, mean performance rating, or mean cluster rating as the second independent variable. The posttest score was, in each case, the dependent variable. The multiple regression tests determined whether the performance ratings contributed to the prediction of the posttest scores given the importance of pretest scores.

In addition to the multiple regression procedure, the Pearson product-moment procedure was used with the performance ratings and the adjusted posttest scores to determine whether a relationship existed between the two. These two procedures, since they are very similar, should yield similar results. With either, the effect of the pretest had to be removed before any conclusions could be made about the predictive value of the performance ratings as they related to the posttest results.

Hypothesis d. There will be no difference in the supervisor ratings of the individual teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested.

Hypothesis e. There will be no difference in the mean supervisor ratings of the teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested.

Hypothesis f. There will be no difference in the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested.

In testing Hypotheses d, e, and f, the single classification analysis of variance procedure was used with the dependent variable being the achievement group (low, middle, or high) and the independent variable

being the adjusted posttest scores for the various norm- and criterion-referenced tests administered. The Scheffé multiple comparison procedure was used for the pairwise comparison of the means when a difference was found.

In testing Hypotheses c and f, clusters of performance criteria grouped into logical performance areas were used rather than the twenty-five individual performance criteria. Four such logical performance areas were identified. The areas and the clusters of performance criteria in each area are as follows:

Area I. Productive Teaching Techniques

1. Demonstrates ability to inspire and to motivate students.
2. Communicates effectively with students.
3. Uses a variety of appropriate evaluation methods which provide students with specific feedback.
4. Prepares appropriate evaluation feedback.
5. Provides opportunities for individual differences.
15. Spends time at the beginning of the learning demonstrating processes to the students (cueing).
16. Uses more controlled (guided) practice before assigning homework (independent practice).
18. Plays a key role in modeling and giving concrete examples.
19. Has higher "expectations."
22. Plans better.
25. Moves more quickly through the curriculum.

Area II. Organized, Structured Classroom Management

6. Effectively uses available materials and resources.
7. Demonstrates evidence of effective planning and organization.
8. Manages student behavior in a constructive manner.
9. Organizes students for effective instruction.
17. Monitors seatwork closely.

Area III. Positive Interpersonal Relations

10. Demonstrates effective interpersonal relations.
11. Promotes positive self-concept.
12. Demonstrates sensitivity in relating to students.
13. Promotes self-discipline and responsibility.

20. Is pleasant, but not affectively extreme.

Area IV. Intellectual Stimulation and Desirable Out of Class Behavior

14. Is involved with the accomplishment of the district and building goals.
21. Has more energy.
23. Wants more feedback from students.
24. Wants more feedback from supervisors and principals.

In carrying out the statistical analysis of the data, the following steps were sequential. First, the individual performance criteria were tested for relationships with student achievement. Secondly, the mean performance ratings were tested for possible correlations. The next step tested the mean ratings on the logical performance areas for relationships with pupil learning. Finally, the mean ratings on individual performance criteria, mean overall performance ratings, and mean ratings on the logical performance areas were tested for differences among the bottom, middle, and top achieving classes as determined by average adjusted posttest scores.

CHAPTER IV. ANALYSIS AND RESEARCH FINDINGS

The primary purposes of this study were to determine whether supervisor ratings of teachers can be used to predict student achievement and to determine which teacher performance criteria and which logical groupings of teacher performance criteria clustered into performance areas are related to higher student achievement in fourth grade reading and mathematics and eighth grade mathematics. This chapter analyzes the data base developed for the fifty-three teachers who were the subjects of the study. The data base consisted of teacher identification numbers (1 to 53), teacher performance criteria ratings for each of the twenty-five criteria, standard pretest scores, standard posttest scores, and adjusted posttest scores for each of the fifty-three teachers.

Testing of the Hypotheses

To answer each of the research questions presented in this study, a specific hypothesis was stated in the null form. All of these null hypotheses were tested at the .05 level of significance. The hypotheses are here presented and discussed in the order of the research questions from Chapter I.

Hypothesis a. There will be no relationship between the supervisor ratings of the individual teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading.

Fourth grade mathematics: Criterion-referenced tests

The hypotheses that the individual teacher performance criteria do not contribute to the prediction of posttest scores on fourth grade

mathematics criterion-referenced tests after the effect of the pretest has been removed was tested using the forward stepwise multiple regression procedure. A t-value is given for the test of the coefficient of the independent variable (in this case, the performance criteria rating) being 0.

Of the twenty-five teacher performance criteria rated, thirteen were found to be significant predictors of posttest scores at the .05 level after the effect of the pretest had been removed. Two of the criteria--9. Organizes students for effective instruction, and 22. Demonstrates effective planning--were highly significant at the .01 level. Table 1 lists the twenty-five performance criteria in order of significance as predictors of posttest scores after pretest effect had been removed.

A regression analysis of the thirteen performance criteria whose ratings were found to be significant predictors of student posttest scores on fourth grade mathematics criterion-referenced tests reveals that the pretest is the best predictor of the posttest score. The pretest was found to account for 51% of the variation in the posttest. When each of the thirteen significant performance criteria was also used as a predictor, it also accounted for an additional percentage of the posttest variation. An equation could be formed using both the pretest score and the individual performance criteria ratings to predict the posttest score. Table 2 shows the additional percentage of posttest variation accounted for by each of the thirteen performance criteria ratings along with the best prediction equation for each criterion.

The Pearson product-moment procedure was used with the teacher

Table 1. Fourth grade mathematics criterion-referenced tests--teacher performance criteria ratings as predictors of student achievement

Teacher performance criteria	t-value
9. Organizes students for effective instruction.	3.042**
22. Demonstrates effective planning.	2.802**
17. Monitors seatwork closely.	2.746*
16. Uses guided practice before independent practice.	2.676*
2. Communicates effectively with students.	2.612*
15. Demonstrates processes at beginning of learning (cueing).	2.565*
8. Manages student behavior in a constructive manner.	2.504*
1. Demonstrates ability to inspire and motivate students.	2.498*
4. Prepares appropriate evaluation feedback.	2.443*
25. Moves quickly through the curriculum.	2.434*
24. Desires feedback from supervisors and principals.	2.319*
6. Effectively uses available materials and resources.	2.233*
10. Demonstrates effective interpersonal relationships.	2.185*
13. Promotes self-discipline and responsibility.	1.982
19. Exhibits high expectations.	1.907
20. Is pleasant, but not affectively extreme.	1.890
3. Uses a variety of evaluation methods with specific feedback.	1.865
5. Provides opportunities for individual differences.	1.840
12. Demonstrates sensitivity in relating to students.	1.721
7. Demonstrates effective planning and organization.	1.696
18. Models and gives concrete examples.	1.672
11. Promotes positive self-concept.	1.664
21. Displays a high energy level.	1.649
23. Desires feedback from students.	1.535
14. Involved with reaching district and building goals.	1.012

*Significant at $p < .05$.

**Significant at $p < .01$.

Table 2. Fourth grade mathematics criterion-referenced tests--regression analysis showing additional posttest variation accounted for and prediction equation for each of the thirteen significant performance criteria ratings

	Percent of variation	Prediction equation
Criterion 9	11.4%	Posttest = .6537 x Pretest + 1.2883 x Criterion 9 + 10.7142
Criterion 3	10.0%	Posttest = .6978 x Pretest + 1.3091 x Criterion 3 + 8.3557
Criterion 17	9.7%	Posttest = .6620 x Pretest + 1.2817 x Criterion 17 + 10.3322
Criterion 16	9.3%	Posttest = .6614 x Pretest + 1.2923 x Criterion 16 + 10.3245
Criterion 2	8.9%	Posttest = .6756 x Pretest + 1.2005 x Criterion 2 + 10.2588
Criterion 15	8.7%	Posttest = .6527 x Pretest + 1.3763 x Criterion 15 + 10.1497
Criterion 8	8.3%	Posttest = .7262 x Pretest + 1.0602 x Criterion 8 + 8.0131
Criterion 1	8.3%	Posttest = .6902 x Pretest + 1.3825 x Criterion 1 + 8.5063
Criterion 4	8.0%	Posttest = .6882 x Pretest + 1.3934 x Criterion 4 + 8.6858
Criterion 25	7.9%	Posttest = .6967 x Pretest + 1.6921 x Criterion 25 + 7.0183
Criterion 24	7.3%	Posttest = .6720 x Pretest + 1.3027 x Criterion 24 + 9.7426
Criterion 6	6.9%	Posttest = .6847 x Pretest + 1.3067 x Criterion 6 + 8.9686
Criterion 10	6.6%	Posttest = .7117 x Pretest + 1.0763 x Criterion 10 + 8.8977

performance criteria ratings and the adjusted posttest scores on the fourth grade mathematics criterion-referenced tests to determine the strength of the relationship between the two. The correlation coefficients are shown in Table 3. As expected, the strength of the relationships is in almost the same order as found with the multiple regression procedure. Performance Criteria 9 and 22 are very significantly related to the adjusted posttest scores, while eleven other of the performance criteria show a significant relationship to the adjusted posttest scores.

Fourth grade reading criterion-referenced tests

The forward stepwise multiple regression procedure was again used with a t-value given for the test of the coefficient of the independent variable (again, the performance criteria rating) being 0. Of the twenty-five teacher performance criteria rated, seven were found to be significant predictors of posttest scores on fourth grade reading criterion-referenced tests at the .05 level after the effect of the pretest had been removed. Three of the criteria--6. Effectively uses available materials and resources, 7. Demonstrates effective planning and organization, and 21. Displays a high energy level--were highly significant at the .01 level. Table 4 lists the twenty-five performance criteria in order of significance as predictors of posttest scores after pretest effect had been removed.

A regression analysis of the seven performance criteria whose

Table 3. Fourth grade mathematics criterion-referenced tests--Pearson product-moment correlations of twenty-five teacher performance criteria ratings with adjusted posttest scores

	Adjusted posttest score
9. Organizes students for effective instruction.	.4506**
22. Demonstrates effective planning.	.4352**
17. Monitors seatwork closely.	.4154*
16. Uses guided practice before independent practice.	.4046*
8. Manages student behavior in a constructive manner.	.4019*
2. Communicates effectively with students.	.4013*
1. Demonstrates ability to inspire and motivate students.	.3907*
15. Demonstrates processes at beginning of learning (cueing).	.3850*
25. Moves quickly through the curriculum.	.3836*
4. Prepares appropriate evaluation feedback.	.3820*
24. Desires feedback from supervisors and principals.	.3601*
10. Demonstrates effective interpersonal relationships.	.3523*
6. Effectively uses available materials and resources.	.3485*
13. Promotes self-discipline and responsibility.	.3106
5. Provides opportunities for individual differences.	.3013
3. Uses a variety of evaluation methods with specific feedback.	.3003
19. Exhibits high expectations.	.2981
20. Is pleasant, but not affectively extreme.	.2923
7. Demonstrates effective planning and organization.	.2769
21. Displays a high energy level.	.2692
12. Demonstrates sensitivity in relating to students.	.2661
18. Models and gives concrete examples.	.2606
11. Promotes positive self-concept.	.2605
23. Desires feedback from students.	.2436
14. Involved with reaching district and building goals.	.1744

*Significant at $p < .05$.

**Significant at $p < .01$.

Table 4. Fourth grade reading criterion-referenced tests--teacher performance criteria ratings as predictors of student achievement

Teacher performance criteria	t-value
21. Displays a high energy level.	3.057**
6. Effectively uses available materials and resources.	2.999**
7. Demonstrates effective planning and organization.	2.811**
3. Uses variety of evaluation methods with specific feedback.	2.212*
9. Organizes students for effective instruction.	2.179*
10. Demonstrates effective interpersonal relationships.	2.106*
2. Communicates effectively with students.	2.065*
25. Moves quickly through the curriculum.	2.027
13. Promotes self-discipline and responsibility.	2.011
8. Manages student behavior in a constructive manner.	2.004
1. Demonstrates ability to inspire and to motivate students.	1.873
23. Desires feedback from students.	1.743
22. Demonstrates effective planning.	1.712
4. Prepares appropriate evaluation feedback.	1.710
19. Exhibits high expectations.	1.519
15. Demonstrates processes at beginning of learning (cueing).	1.447
11. Promotes positive self-concept.	1.386
16. Uses guided practice before independent practice.	1.238
17. Monitors seatwork closely.	1.196
18. Models and gives concrete examples.	1.075
24. Desires feedback from supervisors and principals.	1.064
20. Is pleasant, but not affectively extreme.	0.866
5. Provides opportunities for individual differences.	0.638
12. Demonstrates sensitivity in relating to students.	0.546
14. Involved with reaching district and building goals.	0.406

*Significant at $p < .05$.

