1988

# Anticipatory set bias: effects on teacher observation in evaluation

Diana Marie Bourisaw
*Iowa State University*

Order Number 8909129

# Anticipatory set bias: Effects on teacher observation in evaluation

Bourisaw, Diana Marie, Ph.D.

Iowa State University, 1988

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

# Anticipatory set bias:

# Effects on teacher observation in evaluation

by

Diana Marie Bourisaw

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

1988

TABLE OF CONTENTS

# LIST OF TABLES

# CHAPTER I.

## INTRODUCTION

Education reform is on the move. Since the publication of A Nation at Risk (1983) state and national governing bodies have taken legislative actions shaping many aspects of the reform movement including curriculum, length of the school year, graduation requirements, testing procedures, and teacher evaluation to name a few (Wahlberg, 1986). Of these aspects the one which has received more attention is teacher evaluation, which was the object of this study. What follows is a brief rationale of why it was the object of study and why rater bias was closely scrutinized.

Given the importance of quality teaching, many states directed great attention to the classroom, scrutinizing teaching behaviors and their impact on learning. Attention turned to effective methods of evaluating teacher performance. Ceremonial teacher evaluation--the "walk by" approach--where the evaluator walks by the teacher's classroom, looks in, and then later provides "feedback" to the teacher, is no longer the predominant mode of evaluation. Teacher evaluation is now approached much more systematically. Scheduled classroom observations, pre- and post conferences, formal feedback instruments and performance improvement commitments have been added to this process. Currently forty-six states mandate teacher

evaluation. Performanced based pay and other merit schemes make valid teacher evaluation very important (Duke and Stiggins, 1986).

Now teachers are being critically evaluated on the basis of their students' achievement, knowledge of subject matter, and other measures of effective instruction. Standard evaluation procedures and instruments, based on research on effective instruction, have been developed and implemented in many school districts across the country. Evaluators and teachers often receive training on this process. These evaluation systems and subsequent training are intended to create a fair and objective system for teacher evaluation with the goal being to improve teacher performance (Smith, Peterson, and Micceri, 1987).

The purpose of teacher evaluation is not solely to improve performance (Brandt, 1987). When a teacher is performing below expectation, is given time and assistance yet fails to improve, he can become the subject of a review for termination. Classroom observations reveal the degree of growth the teacher has made. The teacher's behavior is often recorded on an evaluation instrument resulting in a rating of his performance. The evaluator's ratings of that employee can be the final determination between retention or dismissal.

Teachers have also begun to be paid on the basis of performance (Cornett, 1985). Over forty states have developed some type of incentive scheme for their teachers (Olson, 1987). The bulk of these systems base differentiating compensation on the quality of

performance, with emphasis placed on ratings received on classroom performance.

Two elements seem to consistently play a key role in teacher evaluation--classroom observation and rating. In the teacher evaluation process, direct observation in the classroom and the resulting performance appraisal are tied to the need for valid and unbiased assessments made by principals, department chairpersons, and others acting in a supervisory capacity. Supervisors and administrators are expected to not only be knowledgeable and skilled observer/evaluators, they are also expected to be neutral and unbiased in their assessment of teacher performance. Obviously, evaluators who bring biases to the observation/evaluation can invalidate accurate ratings skewing the evaluation in an unfair and unproductive manner. The cost to the teacher can sometimes be reflected in dollars and cents and always influences the extent to which he may grow.

## Statement of the Problem

Appraisal of instruction should be a beneficial process, not just an activity teachers must complete. Performance appraisal requires time, impacts instruction, costs money, and influences relationships. When appraisals are done inaccurately, they not only may be costly and time consuming, but they may influence the relationship between the evaluator and the teacher and can negatively impact instruction. The process, in its best form, benefits students, teachers, administrators, the organization, and of course schools (Redfern,

1983). All reap the reward of improved instruction, fairness and motivation to improve performance.