**Significant at $p < .01$.

ratings were found to be significant predictors of student posttest scores on fourth grade reading criterion-referenced tests reveals that the pretest is the best predictor of the posttest score. The pretest was found to account for 49% of the variation in the posttest. When each of the seven significant performance criteria was also used as a predictor, it accounted for an additional percentage of the posttest variation. An equation could be formed using both the pretest score and the individual performance criteria ratings to predict the posttest score. Table 5 shows the additional percentage of posttest variation accounted for by each of the seven performance criteria ratings along with the best prediction equation for each criterion.

The Pearson product-moment procedure was used with the teacher performance criteria ratings and the adjusted posttest scores on the fourth grade reading criterion-referenced tests to determine the strength of the relationship between the two. The correlation coefficients are shown in Table 6. As expected, the strength of the relationships is in almost the same order and magnitude as found with the multiple regression procedure. Performance Criteria 21, 6, and 7 are very significantly related to the adjusted posttest scores, while four other of the performance criteria show a significant relationship to the adjusted posttest scores.

Eighth grade mathematics: Criterion-referenced tests

Forward stepwise multiple regression was again used with a t-value given for the test of the coefficient of the independent variable

Table 5. Fourth grade reading criterion-referenced tests--regression analysis showing additional posttest variation accounted for and prediction equation for each of the seven significant performance criteria ratings

	Percent of variation	Prediction equation
Criterion 21	11.8%	Posttest = .7013 x Pretest + 1.1484 x Criterion 21 + 10.0167
Criterion 6	11.5%	Posttest = .6667 x Pretest + 1.3924 x Criterion 6 + 10.3997
Criterion 7	10.3%	Posttest = .7471 x Pretest + 1.1589 x Criterion 7 + 7.4643
Criterion 3	6.9%	Posttest = .7100 x Pretest + 1.0964 x Criterion 3 + 9.6972
Criterion 9	6.8%	Posttest = .6785 x Pretest + .8244 x Criterion 9 + 12.5893
Criterion 10	6.4%	Posttest = .6974 x Pretest + .8879 x Criterion 10 + 11.4377
Criterion 2	6.2%	Posttest = .7053 x Pretest + .8170 x Criterion 2 + 11.4393

Table 6. Fourth grade reading criterion-referenced tests--Pearson product-moment correlations of twenty-five teacher performance criteria ratings with adjusted posttest scores

	Adjusted posttest score
21. Displays a high energy level.	.4739**
6. Effectively uses available materials and resources.	.4589**
7. Demonstrates effective planning and organization.	.4488**
3. Uses variety of evaluation methods with specific feedback.	.3618*
9. Organizes students for effective instruction.	.3489*
10. Demonstrates effective interpersonal relationships.	.3426*
2. Communicates effectively with students.	.3386*
25. Moves quickly through the curriculum.	.3355
8. Manages student behavior in a constructive manner.	.3298
13. Promotes self-discipline and responsibility.	.3264
1. Demonstrates ability to inspire and to motivate students.	.3156
23. Desires feedback from students.	.2965
4. Prepares appropriate evaluation feedback.	.2892
22. Demonstrates effective planning.	.2870
19. Exhibits high expectations.	.2572
15. Demonstrates processes at beginning of learning (cueing).	.2439
11. Promotes positive self-concept.	.2332
16. Uses guided practice before independent practice.	.2115
17. Monitors seatwork closely.	.2024
18. Models and gives concrete examples.	.1832
24. Desires feedback from supervisors and principals.	.1806
20. Is pleasant, but not affectively extreme.	.1475
5. Provides opportunities for individual differences.	.1092
12. Demonstrates sensitivity in relating to students.	.0917
14. Involved with reaching district and building goals.	.0698

*Significant at $p < .05$.

**Significant at $p < .01$.

(again, the performance criteria rating) being 0. Of the twenty-five teacher performance criteria rated, eight were found to be significant predictors of posttest scores on eighth grade mathematics criterion-referenced tests at the .05 level after the effect of the pretest had been removed. A single criterion--12. Demonstrates sensitivity in relating to students--was highly significant at the .01 level. Table 7 lists the twenty-five performance criteria in order of significance as predictors of posttest scores after pretest effect had been removed.

A regression analyses of the eight performance criteria whose ratings were found to be significant predictors of student posttest scores on eighth grade mathematics criterion-referenced tests reveals that the pretest is the best predictor of the posttest score. The pretest was found to account for 90% of the variation in the posttest. When each of the eight significant performance criteria was also used as a predictor, it accounted for an additional percentage of the posttest variation. An equation could be formed using both the pretest score and the individual performance criteria ratings to predict the posttest score. Table 8 shows the additional percentage of posttest variation accounted for by each of the eight performance criteria ratings along with the best prediction equation for each criterion.

The Pearson product-moment procedure was used with the teacher performance criteria ratings and the adjusted posttest scores on the eighth grade mathematics criterion-referenced tests to determine the strength of the relationship between the two. The correlation coefficients are shown in Table 9. As expected, the strength of the

Table 7. Eighth grade mathematics criterion-referenced tests--teacher performance criteria ratings as predictors of student achievement

Teacher performance criteria	t-value
12. Demonstrates sensitivity in relating to students.	3.609**
13. Promotes self-discipline and responsibility.	2.729*
22. Demonstrates effective planning.	2.726*
23. Desires feedback from students.	2.693*
14. Involved with reaching district and building goals.	2.577*
1. Demonstrates ability to inspire and to motivate students.	2.314*
10. Demonstrates effective interpersonal relationships.	2.236*
11. Promotes positive self-concept.	2.193*
2. Communicates effectively with students.	2.068
8. Manages student behavior in a constructive manner.	2.023
20. Is pleasant, but not affectively extreme.	1.997
7. Demonstrates effective planning and organization.	1.862
5. Provides opportunities for individual differences.	1.827
19. Exhibits high expectations.	1.822
21. Displays a high energy level.	1.756
6. Effectively uses available materials and resources.	1.648
18. Models and gives concrete examples.	1.615
25. Moves quickly through the curriculum.	1.588
4. Prepares appropriate evaluation feedback.	1.546
16. Uses guided practice before independent practice.	1.479
3. Uses variety of evaluation methods with specific feedback.	1.369
24. Desires feedback from supervisors and principals.	1.321
9. Organizes students for effective instruction.	1.275
15. Demonstrates processes at beginning of learning (cueing).	1.190
17. Monitors seatwork closely.	1.129

*Significant at $p < .05$.

**Significant at $p < .01$.

Table 8. Eighth grade mathematics criterion-referenced tests--regression analysis showing additional posttest variation accounted for and prediction equation for each of the eight significant performance criteria ratings

	Percent of variation	Prediction equation
Criterion 12	4.5%	Posttest = 1.0498 x Pretest + 1.4113 x Criterion 12 - 8.9268
Criterion 13	3.2%	Posttest = .9689 x Pretest + 1.0400 x Criterion 13 - 3.2349
Criterion 22	3.2%	Posttest = 1.0026 x Pretest + 1.0313 x Criterion 22 - 4.8971
Criterion 23	3.1%	Posttest = 1.0043 x Pretest + 1.0331 x Criterion 23 - 4.7691
Criterion 14	3.0%	Posttest = .9698 x Pretest + 1.2921 x Criterion 14 - 4.7509
Criterion 1	2.5%	Posttest = .9991 x Pretest + .8923 x Criterion 1 - 4.0969
Criterion 10	2.4%	Posttest = .9821 x Pretest + .9946 x Criterion 10 - 3.1858
Criterion 11	2.3%	Posttest = .9985 x Pretest + .8685 x Criterion 11 - 3.7069

Table 9. Eighth grade mathematics criterion-referenced tests--Pearson product-moment correlations of twenty-five teacher performance criteria ratings with adjusted posttest scores

	Adjusted posttest score
12. Demonstrates sensitivity in relating to students.	.6648**
22. Demonstrates effective planning.	.5673*
23. Desires feedback from students.	.5614*
13. Promotes self-discipline and responsibility.	.5514*
14. Involved with reaching district and building goals.	.5310*
1. Demonstrates ability to inspire and to motivate students.	.5045*
10. Demonstrates effective interpersonal relationships.	.4809*
11. Promotes positive self-concept.	.4843*
2. Communicates effectively with students.	.4622*
8. Manages student behavior in a constructive manner.	.4409
20. Is pleasant, but not affectively extreme.	.4511
7. Demonstrates effective planning and organization.	.4250
5. Provides opportunities for individual differences.	.4183
19. Exhibits high expectations.	.4091
21. Displays a high energy level.	.4068
6. Effectively uses available materials and resources.	.3871
18. Models and gives concrete examples.	.3788
25. Moves quickly through the curriculum.	.3589
4. Prepares appropriate evaluation feedback.	.3575
16. Uses guided practice before independent practice.	.3334
3. Uses variety of evaluation methods with specific feedback.	.3048
24. Desires feedback from supervisors and principals.	.3132
9. Organizes students for effective instruction.	.3019
15. Demonstrates processes at beginning of learning (cueing).	.2686
17. Monitors seatwork closely.	.2489

*Significant at $p < .05$.

**Significant at $p < .01$.

relationships is in almost the same order and magnitude as found with the multiple regression procedure. Performance criterion 12 was very significantly related to the adjusted posttest scores, while eight other of the performance criteria showed a significant relationship to the adjusted posttest scores.

Fourth grade mathematics and reading and eighth grade mathematics: Norm-referenced tests

The same classes were also given standardized, norm-referenced pretests and posttests in fourth grade mathematics and reading and in eighth grade mathematics. The forward stepwise multiple regression and the Pearson product-moment procedures were used to test Hypothesis a with reference to these norm-referenced tests.

When the twenty-five individual performance criteria ratings were tested with the norm-referenced posttests in fourth and eighth grade mathematics, not one approached being a significant predictor of the posttest score after adjustments were made for the pretest effects. No relationship could be established between posttest results and individual teacher performance criteria ratings.

When the twenty-five individual performance criteria ratings were tested with the norm-referenced posttests in fourth grade reading, a lone criterion was found to be significant. Criterion 6. Effectively uses available materials and resources, was significant with a t-value of 2.232* ($p < .05$). The pretest accounted for approximately 85% of the variance in the posttest with Criterion 6 accounting for an additional 2%. The prediction equation was:

Posttest = .8597 x Pretest + .6175 x Criterion 6 + 3.6096.

The Pearson product moment correlation coefficient between the ratings on performance Criterion 6 and the adjusted posttest score was .3602* ($p < .05$).

Hypothesis b. There will be no relationship between the mean supervisor ratings of the teacher performance criteria and student achievement in fourth and eighth grade mathematics and fourth grade reading.

A mean of the twenty-five performance criteria ratings was calculated for each teacher. Hypothesis b tests whether this mean rating can be used as a predictor of outcomes on the six measures of student achievement used: fourth grade mathematics (both CRT and NRT), eighth grade mathematics (both CRT and NRT), and fourth grade reading (both CRT and NRT). As with the tests of the individual criteria ratings, the forward stepwise multiple regression and the Pearson product-moment correlation procedures were used. The results were similar for both tests.

The mean performance criteria rating was found to be a significant predictor of student achievement in both fourth and eighth grade mathematics on the criterion-referenced measures when using the multiple regression procedure. The mean rating was also significantly related to student outcomes in the same two groups when the Pearson product-moment correlation procedure was used. Both tests were performed on the posttest results after the removal of pretest effects.

No relationship was found between the mean performance criteria rating and student outcomes measured on norm-referenced tests in fourth

and eighth grade mathematics nor on either norm- or criterion-referenced tests in fourth grade reading. Table 10 summarizes the results of both multiple regression procedures and Pearson product-moment correlation procedures used to test the hypothesis.

Table 10. Relationships between mean performance criteria ratings and student achievement

	Mean performance criteria rating	
	t-value	Correlation
Fourth grade mathematics--CRT	2.488*	.3777*
Fourth grade reading--CRT	1.863	.3074
Eighth grade mathematics--CRT	2.340*	.5013*
Fourth grade mathematics--NRT	1.327	.2242
Fourth grade reading--NRT	1.125	.1849
Eighth grade mathematics--NRT	-1.023	-.2470

*Significant at $p < .05$.

With the fourth grade mathematics CRTs, the pretest accounted for approximately 51% of the variation in the posttest and the mean performance rating accounted for an additional 8%. The prediction equation was:

$$\text{Posttest} = .6608 \times \text{Pretest} + 1.4348 \times \text{Mean rating} + 9.6993.$$

With the eighth grade mathematics CRTs, the pretest accounted for approximately 90% of the variation in the posttest and the mean performance rating accounted for an additional 3%. The prediction

equation was:

$$\text{Posttest} = .9842 \times \text{Pretest} + 1.1693 \times \text{Mean rating} - 4.5920.$$

Hypothesis c. There will be no relationship between the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas and student achievement in fourth and eighth grade mathematics and fourth grade reading.

In testing Hypothesis c, clusters of performance criteria grouped into logical performance areas were used rather than the twenty-five individual performance criteria. Four such logical performance areas were identified. The areas and the clusters of performance criteria in each area are as follows:

Area 1. Productive Teaching Techniques

1. Demonstrates ability to inspire and to motivate students.
2. Communicates effectively with students.
3. Uses a variety of appropriate evaluation methods which provide students with specific feedback.
4. Prepares appropriate evaluation feedback.
5. Provides opportunities for individual differences.
15. Spends time at the beginning of the learning demonstrating processes to the students (cueing).
16. Uses more controlled (guided) practice before assigning homework (independent practice).
18. Plays a key role in modeling and giving concrete examples.
19. Has higher "expectations."
22. Plans better.
25. Moves more quickly through the curriculum.

Area 2. Organized, Structured Classroom Management

6. Effectively uses available materials and resources.
7. Demonstrates evidence of effective planning and organization.
8. Manages student behavior in a constructive manner.
9. Organizes students for effective instruction.
17. Monitors seatwork closely.

Area 3. Positive Interpersonal Relations

10. Demonstrates effective interpersonal relations.
11. Promotes positive self-concept.
12. Demonstrates sensitivity in relating to students.
13. Promotes self-discipline and responsibility.
20. Is pleasant, but not affectively extreme.

Area 4. Intellectual Stimulation and Desirable Out of Class Behavior

14. Is involved with the accomplishment of the district and building goals.
21. Has more energy.
23. Wants more feedback from students.
24. Wants more feedback from supervisors and principals.

In testing the logical performance areas for relationships to student outcomes on CRTs and NRTs in fourth grade mathematics and reading and eighth grade mathematics, once again the same procedures were used to perform the tests, namely multiple regression and Pearson product-moment correlation.

When the forward stepwise multiple regression test was run on the fourth grade mathematics criterion-referenced tests, Area 2. Organized, Structured Classroom Management was highly significant in predicting posttest scores after the removal of the pretest effects. The t-value was 2.754** ($p < .01$). Area 1. Productive Teaching Techniques, and Area 3. Positive Interpersonal Relations also contributed significantly to the prediction of the posttest scores. On the fourth grade reading criterion-referenced tests, Area 2. Organized, Structured Classroom Management was found to be the only significant predictor with a t-value of 2.397* ($p < .05$).

With the eighth grade mathematics criterion-referenced tests, Area 3. Positive Interpersonal Relations was the most significant contributor

to the prediction of the posttest scores. Area 4. Intellectual Stimulation and Desirable Out of Class Behavior was also found to be a predictor. None of the four performance areas were found to be significant predictors of student achievement on the norm-referenced tests in fourth grade mathematics and reading or eighth grade mathematics. Table 11 shows the results of these findings.

Table 11. Logical performance areas as predictors of student achievement on norm- and criterion-referenced tests in fourth grade reading and mathematics and eighth grade mathematics

Test	Area 1 t-value	Area 2 t-value	Area 3 t-value	Area 4 t-value
Fourth grade mathematics--CRT	2.565*	2.754**	2.043*	1.927
Fourth grade reading--CRT	1.716	2.397*	1.451	1.806
Eighth grade mathematics--CRT	2.044	1.842	2.866*	2.765*
Fourth grade mathematics--NRT	1.389	1.588	1.024	0.934
Fourth grade reading--NRT	1.155	1.690	0.748	0.602
Eighth grade mathematics--NRT	-1.087	-1.271	-0.920	-0.359

*Significant at $p < .05$.
**Significant at $p < .01$.

A regression analysis of the logical performance areas whose mean ratings were found to be significant predictors of student posttest scores on fourth grade mathematics and reading and eighth grade mathematics criterion-referenced tests reveals that the pretest is the best predictor of the posttest score. The pretest was found to account for 51% of the variation in the posttest in fourth grade mathematics, 49% in fourth grade reading, and 90% in eighth grade mathematics. When each

of the significant logical performance areas was also used as a predictor, it accounted for an additional percentage of the posttest variation. An equation could be formed using both the pretest score and the mean rating on the logical performance areas to predict the posttest score. Table 12 shows the additional percentage of posttest variation accounted for by each of the logical performance area mean ratings along with the best prediction equation for each performance area.

Table 12. Regression analysis showing additional posttest variation accounted for and prediction equation for each of the significant logical performance area mean ratings

	Percent of variation	Prediction equation
		Fourth grade mathematics--CRT
Area 2	9.7%	Posttest = .6697 x Pretest + 1.4257 x Area 2 + 9.1840
Area 1	8.7%	Posttest = .6734 x Pretest + 1.2632 x Area 1 + 9.6315
Area 3	7.9%	Posttest = .6821 x Pretest + 1.4132 x Area 3 + 7.3611
		Fourth grade reading--CRT
Area 2	8.0%	Posttest = .6904 x Pretest + 1.0614 x Area 2 + 10.8185
		Eighth grade mathematics--CRT
Area 3	3.4%	Posttest = .9978 x Pretest + 1.2957 x Area 3 - 5.6564
Area 4	2.8%	Posttest = .9857 x Pretest + 1.1726 x Area 4 - 3.4713

When the Pearson product-moment correlation procedure was used to test the relationships between the logical performance areas and student outcomes, all four areas were found to be related to adjusted posttest

scores on both fourth grade and eighth grade mathematics CRTs. Area 2. Organized, Structured Classroom Management and Area 4. Intellectual Stimulation and Desirable Out of Class Behavior were found to be related to adjusted posttest scores on fourth grade reading CRTs. No significant correlations were found between mean ratings on the logical performance areas and adjusted posttest scores on any of the norm-referenced tests which were used. Table 13 summarizes the results of these tests.

Table 13. Pearson product-moment correlations between logical performance areas and adjusted posttest scores on both norm- and criterion-referenced tests in fourth grade mathematics and eighth grade mathematics

Test	Area 1	Area 2	Area 3	Area 4
Fourth grade mathematics--CRT	.3424*	.3878*	.3332*	.2959*
Fourth grade reading--CRT	.2753	.3537*	.2062	.2969*
Eighth grade mathematics--CRT	.4077*	.3949*	.5100*	.4640*
Fourth grade mathematics--NRT	.2373	.2674	.1721	.1586
Fourth grade reading--NRT	.1912	.2717	.1234	.1018
Eighth grade mathematics--NRT	-.2610	-.3016	-.2241	-.0904

*Significant at $p < .05$.

Hypothesis d. There will be no significant difference in the supervisor ratings of the individual teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested.

This hypothesis was tested using the single classification analysis of variance procedure. If a difference was found, the Scheffé multiple range test was used to identify the achievement levels which produced the

difference. Teachers' classes were grouped according to whether their adjusted posttest means fell in the upper third, middle third, or lower third of the classes which were tested. The significant differences found will be reported singly since there were so few.

Fourth grade mathematics criterion-referenced tests

Criterion 9. Organizes students for effective instruction, was the only performance criterion which showed a difference in mean ratings among the upper, middle, and lower achievement groups. The hypothesis of no significant difference was rejected at the .05 level of significance [$F(2, 31) = 4.25$, $p < .05$]. Additional analysis using the Scheffé multiple range test revealed that the teachers whose adjusted posttest means were in the top third were rated significantly higher than those whose adjusted posttest means were in the bottom third. There were no significant differences in ratings between the top and middle achievement levels or between the middle and bottom levels. The mean ratings on Criterion 9 for each achievement level are shown in Table 14.

Table 14. Fourth grade mathematics criterion-referenced tests--means and standard deviation of ratings on Criterion 9 by achievement level

Achievement level	N	Mean	Standard deviation
Bottom third	10	3.95	1.32
Middle third	15	4.66	1.47
Top third	9	5.70	0.96

Fourth grade reading criterion-referenced tests

Criterion 7. Demonstrates effective planning and organization, and Criterion 21. Displays a high energy level, both showed a difference in their ratings among the three achievement levels. For Criterion 7, the hypothesis of no difference in ratings among the three levels was rejected at the .05 level [$F(2, 31) = 3.70$, $p < .05$]. For Criterion 21, the hypothesis of no difference in ratings among the three levels was also rejected at the .05 level [$F(2, 31) = 3.86$, $p < .05$]. In both cases, the Scheffé test revealed that the only difference in ratings was between the top third and bottom third. The mean ratings for Criterion 7 and Criterion 21 for each level of achievement are shown in Tables 15 and 16.

Table 15. Fourth grade reading criterion-referenced tests--means and standard deviation of ratings on Criterion 7 by achievement level

Achievement level	N	Mean	Standard deviation
Bottom third	9	4.08	1.16
Middle third	14	4.76	1.15
Top third	11	5.45	1.08

Eighth grade mathematics criterion-referenced tests

Criterion 12. Demonstrates sensitivity in relating to students, Criterion 13. Promotes self-discipline and responsibility, and Criterion 14. Involved with reaching district and building goals, all showed a

Table 16. Fourth grade reading criterion-referenced tests--means and standard deviation of ratings on Criterion 21 by achievement level

Achievement level	N	Mean	Standard deviation
Bottom third	9	3.86	1.26
Middle third	14	4.55	1.18
Top third	11	5.37	1.26

difference in their mean ratings among the three achievement levels. For Criterion 12, the hypothesis of no differences of the mean ratings among the three levels of achievement was rejected at the .05 level [$F(2, 16) = 4.17$, $p < .05$]. For Criterion 13, the hypothesis of no differences of the mean ratings among the three levels was rejected at the .05 level [$F(2, 16) = 4.76$, $p < .05$]. For Criterion 14, the hypothesis of no differences of the mean ratings was also rejected at the .05 level [$F(2, 16) = 4.09$, $p < .05$]. In all three cases, the Scheffé test revealed that the only difference in the mean ratings was between the top achievement group and the bottom achievement group. The mean ratings of the three criteria at each level of achievement are shown in Tables 17-19.

Fourth grade mathematics and reading and eighth grade mathematics: Norm-referenced tests

The single classification analysis of variance was used to test Hypothesis d with the levels of achievement being determined by norm-referenced adjusted posttest scores. The analysis produced no significant differences in the mean performance criteria ratings among

Table 17. Eighth grade mathematics criterion-referenced tests--means and standard deviation of ratings on Criterion 12 by achievement level

Achievement level	N	Mean	Standard deviation
Bottom third	6	4.15	1.10
Middle third	8	4.83	0.77
Top third	5	5.62	0.54

Table 18. Eighth grade mathematics criterion-referenced tests--means and standard deviation of ratings on Criterion 13 by achievement level

Achievement level	N	Mean	Standard deviation
Bottom third	6	4.27	1.26
Middle third	8	4.41	0.73
Top third	5	5.88	0.86

Table 19. Eighth grade mathematics criterion-referenced tests--means and standard deviation of ratings on Criterion 14 by achievement level

Achievement level	N	Mean	Standard deviation
Bottom third	6	4.63	0.86
Middle third	8	4.73	0.72
Top third	5	5.80	0.70

the three achievement levels for any of the individual performance criteria on any of the three norm-referenced tests administered.

Hypothesis e. There will be no difference in the mean supervisor ratings of the teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested.

Hypothesis e of no difference in the mean ratings of the performance criteria among the three achievement levels was tested using the single classification analysis of variance procedure. The analyses produced no significant differences in the mean performance criteria ratings among the three levels on any of the norm- or criterion-referenced tests in fourth grade mathematics and reading and eighth grade mathematics.

Hypothesis f. There will be no difference in the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested.

Hypothesis f of no difference in the mean ratings of the logical areas of clusters of performance criteria among the three achievement levels was tested using the single classification analysis of variance. No significant differences were found on either norm- or criterion-referenced measures in fourth grade mathematics, fourth grade reading, or eighth grade mathematics.

CHAPTER V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The primary purpose of this investigation was generally to determine whether supervisor ratings of teachers could be used to predict student achievement and particularly to determine which teacher performance criteria individually and in logical clusters are related to higher student achievement in fourth grade mathematics and reading and eighth grade mathematics. Secondary purposes concerned being able to improve future teacher evaluation instruments and to improve staff development interventions and teacher education programs by identifying teacher behaviors related to effectiveness.

Summary

Teacher evaluation data were collected by having the supervisors of the fifty-three project teachers complete the SIM Teacher Performance Evaluation instrument on each teacher after the completion of the 1983-84 project year. Student achievement data were in the form of class mean scores on pre- and posttests on both norm- and criterion-referenced achievement tests in fourth grade mathematics and reading and eighth grade mathematics.

The study identified criteria whose ratings were able to predict achievement in the subject areas tested over the two grade levels. The detailed findings which resulted from the hypothesis testing were presented in the preceding chapter. This summary will restate each of the research questions and present the answers to them according to the results of the tests of the null hypotheses.