But problems exist. As important as observation is, there are concerns about bias. Like students, no two evaluators are the same. Each individual approaches every situation with personal beliefs or attitudes. If the evaluator allows this to negatively affect the decisions made in the appraisal process, it may result in inappropriate supervisory and personnel decision. These biases can affect the performance appraisal of a teacher resulting in unfairness in the evaluation process, inaccurate targets for improvement, possible lack of career ladder compensation, and even termination.

Bias can occur during teacher observation affecting the observer's interpretation of what he or she sees. Bias may exhibit itself as attitudes and prejudices previously formed. Biases may also result from the amount and kind of training in observation the observer has already received. When these biases result in skewed observations, the entire evaluation process can become marred.

Causes of bias in observation have been researched. Factors influencing observation/evaluation include; age, race, education, and gender but frequently identified biases in educational research include the halo effect and prefatory remarks (Landy and Farr, 1983). The halo effect produces the most common rating error. Raters assign ratings based on global impressions as opposed to individual criterion. Raters may legitimize their ratings by consistently rating performance as "exceeds district standards," "meets district

standards," or "does not meet district standards." These ratings do not provide an accurate picture of teacher performance (Borman, 1975).

Remarks made prior to observing a teacher can create what is known as the expectancy effect, a similar rating bias (Reavis and Shine, 1977). This effect is obtained when teacher observation is prefaced by positive or negative remarks. These comments, or prefatory remarks, are directed to the rater regarding the teacher and can be made by anyone such as a fellow teacher, principal, or parent. They may be in the form of praise such as, "This is the most wonderful teacher I have ever seen." Or may be neutral or negative such as, "She's never caused problems here," or "He's not very good but go ahead and watch him anyway." Prefatory remarks can alter the rater's view or interpretation of the a teacher's performance during observation.

First impression bias is a similar bias that has been studied in business. Psychologists Latham, Wexley, and Pursell studied this phenomenon (1975). In their study, rating error occurred when the observer made judgments based on impressions formed after an initial meeting. The raters viewed a videotape whereby an applicant presented a negative initial impression by her action and answers. The remainder of the videotape showed the applicant to be acceptable for the job; however, the manager continued to act on the basis of his initial impression.

These researchers attempted to train observers to overcome this bias. Trainees participated in group discussion or workshop format training sessions. The workshop format proved most effective. These participants showed no errors in observer ratings six months after training.

Similar research in education is lacking. The topic of initial impression during observation based on teacher behavior has not been systematically studied. It seems likely that the initial impression made by teachers may affect evaluation ratings. If this is true the first few minutes of an observation biases observer ratings.

During the anticipatory set of a lesson the teacher establishes the need and focus of the lesson and instills in the children anticipation for learning. These initial moments of the lesson may give the observer his first impression of the teacher. If the same holds true for education and business, the rater may formulate his or her judgment of the lesson and perception of the ratee's performance within this initial time period. If this remains true during a lesson observation, the rater will slant his entire rating of the lesson based on the initial moments of the lesson.

This study was designed to determine if anticipatory set bias existed in teacher observation and to analyze the effects of this bias on the rating of teacher performance by supervisors. The problem for this study is more specifically addressed by the following questions:

1) To what extent does anticipatory set bias influence performance ratings of teacher evaluators?
2) Does the relative amount of teacher evaluation training reduce anticipatory set bias?
3) Does years of experience as a teacher evaluator reduce anticipatory set bias?

## Purpose

For teachers to receive an evaluation free from bias, evaluators must avoid anticipatory set bias. Valid teacher evaluation, fundamental fairness, and personnel and monetary decisions affecting teachers depend upon the neutrality and fairness of the evaluator. Yet we know little about this possible bias. Thus the intention of this study was to:

1) Assess the level of anticipatory set bias using a sample of evaluators receiving training in teacher evaluation.
2) Assess the effect of teacher evaluation training and teacher evaluation experience on anticipatory set bias.

## Objectives

To accomplish the purpose of this study, it was necessary to do the following:

1) Conduct a thorough review of the literature as it relates to teacher evaluation and bias.
2) Develop a teacher evaluation rating scale to accurately assess anticipatory set bias.

3) Design an effective and ineffective anticipatory set.
4) Develop videotaped lessons to use in conjunction with the rating scale.
5) Administer videotaped lessons and the teacher performance rating scale to a sample population.
6) Assess the extent of anticipatory set bias based on evaluator ratings.
7) Determine if teacher evaluation training is related to anticipatory set bias.
8) Determine if teacher evaluation experience is related to anticipatory set bias.

## Hypotheses

To fulfill the purpose of this study, the following hypotheses were developed and tested:

1) There is no significant difference in lesson ratings of evaluators who observe a teaching lesson with an effective anticipatory set and evaluators who observe a teaching lesson with an ineffective anticipatory set in their rating of overall teacher performance.
2) There is no significant difference between evaluators who observe a teaching lesson with an effective anticipatory set and evaluators who observe a teaching lesson with an ineffective anticipatory set of ratings of 8 out of 12 teaching strategies independent of anticipatory set.

3) There is no significant positive correlation between mean evaluator ratings of 8 out of 12 teaching strategies not related to anticipatory set and amount of teacher evaluation training of the evaluator.
4) There is no significant positive correlation between mean teacher evaluator ratings of 8 out of 12 teaching strategies not related to anticipatory set and amount of teacher evaluation experience of the evaluator.

## Basic Assumptions

This study is based upon the following assumptions:

1) Anticipatory set is a source of bias in teacher observation.
2) An evaluator's rating of a teacher's performance represents a valid measure of a teacher's performance at that point in time.

## Delimitations

This study was intended to determine the effects of anticipatory set bias on the overall rating of teacher performance and the impact of teacher evaluation training and teacher evaluation experience on these ratings. Teacher performance ratings were collected from teacher evaluation training sites in Erie, Pennsylvania; Fort Wayne, Indiana; and Independence, Kansas. Two graduate level administrative supervision classes were selected for this purpose in Ames, Iowa. A total of 106 evaluators participated in the study.

Since evaluators were not required to provide feedback to the teacher, ratings may have been lenient. Although this effect may have existed in this study, leniency did not affect the outcome of the study.

While it is acknowledged that many rater characteristics may have had an effect on the teacher performance ratings, only the characteristics of anticipatory set bias, teacher evaluation training, and teacher evaluation experience were selected for analysis.

## Definition of Terms

Anticipatory set - When the teacher establishes the need and focus of the lesson and instills in the children anticipation for learning.

Rater Bias - Systematic error in the rating of performance which is traced not to actual performance but rather to characteristics of the rater or of the situation in which the rating occurs.

First impression - An initial, perceived image produced by meeting someone for the first time.

Graphic response mode - An evaluation response combining numerical and descriptive responses.

Indicators - A statement of a research based teaching behavior used in making judgments about classroom performance.

Overall rating - A single indicator judging the teacher's total performance.

Prefatory remarks - Statements made regarding a teacher to an evaluator prior to the evaluation.

Reliability - The extent to which measurements (teacher evaluation ratings, in this study) are consistent across time and evaluators.

Scale - An instrument composed of indicators used to rate teacher performance.

Teacher performance - The measurement of research based effective teaching behaviors.

Validity - The degree to which an instrument is truthful in measuring what it purports to measure.

# CHAPTER II.

## REVIEW OF LITERATURE

The validity of teacher evaluation ratings is critical to the effectiveness of teacher performance appraisal systems. These ratings can be contaminated by rater bias. This review of literature assumes that a body of information exists which addresses the concept of bias. The sources for the search consisted of two major areas, one of those being studies of the evaluation of educational personnel. The other major source came from studies of bias both in industry and education.