Question a: Is there a relationship between the supervisor ratings of the individual teacher performance criteria and student achievement in fourth grade mathematics and reading and eighth grade mathematics?

On the fourth grade mathematics criterion-referenced tests, a relationship was found between supervisor ratings of thirteen of the twenty-five criteria and student posttest scores after the removal of pretest effects. On the fourth grade reading criterion-referenced tests, a relationship was found between supervisor ratings of seven of the twenty-five criteria and student achievement. On the eighth grade mathematics criterion-referenced tests, a relationship was found between supervisor ratings of nine of the twenty-five criteria and student posttest scores after the removal of pretest effects.

On the norm-referenced tests in fourth grade mathematics and reading and eighth grade mathematics, only one mean criterion rating was found to be related to student achievement. In fourth grade reading, the criterion which dealt with the teachers' effective use of available materials and resources was related to the mean adjusted posttest scores. No other relationships were identified.

Question b: Is there a relationship between the mean supervisor ratings of the teacher performance criteria and student achievement in fourth grade reading and mathematics and eighth grade mathematics?

The mean supervisor ratings of the twenty-five teacher performance criteria were found to be related to student learning on criterion-referenced tests in fourth and eighth grade mathematics. No such relationship was found on criterion-referenced tests in fourth grade

reading nor on norm-referenced tests in any of the three areas tested.

Question c: Is there a relationship between the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas and student achievement in fourth grade mathematics and reading and eighth grade mathematics?

On the fourth grade mathematics criterion-referenced tests, a relationship was found between the mean supervisor ratings of all four logical performance areas and student posttest scores after the removal of pretest effects. On the fourth grade reading criterion-referenced tests, two performance areas were found to be related to the mean adjusted posttest scores. On the eighth grade mathematics criterion-referenced tests, the statistical procedures employed found that all four logical performance areas were related to student achievement. No relationships were found between the mean supervisor ratings on any of the four logical performance areas and adjusted posttest scores on any of the norm-referenced tests administered.

Question d: Is there a difference in the supervisor ratings of the individual teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested?

The single classification analysis of variance test found a difference in the mean rating among the top, middle, and bottom achievement groups in fourth grade mathematics on only one of the twenty-five performance criteria. On the fourth grade reading criterion-referenced tests, a difference was found among the three achievement

levels on two performance criteria. In each case, the difference found was between the mean ratings of the top and bottom achievement levels. On the eighth grade mathematics criterion-referenced tests, the analysis of variance test revealed a difference among the three achievement levels on three performance criteria. Again, the revealed difference on all three criteria was between the top achievement group and the bottom achievement group.

The one-way analysis of variance test with levels of achievement determined by norm-referenced adjusted posttest means revealed no differences in the mean performance criteria ratings among the three levels of achievement for any of the individual performance criteria ratings on any of the three norm-referenced tests used.

Question e: Is there a difference in the mean supervisor ratings of the teacher performance criteria among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested?

Question f: Is there a difference in the mean supervisor ratings of the clusters of performance criteria grouped into logical performance areas among the teachers whose class achievement means were in the upper third, middle third, and lower third of those tested?

The mean ratings of the individual performance criteria and the mean ratings of the mean ratings of the clusters of performance criteria grouped into logical performance areas were tested among the three achievement levels using the one-way analysis of variance test. In each case, no differences were found among the levels on any of the tests

used, either norm- or criterion-referenced.

Conclusions

After analyzing the findings, several conclusions concerning various relationships between supervisor ratings of teacher performance and student achievement measured on both norm- and criterion-referenced tests in fourth grade mathematics and reading and eighth grade mathematics are offered.

1. On the fourth grade mathematics criterion-referenced tests, a relationship was found between supervisor ratings of thirteen of the twenty-five criteria and student posttest scores after the removal of pretest effects. The thirteen criteria identified were:

- Organizes students for effective instruction.
- Demonstrates effective planning.
- Monitors seatwork closely.
- Uses guided practice before independent practice.
- Communicates effectively with students.
- Demonstrates processes at beginning of learning (cueing).
- Manages student behavior in a constructive manner.
- Demonstrates ability to inspire and motivate students.
- Prepares appropriate evaluation feedback.
- Moves quickly through the curriculum.
- Desires feedback from supervisors and principals.
- Effectively uses available materials and resources.
- Demonstrates effective interpersonal relationships.

2. On the fourth grade reading criterion-referenced tests, a relationship was found between supervisor ratings of seven of the twenty-five criteria and student posttest scores after the removal of pretest effects. The seven criteria were:

- Displays a high energy level.
- Effectively uses available materials and resources.
- Demonstrates effective planning and organization.
- Uses variety of evaluation methods with specific feedback.

- Organizes students for effective instruction.
- Demonstrates effective interpersonal relationships.
- Communicates effectively with students.

3. On the eighth grade mathematics criterion-referenced tests, a relationship was found between supervisor ratings of nine of the twenty-five criteria and student posttest scores after the removal of pretest effects. The nine were:

- Demonstrates sensitivity in relating to students.
- Promotes self-discipline and responsibility.
- Demonstrates effective planning.
- Desires feedback from students.
- Involved with reaching district and building goals.
- Demonstrates ability to inspire and to motivate students.
- Demonstrates effective interpersonal relationships.
- Promotes positive self-concept.
- Communicates effectively with students.

4. On the norm-referenced tests in fourth grade mathematics and reading and eighth grade mathematics, only one mean criterion rating was found to be related to student achievement. In fourth grade reading, the criterion which dealt with the teacher's effective use of available materials and resources was found to be related to the mean adjusted posttest scores. No other relationships were identified.

5. The mean supervisor ratings of the twenty-five teacher performance criteria were found to be related to student learning on criterion-referenced tests in fourth and eighth grade mathematics. No such relationship was found on criterion-referenced tests in fourth grade reading nor on norm-referenced tests in any of the three areas tested.

6. On the fourth grade mathematics criterion-referenced tests, a relationship was found between mean supervisor ratings of all four logical performance areas and student posttest scores after the removal

of pretest effects. The areas were:

- Productive Teaching Techniques
- Organized, Structure Classroom Management
- Positive Interpersonal Relations
- Intellectual Stimulation and Desirable Out of Class Behavior

7. On the fourth grade reading criterion-referenced tests, the statistical procedures employed identified relationships between the second performance area, Organized, Structured Classroom Management, and the fourth performance area, Intellectual Stimulation and Desirable Out of Class Behavior, and student achievement.

8. On the eighth grade mathematics criterion-referenced tests, all four performance areas were identified as being related to student achievement.

9. Therefore, all four of the logical performance areas were found to be related to student achievement on criterion-referenced tests in fourth and eighth grade mathematics. Area 2 and Area 4 were related to achievement measured by criterion-referenced tests in fourth grade reading.

10. No relationships were found between the mean supervisor ratings on any of the four logical performance areas and adjusted posttest scores on any of the norm-referenced tests administered. On the eighth grade mathematics NRTs, all the correlations calculated were negative, although they were not significant.

11. The single classification analysis of variance test found a difference in the mean rating among the top, middle, and bottom achievement groups in fourth grade mathematics, measured on criterion-

referenced tests on only one performance criterion. This lone criterion related to organizing students for effective instruction. The difference was between the mean rating of the top achieving group (5.70) and the bottom achieving group (3.95).

12. On the fourth grade reading criterion-referenced tests, a difference was found among the three achievement levels on two performance criteria:

- Demonstrates effective planning and organization.
- Displays a high energy level.

On both criteria, the difference found was between the mean ratings of the top and bottom achievement levels. On the criterion, demonstrates effective planning and organization, the mean rating of the top group was 5.45, and the mean rating of the bottom group was 4.08. On the criterion displays a high energy level, the mean rating of the top group was 5.37 and the mean rating of the bottom group was 3.86.

13. On the eighth grade mathematics criterion-referenced tests, the analysis of variance test revealed a difference among the three achievement levels on three performance criteria:

- Demonstrates sensitivity in relating to students.
- Promotes self-discipline and responsibility.
- Involved with reaching district and building goals.

Again, the revealed difference on all three criteria was between the top achievement group and the bottom achievement group. On the criterion involving sensitivity, the mean ratings were 5.62 for the top group and 4.15 for the bottom group. Concerning promoting self-discipline and responsibility, the top group mean was 5.88, and the bottom group mean

was 4.27. On goal involvement, the top mean was 5.80, and the bottom mean was 4.63.

14. The one-way analysis of variance test with levels of achievement determined by norm-referenced adjusted posttest means revealed no differences in the mean performance criteria ratings among the three levels of achievement for any of the individual performance criteria ratings on any of the three norm-referenced tests administered.

15. The mean ratings of the individual performance criteria were tested among the three achievement levels using the single classification analysis of variance test. No differences were found in the mean ratings among the three levels on any of the norm- or criterion-referenced tests. Also, the mean ratings of the clusters of performance criteria grouped into logical performance areas were tested among the three levels using the same one-way analysis of variance procedure, and no differences were found among the levels on any of the tests used, either norm- or criterion-referenced.

Limitations

In examining the findings and conclusions herein presented, the reader should be aware of the following limitations imposed upon this investigation:

1. Data used in this study were collected from five school organizations with only fifty-three teachers and their classes representing all five organizations.

2. All teacher evaluation and student achievement data were

collected during the 1983-84 school year, preventing the analysis of findings in a longitudinal study beyond that time period.

3. Many variables not involved in the present study likely affected both supervisor ratings of teacher performance as well as student learning.

4. Findings of this investigation were based on supervisor ratings and, therefore, perceptions of teacher performance and not on actual teacher performance.

5. Only teachers and classes of fourth grade mathematics and reading and eighth grade mathematics were involved in this study, and caution should be observed in generalizing to other grade levels and subject areas.

6. Teacher evaluation data used in this investigation were not shared with the project teachers. Teacher performance ratings which were shared with the teachers were not available for this study and, therefore, a comparison between the two was not possible.

7. Age, race, sex, or years experience were not considered among either the teachers or supervisors.

8. Student achievement scores were not disaggregated according to race, sex, or socioeconomic status.

9. Project teachers were characterized by various levels of expertise in Effective Elements of Instruction.

10. Supervisors were characterized by various levels of expertise in observing and evaluating teachers.

Discussion

This investigation has attempted to add to the body of knowledge regarding performance criteria which are related to teacher effectiveness. Although the difficulties encountered in rating teacher performance and in assessing student achievement provide real obstacles in the study of teacher behaviors and student learning, this investigation does shed much insight into effective teaching behaviors. It is appropriate, however, to caution readers not to go beyond the research literature in efforts to generalize present findings.

Consider, for example, the types of tests used to measure student learning. Both external, standardized norm-referenced measures and internally developed criterion-referenced measures were used. Twenty-one of the twenty-five individual teacher performance criteria were found to be related to student achievement on one or more of the three criterion-referenced measures in fourth grade mathematics and reading and eighth grade mathematics. These relationships are shown in Table 20. On the norm-referenced tests, only one criterion was identified as being related to pupil learning, and that was on only one of the three norm-referenced tests. It is interesting that several studies have found that student opportunity to learn criterion material, content coverage, and curriculum-test congruence are important variables which are able to discriminate between more and less effective teachers (Berliner, 1980; Cooley & Leinhardt, 1978; Rosenshine & Furst, 1971). These variables would also seem to differentiate between norm-referenced and criterion-referenced test results. It is evident that there is no curriculum-test

Table 20. Predictors of student achievement

Teacher performance criteria	4th math	4th reading	8th math
I. Productive Teaching Techniques			
(1) Inspires and motivates students.	*		*
(2) Communicates effectively with students.	*	*	*
(3) Variety of evaluation methods w/feedback.		*	
(4) Prepares appropriate evaluation feedback.	*		
(5) Provides for individual differences.			
II. Effective Classroom Management			
(6) Uses available materials & resources.	*	**	
(7) Effective planning & organization.		**	
(8) Constructive management of behavior.	*		
(9) Effective organization for instruction.	**	*	
III. Positive Interpersonal Relations			
(10) Effective interpersonal relationships.	*	*	*
(11) Promotes positive self-concept.			*
(12) Sensitive in relating to students.			**
(13) Promotes self-discipline & responsibility.			*
IV. Professional Responsibilities			
(14) Promotes district & building goals.			*
V. High Gain Teacher Characteristics			
(15) Uses cueing.	*		
(16) Guided before independent practice.	*		
(17) Monitors seatwork closely.	*		
(18) Models and gives concrete examples.			
(19) Exhibits high expectations.			
(20) Pleasant, but not affectively extreme.			
(21) Displays high energy level.		**	
(22) Demonstrates effective planning.	**		*
(23) Desires feedback from students.			*
(24) Desires feedback from supervisors.	*		
(25) Moves quickly through curriculum.	*		
Mean rating on all 25 criteria.	*		*

* PROBABILITY < .05
** PROBABILITY < .01

congruence related to norm-referenced tests in that the NRTs do not test what is being taught. The norm-referenced tests also do not test the desired learning objectives. Therefore, the outcomes of the NRTs are not sensitive to differences in classroom instruction.