This review of literature within these categories attempted to: (1) provide a brief background and current information on the process of teacher evaluation, (2) describe bias as it relates to performance appraisal, (3) describe bias as it relates to classroom observation and (4) identify anticipatory set bias as it relates to business and industry and the effect on the validity and reliability of performance evaluation.

### Background

Less than a decade ago, many school systems were spending a major amount of time and money evaluating student achievement and a minimal amount checking on the teaching process (Buttram and Wilson, 1987). Student test scores were monitored and reading comprehension

was checked. Papers were shuffled and grades recorded. Students were labeled "unmotivated" if they received a failing grade and "incorrigible" if they were suspended for disciplinary reasons. But today teachers are being held accountable for their students' success. Instead of blaming students for their success or failure, teachers are held responsible for student performance. In addition, changes in the teacher evaluation process have come on the heels of changes in the way accountability for learning is viewed.

The appraisal of teaching, in some form or other, has been in existence for many years (Fletcher and Williams, 1985). Administrators typically judged teacher performance based on the number of discipline referrals to the office, few or no referrals equating to a positive appraisal. Teacher evaluation forms were left in teachers' mailboxes awaiting the teachers' signatures. Only on rare occasions did administrators and teachers discuss the content of the documents. If this discussion occurred, teachers looked forward to the end of this annual ritual so they could get back to the classroom and teach behind closed doors. New views on teacher evaluation have opened these doors. The teacher is no longer the only adult in the classroom.

Studies pointing to the schoolhouse and away from the school children have increasingly placed attention on teacher performance (National Commission, 1983). Jointly teachers and administrators are developing appraisal systems that focus on instructional improvement (Holdzkom, 1987). This is not an easy task. The goal of developing

an evaluation instrument based on the latest in effective schools/effective teaching research is an awesome feat in and of itself. This goal is often accomplished in a setting where administrators and teachers work together. In such a setting, the state of North Carolina narrowed the effective teaching research down to the following eight basic teaching functions:

- Management of instructional time
- Management of student behavior
- Instructional presentation
- Instructional monitoring of student performance
- Instructional feedback
- Facilitating instruction
- Communicating within the educational environment
- Performing non-instructional duties

Based on these eight items, a system for active data collection, feedback, and general evaluation was put into place in over forty school districts across the state.

Like most school systems that develop such a model, the goal throughout the development of North Carolina's performance appraisal process was the improvement of performance, not the elimination of personnel. Information regarding effective teaching practices was provided to districts throughout the state, teachers and administrators received training in Teacher Performance Appraisal, and objective evaluation instruments with graphic response modes were used. During training sessions teachers and administrators observed

videotaped lessons, recorded data and evaluated teaching lessons. A total of twenty-four hours was spent in such training. North Carolina has developed several other training modules to further this training process.

North Carolina is not alone in its efforts to train teacher evaluators. Training in observation techniques is common to all objective observation systems. Districts in Florida using the Florida Performance Measurement System receive observer training (Smith, Peterson, and Micceri, 1987). The question remains: Does this training produce evaluators who will make valid, reliable, unbiased judgments? Does observer training actually make a difference? Does the trained rater still hold biases despite training?

## Bias in Performance Appraisal

Since 1950, the study of bias in relationship to performance appraisal has been a major topic of discussion in the literature of business and industry (Rice, 1985). Over 300 studies have appeared in educational and business journals focusing on bias, particularly in the areas of race and sex. These concerns emerged from the adoption of the Equal Employment Opportunity guidelines issued during 1969 and 1970.

The research on sexual bias shows no consistent pattern (Wexley and Pulakos, 1983). Kenneth N. Wexley, a psychologist at Michigan State, studied 300 manager-subordinate pairs and discovered nothing to substantiate sexual bias in performance ratings. Psychologist

William H. Mobley confirmed this finding in studies of over 1,000 employees finding that women typically ranked higher than men regardless of the sex of the supervisor (Mobley, 1982).