Of all the twenty-five teacher performance criteria, no significant relationship was found between any of the student achievement measures and four of the criteria. The four were:

Criterion 5. Provides opportunities for individual differences.

Criterion 18. Models and gives concrete examples.

Criterion 19. Exhibits high expectations.

Criterion 20. Is pleasant, but not affectively extreme.

Of the four, criteria 5, 19, and 20 are difficult to observe, document, and rate accurately. Criterion 18, which deals with modeling, also was not sensitive to differences in teacher effectiveness even though it can be documented more accurately than the other three. Most of the staff development interventions which address teacher effectiveness stress the importance of modeling, and, since it is not a particularly difficult behavior to add to one's teaching repertory, less effective teachers demonstrated it as well as those who were more effective. The mean rating of this criterion was the highest of all the 25 which were rated.

The SIM Teacher Performance Evaluation instrument used in this investigation is a relatively high inference tool to be used by appraisers. Many of the criteria do not focus on specific behaviors and involve more subjectivity on the part of the raters. Studies have shown that evaluation instruments requiring a higher degree of inference on the

part of the appraiser may well produce ratings with higher levels of error due to personal bias than those using low inference items (Brophy & Good, 1986). However, intensive staff development interventions were provided for all the teachers and administrators in Madeline Hunter's Essential Elements of Instruction by Dr. Hunter and some of her laboratory school instructors. This, along with other staff development training provided for project participants, should have served to reduce the error level often associated with high inference evaluation instruments.

Current research provides significant insight into teacher behavior variables which are related to student learning. Among these are:

- teacher clarity
- teacher enthusiasm
- use of student ideas
- probing
- teacher variability
- task-oriented behaviors
- use of structuring comments
- teaching strategies
- student opportunity to learn criterion material
- effective communication of expectations to students
- accept responsibility for student outcomes
- effective use of materials and resources
- thoughtful and reflective practitioners

These teacher behavior variables represent a consensus opinion of those behaviors showing the strongest evidence of relationship to student achievement (Berliner, 1984; Brophy & Good, 1986; Coker et al., 1980; Duttweiler, 1988; Porter & Brophy, 1988; Rosenshine & Furst, 1971).

The criteria used in this study have been found to be valid, reliable, and legally discriminating (Manatt & Stow, 1984). The findings of the present investigation parallel those of earlier studies and contribute to the body of knowledge which supports the definition of teacher effectiveness. The twenty-one criteria found by this study to be

related to one or more of the student outcome measures used are congruent with or can at least be coupled with one or more of the teacher effectiveness behaviors identified by the extant research.

It should be noted that the teacher effectiveness behaviors identified by this investigation are not mutually exclusive. Due to the homogeneity of job tasks, there is a natural correlation between performance criteria which describe job behavior (Landy, 1985). Therefore, the behaviors found by this study to be related to student achievement are likely related to each other and can also be paired with one or more of the teacher effectiveness criteria identified by previous research efforts.

It is interesting to note that of the twenty-one criteria found to be related to student achievement on at least one of the achievement measures, only two were related to criterion-referenced outcomes on all three tests. These were the criteria related to effective communication and effective interpersonal relations. Four other criteria were related to student achievement on two of the outcome measures, while the other fifteen significant relationships occurred with only one of the achievement tests. Also, in fourth and eighth grade mathematics, there were thirteen and nine teacher performance criteria, respectively, which were related to student achievement. In fourth grade reading, there were only seven significant relationships identified. When the logical performance areas were tested, all four mean ratings were related to student achievement on the criterion-referenced tests in fourth and eighth grade mathematics, while only two areas had mean ratings related

to outcomes measures on the criterion-referenced reading tests. It would seem that the performance criteria ratings are more sensitive to higher student outcomes in mathematics than in reading.

Some might suggest that there is a Gestalt effect, that supervisors rated the teachers based on a single impression rather than concentrating on the individual performance criteria, and that perhaps surgency affected the ratings. The findings that different individual criteria ratings, as well as different performance area ratings and different mean overall ratings, were related to different measures of student outcomes are sufficient to conclude that surgency and Gestalt effects were not present to a significant degree in this investigation.

This supports the conclusion many have reached that teaching is a highly contextual activity--being both grade level and subject matter specific. There were four criteria, however, whose mean ratings were related to higher student achievement in both reading and mathematics. The four dealt with effective communication, use of materials and resources, organization of students for instruction, and interpersonal relationships.

Recommendations for Practitioners

In addition to adding to the body of knowledge concerning the effectiveness of teachers, the results of this investigation have implications for those with the responsibility of implementing teacher performance evaluation systems and programs to improve teacher effects on student outcomes. The following recommendations are submitted:

1. Evaluation instruments consisting of the twenty-one criteria whose mean ratings were found to be related to higher student learning outcomes should be utilized, especially if teachers of different subjects are being rated. Efforts to develop evaluation instruments which discriminate among levels of teacher performance should be continued.

2. Inservice training of evaluators should be a continual process in order to improve the reliability of the appraisal process and appraisal ratings across raters. Training should be implemented to develop the teacher performance evaluation skills of all who supervise teaching.

3. Staff development interventions should be initiated to train teachers in the development of research-based teacher effectiveness behaviors. Elements determined to be essential to instruction need to be regularly reviewed and revised to reflect current research findings concerning effective practices.

4. Teacher education programs should be revised to train prospective teachers to develop skills and behaviors which relate to improved student outcomes.

Recommendations for Further Research

The following recommendations are submitted for the consideration of other researchers doing investigations in the area of teacher effectiveness:

1. The present study should be replicated in other school organizations with as many as three or four hundred teachers involved.

This would allow a factor analysis and, therefore, rational groupings of the performance criteria which were not possible in the present investigation.

2. Teacher evaluation and student achievement data should be collected over a three or four year period in a longitudinal study to remove many of the variations resulting from a short, one-year time frame.

3. Other grade levels in addition to fourth and eighth grades and other subject areas than mathematics and reading should be considered for identifying relationships to teacher performance.

4. In addition to confidential summative evaluation data, teacher performance ratings which are shared with the teachers should also be collected to allow an analysis of differences in ratings based upon the purpose for which the evaluation is used.

5. Teacher evaluation ratings should be categorized according to age, race, sex, and years experience of both teachers and supervisors to allow comparisons across these variables.

6. Student achievement scores should be disaggregated according to race, sex, and socioeconomic status to permit an analysis of differences based upon these important student variables.

"The effect of schooling on the individual pupil depends to a considerable extent on who his teacher is" (Medley, 1979, p. 11). As this investigation and others have shown, the effect of such schooling also depends on what the teacher does. With teaching being the complex, contextual, multi-faceted activity that it is, to justify significant

changes in our teacher evaluation systems, methods of instruction, and methods of assessing pupil learning, more empirical findings of the type reported here are necessary.

BIBLIOGRAPHY

Anderson, H. M. (1954). A study of certain criteria of teaching effectiveness. Journal of Experimental Education, 23, 41-71.

Barnes, S. (1981). Synthesis of selected research on teacher findings (Report No. 9009). Austin, TX: Texas University, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 223 569)

Barr, A. S., & Emans, L. M. (1930). What qualities are prerequisite to success in teaching? The Nation's Schools, 6(3), 60-64.

Barr, A. S., Torgerson, T. L., Johnson, C. E., Lyon, V. E., & Walvoord, A. C. (1935). The validity of certain instruments employed in the measurement of teaching ability. In H. M. Walker (Ed.), The measurement of teaching efficiency (Ch. IV). New York: Macmillan.

Barrett, R. S. (1966). Performance rating. Chicago: Science Research Associates.

Berk, R. A. (1984). The use of student achievement test scores as criteria for allocation of teacher merit pay. Paper presented at the National Conference on Merit Pay for Teachers, Sarasota, FL.

Berk, R. A. (in press). Limitations of using student achievement data for career ladder promotions and merit pay decisions. In J. V. Mitchell, Jr., B. S. Plake, & S. L. Wise (Eds.), Assessment of teaching. Hillsdale, NJ: Erlbaum.

Berliner, D. C. (1980). Using research on teaching for the improvement of classroom practice. Theory into Practice, 19, 302-308.

Berliner, D. C. (1984). The half-full glass: A review of research on teaching. In P. L. Hosford (Ed.), Using what we know about teaching (pp. 51-77). Alexandria, VA: Association for Supervision and Curriculum Development.

Blumberg, A. (1974). Supervisors and teachers: A private cold war. Berkeley: McCutchan.

Borg, W. R. (1979). Teacher coverage of academic content and pupil achievement. Journal of Educational Psychology, 71, 635-645.

Borg, W. R., & Gall, M. D. (1983). Educational research (4th Ed.). New York: Longman.

Borich, G. D. (1977). The appraisal of teaching: Concepts and process. Boston: Addison-Wesley.

Borich, G. D., & Kash, M. M. (1979). What the teacher effectiveness research has to say about teaching practices and student performance (R & D Report No. 5069). Austin, TX: Texas University, Research and Development Center for Teacher Education. Southwest Educational Development Laboratory. (ERIC Document Reproduction Service No. ED 189 077)

Boyce, A. C. (1915). Methods of measuring teachers' efficiency. Fourteenth Yearbook of the National Society for the Study of Education (Part II). Bloomington, IL: Public School Publishing.

Brookover, W. B. (1945). The relation of social factors to teaching ability. Journal of Experimental Education, 13, 191-205.

Brookover, W. B., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). School social systems and student achievement: Schools can make a difference. Brooklyn: J. F. Bergin.

Brophy, J. E. (1979). Teacher behavior and its effects. Journal of Educational Psychology, 71, 733-750.

Brophy, J. E. (1986). Classroom management techniques. Education and Urban Society, 18, 182-194.

Brophy, J. E. (1987). Teacher effects research and teacher quality. Journal of Classroom Interaction, 22(1), 14-23.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 328-375). New York: Macmillan.

Cangelosi, J. S. (1984). Evaluating teaching: A suggestion for principals. NASSP Bulletin, 68(475), 19-23.

Capie, W., & Cronin, L. (1986). How many teacher performance criteria should there be? Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 270 465)

Career Ladder Clearinghouse. (1988). Is "paying for performance" changing schools? Atlanta: Southern Regional Education Board.

Coker, H., Medley, D. M., & Soar, R. S. (1980). How valid are expert opinions about effective teaching? Phi Delta Kappan, 62, 131-134, 149.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality of educational opportunity. Washington, DC: U. S. Department of Health, Education, and Welfare, Office of Education.

Cooley, W. W., & Leinhardt, G. (1978). The instructional dimensions study: Final report. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Dangel, R. F., & Hopkins, B. L. (1979). The effects of different-length assignments on classroom deportment and academic performance. Journal of Educational Research, 72, 303-309.

Decotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. Academy of Management Review, 3, 635-646.

Doyle, W. (1985). Content representation in teachers' definitions of academic work (R & D Report No. 6161). Austin: University of Texas, Research and Development Center for Teacher Education.

Duttweiler, P. C. (1988). Organizing for excellence. Austin: Southwest Educational Development Laboratory.

Evertson, C. M. (1986). Do teachers make a difference? Issues for the eighties. Education and Urban Society, 18, 195-210.

Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. American Educational Research Journal, 17, 43-60.

Evertson, C. M., Emmer, E. T., & Brophy, J. E. (1980). Predictors of effective teaching in junior high mathematics classrooms. Journal for Research in Mathematics Education, 11, 167-178.

Feldvebel, A. M. (1980). Teacher evaluation: Ingredients of a credible model. The Clearing House, 53, 416-420.

Ferrera, R. J. (1974). Perceptions of school personnel and board members regarding teacher evaluation based on performance criteria. Unpublished doctoral dissertation, University of Southern California, Los Angeles, CA.

Fisher, C., Filby, N., Marliave, R., Cohen, L., Dishow, M., Moore, J., & Berliner, D. (1978). Teaching behaviors, academic learning time, and student achievement: Final Report of Phase III-B, Beginning Teacher Evaluation Study, Technical Report V-I. San Francisco: Far West Laboratory for Educational Research and Development.

Gage, N. L., & Giaconia, R. (1981). Teaching practices and student achievement: Causal connections. New York University Education Quarterly, 12(3), 2-9.

Gall, M. D. (1977). The importance of context variables in research on teaching skills. Journal of Teacher Education, 28, 43-48.

Glick, I. D. (1982). The search for excellence in teaching: An annotated bibliography. Washington, DC: ERIC Clearinghouse on Teacher Education. (ERIC Document Reproduction Service No. 213 660)

Good, T. L. (1979). Teacher effectiveness in the elementary school: What we know about it now. Journal of Teacher Education, 30, 52-64.