On the contrary, racial bias has been shown to exist. Supervisors tend to give higher ratings to those of like race (Mobley, 1982). White supervisors tend to rate white employees higher than black and black supervisors tend to rate black employees higher than white.

As far as many ratees are concerned, every characteristic that the rater brings to the task is a variation on a personal bias. The color of one's skin, gender, age and education all play a role in evaluation. These characteristics affect one's interpersonal relationships with all individuals, not just those in the work setting (Rice, 1985).

But bias is not restricted to race, sex, age and education. Bias cuts across many boundaries including religion, national origin, and physical attractiveness (Landy and Farr, 1983). Bias in these areas is also known as prejudice and bigotry. These biases are often obvious and intense. Unfortunately the broader picture of bias is not that simple.

Types of raters can also have an effect on the outcome of an evaluation (Borman, 1979). Supervisors, peers, and subordinates tend to rate the same employee differently. While supervisors tend to be harder on subordinates, peers are more consistent in their ratings (Rice, 1985). The length of experience on the job also can effect a

rating as well as the job performance level of the rater. Those evaluators with more experience and a job performance close to that of the ratee tend to be more accurate in their rating. These additional biases serve to compound the issue of bias in performance evaluation.

Skill and ability are not the only factors used to evaluate employees. Psychologists Rutlege Jay and James Copes reviewed 47 studies of employees and job longevity. The studies showed employees with more experience consistently received higher ratings than those with less experience and equal ability (as cited by Rice, 1985). This literature review implied that company loyalty carried a high price tag.

The study of bias had a major impact on performance appraisal in business and industry. In the past several decades, education began to see the need for similar research. Many studies of bias in business were duplicated in education and new variables unique to teaching were also addressed. At the University of Pittsburgh, Dr. Mary Jo Retzer analyzed four types of bias in teacher evaluation (Retzer, 1980). These areas were personality, classroom preparation, technique of instruction, and pupil reaction. One hundred subjects were randomly divided into four treatment groups. Before watching a videotape of an elementary reading lesson, each group received one of the following four treatments: positively biased comments about the teacher's performance, negatively biased comments about the teacher's performance, neutrally biased comments about the teacher's

performance, or no information about the teacher's performance. No significant difference was noted in the ratings of the four criteria.

Rucker (1981) completed a study in the area of teaching style bias. In his study it was hypothesized that principals with a preference for a certain teaching style would rate those teachers with a similar style higher than those with dissimilar teaching styles. Four basic teaching styles were identified: personal, social interaction, information processing, and behavior modification. After analyzing the data, Rucker concluded that teaching style preference is not a source of bias in teacher evaluation.

Geosits (1978) addressed the possible rater bias of an open approach to instruction verses the traditional approach. Her study concluded that there is no evidence to substantiate bias in the ratings against open style teaching. Yet in a study of the perception of principals on teacher behaviors, Tuckman and Others (1977) confirmed that principals at different levels differ significantly in terms of their perception of teacher effectiveness. Elementary principals appear to prefer teachers who are very warm and accepting, highly organized and creative. Intermediate principals prefer very organized, in control, warm, sociable, fair, imaginative, creative, and dynamic teachers. And senior high principals appear to prefer highly systematic, organized, structured, and task oriented individuals. This teacher preference may affect ratings of teachers by principals across grade levels.

Millman (1981) looked at the correlation between an evaluator's emphasis on paperwork and report deadlines and teacher ratings. An evaluator with this characteristic will often rate a teacher negatively if he is adequate in teaching performance but late with paperwork. Millman stated, "To evaluate teachers and to conduct effective appraisal interviews, it is vitally important that evaluators understand how their values affect their judgements of teaching competence" (Millman, 1981, p. 53).