Good, T. L., & Beckerman, T. M. (1978). An examination of teachers' effects on high, middle, and low aptitude students' performance on a standardized achievement test. American Educational Research Journal, 15, 477-482.

Good, T. L., & Grouws, D. A. (1977). Teaching effects: A process-product study in fourth-grade mathematics classrooms. Journal of Teacher Education, 28, 49-54.

Good, T. L., & Mulryan, C. (in press). Teacher ratings: A call for teacher control and self-evaluation. In J. Millman & L. Darling-Hammond (Eds.), Handbook of teacher evaluation (2nd ed.). National Council on Measurement in Education. Beverly Hills, CA: Sage.

Gotham, R. E. (1945). Personality and teaching efficiency. Journal of Experimental Education, 14, 157-65.

Haertel, E. (1986). The valid use of student performance measures for teacher evaluation. Educational Evaluation and Policy Analysis, 8, 45-60.

Hellfritzsch, A. G. (1945). A factor analysis of teacher abilities. Journal of Experimental Education, 14, 166-199.

Henderson, R. (1984). Performance appraisal. Reston, VA: Reston Publishing Company, Inc.

Hill, C. W. (1921). The efficiency ratings of teachers. Elementary School Journal, 21, 438-443.

Hines, C. V., Cruickshank, D. R., & Kennedy, J. J. (1985). Teacher clarity and its relationship to student achievement and satisfaction. American Educational Research Journal, 22, 87-99.

Ilgen, D. (1983). Gender issues in performance appraisal: A discussion of O'Leary and Hansen. In F. Landy, S. Zedek, & J. Cleveland (Eds.), Performance measurement and theory (pp. 219-230). Hillsdale, NJ: Erlbaum.

Jayne, C. D. (1945). A study of the relationship between teaching procedures and educational outcomes. Journal of Experimental Education, 14, 101-134.

Jones, R. D. (1946). The prediction of teaching efficiency from objective measures. Journal of Experimental Education, 15, 85-99.

LaDuke, C. V. (1945). The measurement of teaching ability. Journal of Experimental Education, 14, 75-100.

Landy, F. (1985). Psychology of work behavior. Homewood, IL: Dorsey.

Lins, L. J. (1946). The prediction of teaching efficiency. Journal of Experimental Education, 15, 2-60.

Manatt, R. P., Palmer, K. L., Hidlebaugh, E. (1976). Evaluating teacher performance with improved rating scales. NASSP Bulletin, 60(401), 21-24.

Manatt, R. P., & Stow, S. B. (1984). Clinical manual for teacher performance evaluation. Ames, IA: Iowa State University Research Foundation.

Manatt, R. P., & Stow, S. B. (1986). Developing and testing a model for measuring and improving educational outcomes of K-12 schools: Technical report. Ames, IA: School Improvement Model (SIM), College of Education, Iowa State University.

Martin, W. R. (1979). Teacher behaviors--Do they make a difference? A review of the research. Kappa Delta Pi Record, 16, 48-50.

McGreal, T. L. (1983). Successful teacher evaluation. Alexandria, VA: Association for Supervision and Curriculum Development.

Medley, D. M. (1979). The effectiveness of teachers. In P. L. Peterson & H. J. Walberg (Eds.), Research on teaching (pp. 11-27). Berkeley, CA: McCutchan.

Medley, D. M. (1985). Issues and problems in the validation of teaching and teacher professional behaviors. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 261 085)

Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. Journal of Educational Research, 80, 242-247.

Medley, D. M., & Crook, P. R. (1980). Research in teacher competency and teaching tasks. Theory into Practice, 19, 294-301.

Medley, D. M., & Mitzel, H. E. (1959). Some behavioral correlates of teacher effectiveness. Journal of Educational Psychology, 50, 239-246.

Millman, J. (1981). Student achievement as a measure of teacher competence. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 146-166). National Council on Measurement in Education, Beverly Hills, CA: Sage.

Mintzes, J. J. (1982). Relationships between student perceptions of teacher behavior and learning outcomes in college biology. Journal of Research in Science Teaching, 19, 789-794.

Mohrman, A. M., & Lawler, E. E. (1983). Motivation and performance-appraisal behavior. In F. Landy, S. Zedek, & J. Cleveland (Eds.), Performance measurement and theory (pp. 173-189). Hillsdale, NJ: Erlbaum.

Murphey, K., Balzer, W., Kellam, K., & Armstrong, J. (1984). Effects of the purpose of rating on accuracy in observing teacher behavior and evaluating teaching performance. Journal of Educational Psychology, 76, 45-54.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for education and reform. Washington, D.C.: U.S. Government Printing Office.

Norusis, M. J. (1983). SPSS^{x} introductory statistics guide. Chicago: SPSS Inc.

Ornstein, A. C. (1982). A look at the research: How good are teachers in effecting student outcomes? NASSP Bulletin, 67(458), 61-70.

Ornstein, A. C. (1983). A look at the research--Part II: How good are teachers in effecting student outcomes? NASSP Bulletin, 67(459), 73-80.

Ornstein, A. C. (1986a). Research on teaching: Measurements and methods. Education and Urban Society, 18, 176-81.

Ornstein, A. C. (1986b). Teacher effectiveness: Current research and issues. Education and Urban Society, 18, 139-41.

Ornstein, A. C. (1986c). Teacher effectiveness research: Some ideas and issues. Education and Urban Society, 18, 168-175.

Otto, P. B., & Schuck, R. F. (1983). The effect of a teacher questioning strategy training program on teaching behavior, student achievement, and retention. Journal of Research in Science Teaching, 20, 521-528.

Ouzts, D. T. (1986). Teacher expectation: Implications for achievement. Reading Horizons, 26, 133-139.

Peterson, D. W. (1988). Rating teacher performance to determine career ladder advancement: An analysis of bias and reliability. Unpublished doctoral dissertation, Iowa State University, Ames, IA.

Popham, W. J. (1971). Teaching skill under scrutiny. Phi Delta Kappan, 52, 599-602.

Popham, W. J. (1975). Educational evaluation. Englewood Cliffs, NJ: Prentice-Hall.

Porter, A. C., & Brophy, J. E. (1988). Synthesis of research on good teaching: Insights from work of the Institute for Research on Teaching. Educational Leadership, 45(8), 74-85.

Rose, J. S., & Medway, F. J. (1981). Teacher locus of control, teacher behavior, and student behavior as determinants of student achievement. Journal of Educational Research, 74, 375-381.

Rosenshine, B. V. (1970). The stability of teacher effects upon student achievement. Review of Educational Research, 40, 647-662.

Rosenshine, B. V. (1976). Recent research on teaching behaviors and student achievement. Journal of Teacher Education, 27, 61-65.

Rosenshine, B. V. (1982). Teaching functions in instructional programs. Paper presented at the National Invitational Conference, "Research on Teaching: Implications for Practice," Warrenton, VA.

Rosenshine, B. V., & Furst, N. F. (1971). Research on teacher performance criteria. In B. O. Smith (Ed.), Research in Teacher Education (pp. 37-72). Englewood Cliffs, NJ: Prentice-Hall.

Ryans, D. G. (1949). The criteria of teaching effectiveness. Journal of Educational Research, 42, 690-699.

Schein, M. W. (1985). Student achievement as a measure of teaching effectiveness. Journal of College Science Teaching, 14, 471-474.

Seifert, E. H., & Beck, J. J. (1984). Relationships between task time and learning gains in secondary schools. Journal of Educational Research, 78(1), 5-10.

Shavelson, R. J., Webb, N. M., & Burstein, L. (1986). Measurement of teaching. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 50-91). New York: Macmillan.

Sherman, T. M., & Giles, M. B. (1983). An analytic review of a process-product variable: Teacher clarity. Educational Research Quarterly, 8(2), 26-37.

Smith, L. R. (1977). Aspects of teacher discourse and student achievement in mathematics. Journal for Research in Mathematics Education, 8, 195-204.

Smith, L. R. (1979). Task-oriented lessons and student achievement. Journal of Educational Research, 73, 16-19.

Soar, R. S., Medley, D. M., & Coker, H. (1983). Teacher evaluation: A critique of currently used methods. Phi Delta Kappan, 65, 239-246.

Squires, D. A., Huitt, W. G., & Segars, J. K. (1984). Effective schools and classrooms: A research-based perspective. Alexandria, VA: Association for Supervision and Curriculum Development.

Stallings, J. A. (1980). Allocated academic learning time revisited, or beyond time on task. Educational Researcher, 9(11), 11-16.

Stallings, J. A. (1981). Changing teacher behavior--A challenge for the 1980's. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA.

Stallings, J. A. (1985). Effective elementary classroom practices. In R. J. J. Kyle (Ed.), Reaching for excellence. Washington, DC: National Institute of Education.

Sweeney, J. E. (1988). Tips for improving school climate. Arlington, VA: American Association of School Administrators.

Taft, R. (1971). The ability to judge people. In W. W. Ronan & E. P. Prien (Eds.), Perspectives on the measurement of human performance (pp. 165-180). New York: Appleton-Century-Crofts.

Tobin, K. G. (1980). The effect of an extended teacher wait-time on science achievement. Journal of Research in Science Teaching, 17, 469-475.

Tobin, K. G. (1986). Validating teacher performance measures against student engagement and achievement in middle school science. Science Education, 70, 539-547.

Tobin, K. G., & Capie, W. (1982). Relationships between classroom process variables and middle school science achievement. Journal of Educational Psychology, 74, 441-454.

Tracz, S. M., & Gibson, S. (1986). Effects of efficacy on academic achievement. Paper presented at the Annual Meeting of the California Educational Research Association, Marina del Rey, CA.

Travers, R. M. W. (1981). Criteria of good teaching. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 14-22). National Council on Measurement in Education. Beverly Hills, CA: Sage.

Van Schaack, H. (1982). A qualitative study of excellence in teaching. Washington, DC: ERIC Clearinghouse on Teacher Education. (ERIC Document Reproduction Service No. 213 660)

Veldman, D. J., & Brophy, J. E. (1974). Measuring teacher effects on pupil achievement. Journal of Educational Psychology, 66, 319-324.

Walberg, H. J. (1988). Productive time and subject matter learning. In D. Tanner & J. W. Keefe (Eds.), Improving the curriculum: The principal's challenge. Reston, VA: National Association of Secondary School Principals.

Waxman, H. C., & Walberg, H. J. (1982). The relation of teaching and learning. Contemporary Education Review, 2, 103-120.

ACKNOWLEDGMENTS

Many debts of gratitude are owed to many people for their continued encouragement of my efforts during this doctoral program. My family consented to be uprooted from a relatively comfortable lifestyle to become the family of a graduate student for two years. My wife, Becky, has been unwavering in her support through the "down times" as well as the "up times." My sons, Danny and Bryce, made every adjustment necessary and rarely complained. This was a valuable learning experience for them, as well as for me. They were my reason for being here and are my reason for completing this undertaking.

Appreciation is also extended to the members of my doctoral committee, Drs. Shirley Stow, Detroy Green, Dick Manatt, Tony Netusil, and Jim Sweeney. My major professor, Dr. Dick Manatt, provided the expertise and experience that were invaluable. Whether he was in Blytheville, Ames, Chicago, Florida, or New England, he was always "there" when I needed him, and that was quite often--from the time I listened to his first "recruiting" talk until now. He issued the challenge. He set the high expectations and helped me to meet them. Dr. Jim Sweeney, for whom I have worked the past two years, has been an inspiration. If my mind could only go half as fast and in half as many directions and maintain half the intensity as his, it would be for me a great accomplishment. Dr. Tony Netusil taught me much more than statistics; he became a friend and provided constant support and encouragement throughout this endeavor. Dr. Shirley Stow is a devoted

teacher as well as student of teaching, and observing her teach and being taught by her have been a pleasure. In my contacts with Dr. Detroy Green, I have been much impressed by his dedication, his insights, and his abilities related to teaching and administration. The role models provided by these committee members are impressive, if not overwhelming. It has been my privilege to work with them.

Appreciation is expressed to my colleagues and members of the educational administration family. The collegial support group made up of these individuals has been a major strength of the program. The sense that "we are all in this together" has kept the ship afloat instead of causing it to sink.

I also wish to thank Carolyn Taylor for her willingness to work with some very short timelines in the typing of my drafts and final copy. And finally, if I had all this to do over again--I would.

APPENDIX A. SIM TEACHER PERFORMANCE EVALUATION

SIM Teacher Performance Evaluation

Evaluator's Name_____

District_____

Building_____

Teacher _____

Directions: Place check (✓) at appropriate location for each evaluative criterion of the above teacher. Scale ranges: 1 = low performance to 7 = high performance

Example: Maintains high expectations for students.

1 . . . . . 2 . . . . . 3 . . ✓ . . . . . 4 . . . . . 5 . . . . . 6 7
Needs Improvement | Meets Standard | Exceeds Standard

I. Productive Teaching Techniques

(1) **1. Demonstrates ability to inspire and to motivate students.**

1 . . . . . 2 . . . . . 3 . . . . . 4 . . . . . 5 . . . . . 6 7
Needs Improvement | Meets Standard | Exceeds Standard

(2) **2. Communicates effectively with students.**

1 . . . . . 2 . . . . . 3 . . . . . 4 . . . . . 5 . . . . . 6 7
Needs Improvement | Meets Standard | Exceeds Standard

(3) **3. Uses a variety of appropriate evaluation methods which provide students with specific feedback.**

1 . . . . . 2 . . . . . 3 . . . . . 4 . . . . . 5 . . . . . 6 7
Needs Improvement | Meets Standard | Exceeds Standard

(4) **4. Prepares appropriate evaluation feedback.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(5) **5. Provides opportunities for individual differences.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

II. Organized, Structured Classroom Management

(6) **1. Effectively uses available materials and resources.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(7) **2. Demonstrates evidence of effective planning and organization.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(8) **3. Manages student behavior in a constructive manner.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(9) **4. Organizes students for effective instruction.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

III. Positive Interpersonal Relations

(10) **1. Demonstrates effective interpersonal relationships.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(11) **2. Promotes positive self-concept.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(12) **3. Demonstrates sensitivity in relating to students.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(13) **4. Promotes self-discipline and responsibility.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

IV. Professional Responsibilities

(14) **1. Is involved with the accomplishment of the district and building goals.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

The following items have been used by Good, Brophy, Berliner, Medley, Hunter, et al. as descriptors of "high gain teachers." Please rate the teacher you are evaluating on these criteria too.

---

(15) **1. Spends time at the beginning of the learning demonstrating processes to the students (cueing).**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(16) **2. Uses more controlled (guided) practice before assigning homework (independent practice).**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(17) **3. Monitors seatwork closely.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(18) **4. Plays a key role in modeling and giving concrete examples.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(19) **5. Has higher "expectations."**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(20) **6. Is pleasant, but not affectively extreme.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(21) **7. Has more energy.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(22) **8. Plans better.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(23) **9. Wants more feedback from students.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(24) **10. Wants more feedback from supervisors and principals.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

(25) **11. Moves more quickly through the curriculum.**

1.....2.....3.....4.....5.....6.....7
Needs Improvement | Meets Standard | Exceeds Standard

APPENDIX B. TEACHER PERFORMANCE EVALUATION DATA

								TEACHER PERFORMANCE EVALUATION DATA																		PAGE 1	
TCHR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	MEAN
1	4.5	4.4	4.4	4.3	4.4	4.3	4.4	4.5	4.4	3.4	3.6	3.5	3.6	5.5	5.3	4.4	4.5	4.5	3.5	4.4	4.5	4.4	3.6	4.6	3.4	4.25
2	5.4	5.4	5.4	5.5	4.7	5.5	6.4	4.5	4.6	4.5	4.6	4.6	5.4	4.6	5.5	5.5	4.6	5.4	5.5	4.5	4.4	4.4	5.5	3.4	4.4	4.97
3	5.5	5.4	5.4	5.4	4.5	5.5	5.6	6.4	6.4	4.6	5.7	5.6	6.5	5.5	5.5	5.5	5.4	5.4	5.5	5.5	4.5	4.4	5.5	3.5	4.5	5.33
4	3.5	3.5	4.7	4.5	4.5	4.6	4.6	3.9	4.6	3.3	3.5	4.0	4.6	4.6	4.7	4.6	4.7	4.7	4.7	3.2	3.3	4.0	4.0	3.2	4.0	4.14
5	3.2	3.2	3.4	3.3	4.4	4.3	4.5	3.2	3.2	3.2	4.5	4.6	3.2	2.7	3.2	3.2	3.2	4.5	4.6	4.5	4.5	3.2	4.7	4.6	3.2	3.77
6	3.4	2.6	3.6	3.3	3.4	2.5	2.7	2.5	2.7	4.4	3.4	3.5	2.5	4.4	3.4	3.5	2.6	2.7	3.5	3.4	3.4	3.3	4.4	3.5	3.6	3.29
7	5.5	5.5	5.5	5.4	5.5	5.5	4.4	6.4	5.4	5.4	5.4	5.4	5.5	4.5	5.5	4.5	4.5	5.3	5.4	5.4	4.5	4.5	5.4	4.5	4.5	5.17
8	5.0	5.0	6.0	5.0	5.0	5.0	5.0	5.0	6.0	4.0	4.0	4.0	5.0	6.0	7.0	7.0	6.0	6.0	6.0	5.0	5.0	5.0	3.0	3.0	6.0	5.16
9	4.6	4.5	5.3	4.6	4.6	5.3	4.6	4.7	4.6	4.6	5.3	4.5	4.5	4.7	4.6	4.7	4.6	5.4	5.2	4.5	4.9	5.0	4.9	4.9	4.7	4.79
10	4.5	4.7	4.7	4.7	4.6	4.7	5.1	3.3	3.2	2.8	4.5	5.3	5.2	4.6	4.5	4.5	4.4	4.3	4.2	4.4	3.3	4.4	4.4	3.2	3.4	4.28
11	5.4	5.5	5.5	5.3	5.4	5.4	4.6	5.2	5.3	5.3	5.4	5.5	5.4	5.5	5.5	5.3	6.2	5.6	6.2	6.4	6.2	6.1	5.7	6.1	5.5	5.58
12	5.2	5.4	5.3	4.5	5.4	5.3	5.3	6.4	5.4	5.4	5.4	5.5	5.4	4.6	5.3	5.3	5.3	5.3	5.4	5.3	5.3	5.3	5.2	4.5	4.7	5.26
13	4.5	2.6	3.3	3.4	3.5	4.6	4.4	3.3	3.5	3.4	3.5	2.5	3.4	3.5	4.4	4.2	3.3	3.4	4.4	3.2	3.3	3.4	3.3	3.4	3.4	3.56
14	3.4	3.4	3.4	3.2	3.3	3.3	3.3	3.3	3.4	3.2	2.7	3.4	3.2	3.3	4.3	3.2	2.6	3.3	2.5	3.2	2.4	3.3	3.3	3.3	3.2	3.22
15	5.2	5.3	4.6	5.3	5.2	5.4	4.6	5.4	4.5	3.4	4.6	4.6	5.4	5.3	5.2	4.7	4.4	5.2	5.2	4.3	5.2	5.2	5.2	4.5	5.2	4.92
16	3.7	3.8	3.8	4.3	4.3	5.4	4.7	5.6	5.3	5.4	4.7	5.5	4.5	5.3	4.7	4.6	4.6	4.7	5.4	4.7	4.5	5.4	4.6	5.5	4.6	4.78
17	2.7	2.7	4.5	3.4	3.4	1.5	2.7	1.5	3.4	2.6	2.8	3.3	2.8	3.6	3.7	3.4	3.4	3.5	3.4	3.3	2.8	2.7	3.6	5.2	3.6	3.18
18	4.6	4.6	4.6	4.6	4.5	4.4	3.3	3.5	3.6	3.6	3.4	5.3	4.2	4.3	3.6	3.5	4.5	4.3	4.3	4.4	4.3	4.3	4.3	5.6	4.4	4.24
19	5.2	4.7	5.1	5.1	3.7	5.2	5.3	5.9	4.8	4.3	4.2	4.2	5.0	5.0	5.8	5.9	5.8	5.2	5.2	5.2	5.4	5.5	3.9	5.5	5.6	5.07
20	4.0	4.0	4.0	3.9	3.1	4.0	4.0	3.2	3.9	3.9	4.0	3.0	4.0	4.9	3.4	3.9	4.0	3.9	4.0	4.0	4.0	3.9	4.0	3.9	3.9	3.87
21	3.0	4.0	3.0	4.0	3.1	4.0	4.0	3.2	3.9	3.0	3.0	4.0	3.1	4.0	3.0	4.0	3.0	3.9	4.0	3.9	3.9	3.1	3.0	3.0	4.0	3.52
22	5.9	5.0	4.0	4.9	6.0	5.9	5.0	4.0	5.0	4.0	5.0	5.0	4.0	5.0	4.0	4.0	3.9	5.9	4.0	5.0	5.9	6.0	4.0	4.0	4.0	4.78
23	4.9	4.9	4.9	5.0	4.0	4.9	5.0	4.9	4.9	3.9	4.0	4.0	4.0	4.0	4.0	4.0	4.9	4.8	3.8	3.9	4.9	4.0	4.0	4.0	3.0	4.34
24	3.0	3.1	3.0	3.8	3.9	4.0	4.0	3.0	2.9	2.1	2.2	2.1	2.1	3.1	3.9	3.8	4.0	3.0	4.0	2.0	3.9	3.9	4.0	5.0	3.9	3.35
25	2.9	3.0	3.0	4.0	4.0	3.0	1.9	3.0	2.8	4.0	4.0	4.0	3.0	3.0	4.0	3.8	4.0	2.9	4.0	3.9	2.0	3.0	3.0	3.0	4.0	3.33
26	6.0	6.0	6.0	6.0	7.0	6.0	7.0	7.0	7.0	7.0	6.0	5.0	6.0	6.0	6.0	6.0	6.0	6.0	5.8	6.0	7.0	7.0	6.0	3.9	6.0	6.15
27	4.0	2.0	4.0	3.0	4.0	3.0	4.0	2.0	2.0	4.0	3.0	3.0	2.0	5.0	5.0	5.0	4.0	5.0	4.0	4.0	4.0	5.0	5.0	3.0	4.0	3.72