Christner (1981) tested for possible biases in school administrators' evaluation of staff members in the Austin, Texas, Independent School District. Variables tested included the evaluator's contractual status, highest degree held, as well as the evaluator's gender and ethnic status. Bias was noted during all three years of this study. Males, blacks, secondary level teachers and other professionals, inexperienced teachers, and those with bachelors degrees and less permanent contracts consistently received lower ratings. These results indicated the need to consider possible rater biases in the development, implementation, and use of teacher evaluation systems.

Bias does exist in business as well as education (Weitzul, 1983). Evaluators need to be aware of these biases in order to effectively address their impact on evaluation.

## Bias in Classroom Observation

> "The most prevalent technique for collecting information about classroom processes is classroom observation."
> (Levin and Long, 1981, p.39)

Classroom observation is the most common method for collecting the necessary data to make accurate evaluative ratings. The anthology Mirrors for Behaviors contains over 92 observational techniques (Simon and Boyer, 1970).

The late Robert Goldhammer believed the purpose of observation was to collect objective data in order to reconstruct the lesson and analyze its contents (Goldhammer, 1969). Observations of teacher performance have been done for decades and possibly centuries (Wiersma and Gibney, 1985). The method for recording data has run the spectrum of tabulation on complex matrices to anecdotal data subjectively perceived by the observer. These methods of observation have been developed to assist the observer in eliminating distortions of perception, to see and hear objectively (Goldhammer, 1969).

The University of Toledo in the 1970s and early 1980s published low inference and high inference observation inventories (Wiersma and Gibney, 1985). The low inference observation inventory involves a behavioristic approach where observers record the occurrence of specific behaviors. The instrument, published by the University of Toledo, is called the Classroom Observation Keyed for Effectiveness Research or COKER. The high inference observation inventory forces the observer to make a judgment or rating regarding teaching behaviors. This inventory is known as the Teacher Performance Assessment Instrument or TPAI.

Numerous reliability studies have been conducted using these inventories. They have proven to be a viable approach to evaluation.

Since many states require classroom observation to be a primary component of teacher evaluation, it is imperative that these observations be as objective as possible (Valentine, 1984). The COKER and TPAI have attempted to eliminate bias from the observation process.

Memory and perception are important components of observation (Goldhammer, 1969). While we tend to acknowledge the distortions of our memory, we do not consciously acknowledge that perceptual distortions operate from moment to moment in all of us. These perceptual distortions are known as biases.

Soar and colleagues (1983) believe that the chief existing teacher evaluation methods--measuring teacher characteristics, student achievement, test scores and teacher performance rating scales--are subjective and open to bias. These researchers advocate evaluation procedures that are performance based and empirically tested. However, when instruments that are performance based and empirically tested are used, bias still exists.

Although rating instruments have been developed and implemented for the purpose of reducing or eliminating bias in teacher observation, psychologists have grown tired of efforts to improve rating scales or experiment with sources of bias (Landy and Farr, 1983). Even when those innovations do reduce error, it is often so minimal that the improvement is considered insignificant. If this is the case, how do evaluators reduce their own personal biases in order to be fair and objective in rating employees?

Goldhammer maintained the following:

> "If one is ready to believe that self knowledge, particularly knowledge of one's own values and biases already constitutes some measure of control over such biases, then a commitment to knowing oneself in these terms would be appropriate to demand of clinical supervisors."
>
> (Goldhammer, 1969, p.294)

This statement would dispute Landy and Farr's moratorium on the study of biases and encourages self-knowledge in this area. Numerous purported biases have been studied and confirmed or denied. Unfortunately not all have been addressed in the literature.

## Anticipatory Set Bias

Numerous studies on bias in business have been paralleled in education, studies on race, gender, and age to name a few. But one remains to be confirmed, that is, the initial impression the teacher makes on the evaluator at the beginning of the lesson. In business this is called making a first impression--in education it is labeled the "anticipatory set."