								TEACHER PERFORMANCE EVALUATION DATA																	PAGE 2	
TCHR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	MEAN
28	6.1	6.1	6.8	6.0	5.1	6.0	6.0	6.0	6.0	5.9	6.0	6.1	6.1	6.8	6.0	6.8	6.0	6.1	6.1	6.1	6.0	6.0	6.0	6.1	6.0	6.09
29	6.1	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.7	8.7	6.7	6.8	6.8	6.0	6.8	6.2	6.1	5.1	6.61
30	5.0	5.8	5.9	6.1	6.0	6.1	6.1	6.0	6.1	6.0	6.1	6.8	6.1	6.1	6.7	6.0	6.1	6.7	6.1	6.8	5.0	5.1	6.0	5.0	5.0	5.95
31	6.1	6.1	6.2	6.0	6.1	6.1	6.1	6.9	6.2	6.8	6.7	6.7	6.8	6.8	6.1	6.8	6.8	6.7	6.8	6.8	5.8	5.9	5.8	5.1	5.1	6.29
32	3.0	2.1	3.0	3.1	3.1	2.1	4.0	4.0	3.0	2.8	3.0	3.0	4.0	3.8	2.0	2.0	3.0	3.0	3.0	3.0	3.0	3.0	3.9	3.0	4.1	3.08
33	4.0	4.1	4.1	3.0	3.0	4.1	3.0	3.0	3.0	4.0	4.1	3.0	2.1	3.0	4.0	3.0	2.8	3.0	3.0	4.0	4.0	3.0	3.9	3.8	3.0	3.40
34	4.1	4.5	4.8	4.2	4.5	5.4	4.5	5.1	4.9	3.6	4.1	4.5	4.5	4.0	4.2	4.4	4.5	4.5	5.0	5.0	3.6	4.4	4.4	4.9	4.5	4.48
35	3.0	2.9	3.3	2.9	2.7	3.4	2.9	2.8	3.0	2.4	2.4	3.1	3.3	2.4	2.5	2.6	2.6	3.5	2.8	2.7	2.0	2.9	2.2	3.5	2.7	2.82
36	5.2	4.9	5.2	4.8	5.2	4.2	5.2	4.5	5.2	5.2	5.4	5.7	5.6	5.8	5.2	4.8	4.8	5.1	4.9	4.8	5.3	5.2	5.1	4.9	4.7	5.08
37	4.3	3.5	3.9	3.8	4.2	4.4	4.4	4.6	4.4	3.8	4.5	4.6	4.7	4.7	4.5	3.7	3.5	4.6	4.7	4.8	4.3	4.6	4.3	4.5	4.2	4.30
38	4.0	5.0	4.8	4.8	4.1	4.1	4.1	5.8	5.8	4.1	4.1	4.1	5.1	5.1	4.1	4.1	4.9	4.8	4.0	4.0	3.1	4.1	4.0	5.1	4.0	4.45
39	4.8	4.8	4.8	4.7	4.8	4.8	4.1	4.8	4.8	4.8	4.7	5.4	4.7	5.2	5.4	5.5	5.5	4.7	4.7	5.4	3.7	5.2	4.7	4.7	4.8	4.86
40	6.9	6.9	6.9	6.9	5.1	5.1	6.8	6.8	6.9	5.0	6.2	5.1	6.9	6.1	6.1	6.1	5.1	6.1	7.0	5.2	6.2	7.0	7.0	4.2	6.3	6.16
41	6.1	6.1	6.0	5.2	5.1	6.1	6.9	6.5	6.1	5.1	6.2	6.1	6.2	5.2	6.6	6.2	5.5	6.2	6.8	6.2	5.0	6.0	6.1	5.1	5.3	5.92
42	5.6	5.7	4.8	4.8	5.3	5.7	5.7	5.1	5.7	5.6	5.7	5.3	4.8	5.8	5.3	5.2	4.8	5.2	5.2	4.7	5.3	5.3	4.8	4.2	4.7	5.21
43	5.8	5.3	5.8	5.7	5.7	5.5	5.6	5.8	5.6	5.3	5.6	5.5	5.3	4.6	5.4	5.3	4.8	5.5	5.5	5.1	5.3	5.2	5.2	4.4	5.3	5.36
44	6.1	6.2	5.4	5.1	5.1	5.1	6.2	5.1	5.1	5.2	5.1	5.1	5.7	6.7	5.1	5.1	6.1	6.1	6.1	6.1	5.2	6.1	5.1	5.1	6.2	5.58
45	5.3	5.3	4.8	5.5	5.5	5.8	6.2	5.2	4.6	5.4	4.6	5.3	4.7	5.2	5.0	5.8	4.9	5.6	5.6	4.7	4.9	4.9	5.3	5.0	5.9	5.24
46	6.2	5.8	5.7	5.3	4.6	6.7	6.5	6.6	5.4	5.7	6.3	6.2	6.4	6.2	6.0	6.5	4.5	6.2	6.2	5.6	5.1	5.4	6.4	5.6	5.5	5.86
47	6.8	5.6	4.4	5.0	6.5	5.0	6.0	6.8	6.9	6.0	6.0	6.0	5.0	5.0	5.0	5.0	6.0	6.0	6.0	6.0	6.0	7.0	5.0	6.0	5.0	5.76
48	4.0	4.0	4.0	4.0	4.0	5.0	5.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	5.0	5.0	5.0	5.0	4.0	4.0	4.0	4.0	5.0	5.0	5.0	4.36
49	5.8	5.2	5.8	4.2	5.2	5.2	5.2	6.4	5.8	5.1	5.2	6.2	6.1	6.3	6.2	6.8	5.3	6.7	6.8	6.2	6.2	6.2	5.2	5.2	6.2	5.79
50	2.9	2.8	4.0	2.8	2.9	4.7	3.8	4.2	2.8	4.5	3.2	1.4	3.5	3.5	2.7	1.7	1.8	2.3	2.4	2.4	4.3	3.0	2.5	2.0	3.3	3.02
51	4.8	4.7	5.4	4.1	2.7	5.1	5.1	5.0	4.6	4.6	3.3	3.5	4.5	2.5	4.4	4.3	3.9	4.0	4.4	4.0	6.0	4.5	4.4	5.3	4.7	4.39
52	5.3	5.5	5.4	5.6	7.0	6.0	4.4	6.4	6.6	6.9	6.8	7.0	7.0	6.1	6.4	6.4	6.0	7.0	6.9	5.9	7.0	7.0	6.4	7.0	5.6	6.30
53	5.8	6.3	5.6	5.9	5.2	6.0	6.6	5.2	6.5	5.3	6.9	5.3	6.6	5.0	6.5	5.8	6.4	7.0	7.0	7.0	7.0	6.3	6.0	6.0	5.2	6.10

TEACHER PERFORMANCE CRITERIA RATINGS
DESCRIPTIVE STATISTICS

Criteria	Minimum	Maximum	Variance	Mean
1	2.7	6.9	1.3	4.8
2	2.0	6.9	1.5	4.6
3	3.0	6.9	1.1	4.8
4	2.8	6.9	1.0	4.6
5	2.7	7.0	1.2	4.6
6	1.5	6.8	1.2	4.8
7	1.9	7.0	1.4	4.9
8	1.5	7.0	2.0	4.8
9	2.0	7.0	1.7	4.8
10	2.1	7.0	1.4	4.5
11	2.2	6.9	1.5	4.6
12	1.4	7.0	1.6	4.7
13	2.0	7.0	1.8	4.7
14	2.4	6.8	1.3	4.8
15	2.0	7.0	1.4	4.9
16	1.7	7.0	1.6	4.8
17	1.8	6.8	1.4	4.6
18	2.3	7.0	1.5	5.0
19	2.4	7.0	1.5	4.9
20	2.0	7.0	1.4	4.7
21	2.0	7.0	1.5	4.7
22	2.7	7.0	1.5	4.8
23	2.2	7.0	1.1	4.7
24	2.0	7.0	1.1	4.5
25	2.7	6.3	0.9	4.6

APPENDIX C. STUDENT ACHIEVEMENT DATA

STUDENT ACHIEVEMENT DATA--FOURTH GRADE MATHEMATICS
PRETEST AND POSTTEST SCORES

TEACHER ID NO.	FOURTH GRADE MATH--CRT PRETEST	FOURTH GRADE MATH--CRT POSTTEST	FOURTH GRADE MATH--NRT PRETEST	FOURTH GRADE MATH--NRT POSTTEST
1	50.7	44.7	52.7	47.8
2	49.6	50.2	50.2	51.0
3	50.1	50.0	49.8	47.1
4	56.5	48.6	49.5	46.2
5	47.3	38.3	52.0	43.9
6	53.1	51.0	54.6	51.4
7	48.9	46.0	45.8	46.7
11	50.7	52.9	49.0	52.4
12	52.0	52.7	52.4	50.2
13	51.0	51.2	47.4	50.1
19	49.7	51.6	53.7	51.4
24	43.3	46.3	44.0	47.6
25	47.6	47.4	47.3	46.4
26	45.8	52.5	48.6	50.8
27	45.7	43.5	42.4	41.4
28	49.6	44.5	47.7	44.2
29	49.2	52.1	46.4	46.8
30	58.6	56.5	57.1	55.5
31	54.5	53.3	53.0	54.6
32	49.0	41.7	52.1	47.6
33	52.5	48.8	52.2	48.0
34	60.0	60.6	61.2	59.5
35	47.4	46.7	48.4	49.8
36	55.7	56.6	56.1	55.6
37	47.8	49.0	52.5	51.2
38	42.3	48.2	51.0	53.1
39	43.8	48.8	45.2	45.9
44	51.8	47.4	50.9	48.0
47	51.2	54.0	52.9	51.1
48	55.1	55.2	53.5	53.7
50	42.4	39.0	43.5	44.6
51	51.5	54.4	44.9	45.4
52	56.6	54.4	58.7	56.9
53	59.9	58.2	59.0	56.6

STUDENT ACHIEVEMENT DATA--FOURTH GRADE MATHEMATICS
ADJUSTED POSTTEST SCORES

TEACHER ID NO.	FOURTH GRADE MATH--CRT ADJUSTED POSTTEST	FOURTH GRADE MATH--NRT ADJUSTED POSTTEST
1	44.15	45.73
2	50.51	50.85
3	49.92	47.25
4	43.48	46.58
5	40.43	42.37
6	48.56	47.87
7	46.87	49.92
11	52.35	53.17
12	51.13	48.36
13	50.41	52.09
19	51.84	48.56
24	51.58	52.20
25	49.29	48.47
26	55.81	51.87
27	46.89	47.23
28	44.81	45.96
29	52.73	50.91
30	49.73	50.06
31	49.76	52.30
32	42.49	45.99
33	46.83	46.31
34	52.73	50.91
35	48.75	51.03
36	51.11	50.92
37	50.73	49.28
38	54.26	52.33
39	53.68	49.58
44	45.98	47.31
47	53.06	48.88
48	51.18	51.02
50	44.98	49.58
51	53.22	49.31
52	49.20	50.23
53	50.41	49.70

STUDENT ACHIEVEMENT DATA--FOURTH GRADE READING
PRETEST AND POSTTEST SCORES

TEACHER ID NO.	FOURTH GRADE READING--CRT		FOURTH GRADE READING--NRT	
	PRETEST	POSTTEST	PRETEST	POSTTEST
1	51.6	45.8	49.8	47.4
2	45.3	50.5	49.2	50.4
3	48.9	53.3	47.3	47.7
4	54.3	52.9	55.4	52.2
5	46.8	42.4	45.4	41.9
6	51.6	50.5	50.6	49.8
7	45.0	43.2	41.8	42.0
11	54.2	54.5	51.8	52.7
12	55.5	54.1	54.0	53.6
13	52.3	51.1	50.5	53.6
19	50.6	49.9	52.1	52.8
24	44.1	48.3	48.8	48.9
25	51.3	46.6	50.2	48.0
26	49.7	51.7	52.6	51.6
27	46.5	44.4	41.4	41.0
28	43.8	47.1	45.8	43.4
29	47.7	49.7	48.0	50.0
30	54.0	56.9	56.3	56.2
31	53.0	50.7	52.2	51.6
32	40.6	41.7	50.7	45.6
33	48.2	48.9	45.2	44.2
34	57.7	55.5	59.2	57.7
35	48.9	45.9	48.3	47.7
36	55.4	54.0	59.1	56.9
37	52.7	50.0	53.0	53.6
38	53.2	50.6	51.3	52.9
39	45.9	46.7	46.0	45.7
44	51.8	50.2	52.3	51.0
47	49.4	53.8	49.5	51.8
48	54.8	55.5	53.0	53.3
50	48.5	55.5	43.4	45.9
51	52.0	58.9	46.3	46.8
52	56.9	57.1	59.2	56.5
53	49.3	53.9	57.9	55.8

STUDENT ACHIEVEMENT DATA--FOURTH GRADE READING
ADJUSTED POSTTEST SCORES

TEACHER ID NO.	FOURTH GRADE READING--CRT ADJUSTED POSTTEST	FOURTH GRADE READING--NRT ADJUSTED POSTTEST
1	44.59	47.58
2	54.05	51.11
3	54.13	50.11
4	49.65	47.37
5	44.82	46.01
6	49.29	49.26
7	46.98	49.33
11	51.33	51.09
12	49.94	50.03
13	49.36	53.15
19	49.45	50.92
24	52.76	49.97
25	45.62	47.82
26	51.93	49.28
27	47.05	48.68
28	51.79	47.15
29	51.44	51.79
30	53.88	50.57
31	48.43	49.63
32	48.81	44.97
33	50.26	48.40
34	49.68	49.48
35	46.73	49.22
36	49.92	48.77
37	47.96	50.92
38	48.18	51.74
39	49.80	49.27
44	48.84	48.94
47	54.25	52.25
48	51.87	50.62
50	56.63	51.80
51	57.39	50.11
52	51.88	48.28
53	54.43	48.74

STUDENT ACHIEVEMENT DATA--EIGHTH GRADE MATHEMATICS
PRETEST AND POSTTEST SCORES

TEACHER ID NO.	EIGHTH GRADE MATH--CRT		EIGHTH GRADE MATH--NRT	
	PRETEST	POSTTEST	PRETEST	POSTTEST
8	56.8	54.5	56.2	54.8
9	52.6	51.2	54.2	53.4
10	48.7	51.2	51.7	51.4
14	47.5	46.9	48.5	50.2
15	52.4	51.2	52.5	51.4
16	48.3	49.9	47.0	48.9
17	42.7	40.5	45.9	47.4
18	41.5	42.9	44.1	44.5
20	52.0	50.7	50.7	50.1
21	36.7	34.4	36.8	35.2
22	34.3	34.5	31.8	31.8
23	52.2	51.0	51.3	51.1
40	48.2	51.8	47.5	49.0
41	38.9	37.0	33.0	30.7
42	48.9	51.3	61.5	60.2
43	53.7	53.7	51.1	50.9
45	49.3	50.6	44.9	45.6
46	45.7	47.8	54.6	54.3
49	49.2	53.0	52.8	52.5

STUDENT ACHIEVEMENT DATA--EIGHTH GRADE MATHEMATICS
ADJUSTED POSTTEST SCORES

TEACHER ID NO.	EIGHTH GRADE MATH--CRT ADJUSTED POSTTEST	EIGHTH GRADE MATH--NRT ADJUSTED POSTTEST
8	47.59	48.61
9	48.56	49.21
10	52.52	49.70
14	49.44	51.70
15	48.76	48.90
16	51.63	51.89
17	47.92	48.37
18	51.54	50.39
20	48.67	49.40
21	47.92	48.37
22	50.46	49.96
23	48.76	49.80
40	53.63	51.50
41	48.28	47.67
42	52.42	48.72
43	49.94	49.80
45	51.31	50.69
46	52.17	49.71
49	53.81	49.71