First impression bias has been researched in business and industry (Latham, Wexley, and Pursell 1975). This bias occurs when an observer evaluates someone on the basis of judgments made after an initial meeting. In this study, manager trainees were provided a job description and requirements for an insurance rater. Each then viewed a videotaped interview in which the applicant presented a poor impression through her appearance, action and answers. The remainder of the videotape showed the applicant to be acceptable for the job.

However, the manager trainees continued to act on the basis of their first impressions.

Another part of the study tested first impression bias by presenting a videotaped interview of a woman who presented an unfavorable impression at first but later in the interview presented a favorable impression. A second tape showed the woman presenting a favorable-unfavorable response. The responses by the woman were identical but the sequence of the questions and answers in the interview were reversed. Again the manager trainees rated the applicant based on their first impressions. The managers in these groups frequently evaluated employees and were present during this study for the purpose of receiving training in employee evaluation. The managers were divided into three groups; a control group, a discussion group, and those that participated in a workshop. Six months after training trainees in the control group and those in the discussion group committed numerous rating errors while those in the workshop group committed no errors.

Weitzul (1983) also studied the effects of first impressions in the insurance industry, this time between insurance agents and their clients. Because insurance agents spend so little time with their clients, it is important for them to obtain an accurate first impression. Inaccurate initial perceptions of the client can be costly to both the agent and the company. In this study, when insurance agents did reach false conclusions about their clients, they rationalized these conclusions based on their first impression.

The closest educational correlate to this study of first impression bias have been studies conducted on the use of prefatory remarks. Prior to Reavis' and Shine's 1977 study of prefatory remarks and their effect on teacher evaluation, no studies in this area were found. Thirty-four graduate students who had just completed nine hours of instruction on rater bias participated in this study. The graduate students were randomly divided into two groups to view the same videotape of a teaching lesson. Prior to viewing the tape one group received a positive comment about the teacher's teaching capabilities and the other group received a mildly negative comment regarding her competency. After viewing the tape each participant completed a teacher performance rating scale.

The results of the study showed that the group receiving the positive comment rated the teacher significantly higher on the rating scale then the group receiving the mildly negative comment. Even after nine hours of training on biases, these evaluators continued to be influenced by prefatory comments. It was concluded that teacher ratings can be altered by these remarks.

First impression in business and prefatory remarks in education can be equated to an evaluator's first few minutes in a teacher's classroom. This important initial perception by the observer is one of the most critical times in the lesson (Valentine, 1985). It is during the anticipatory set portion of the lesson that the evaluator begins to form impressions of the effectiveness of the teacher. This study attempts to prove that first impression bias exists in

education in the form of anticipatory set bias. Unawareness of this bias by evaluators can only be a detriment to teachers in the evaluation process.

## Summary

Teacher evaluation is one of the more emotional issues around today. Teachers and administrators are both expected to be knowledgeable in the areas of effective schools and effective teaching yet very little time or resources are allocated for this purpose. School districts often receive performance appraisal systems in the mail on one day and are expected to implement them the next. It is not unusual for administrators to evaluate staff without any training at all.

When money and tenure are on the line, it only makes good sense to investigate all the possible problems before implementing a performance appraisal system that could have such a tremendous impact. When inaccurate judgments are made about classroom performance, the learner is the one who ultimately suffers. The purpose of performance appraisal is professional growth. Inaccurate professional growth plans are not only a time waster for the teacher, but can be detrimental to the learner as well. And if the evaluator, the identified instructional leader in the building, is prescribing the cure for a misdiagnosed problem, the career ladder teacher may spend a great deal of time working toward improving an area that does not enhance her opportunities for that financial incentive.

Biases have always existed in performance appraisal in business, industry, and education. Human perceptions and error are inevitable. Although many biases have been identified, many remain unknown. The awareness of such biases can only lead to improved evaluation.

# CHAPTER III.

# METHODS AND PROCEDURES

## Introduction

This chapter describes the procedures followed to conduct the study. The field test and development of instruments and media are described in the first section of this chapter. The sample is described in the second. The third section of this chapter describes the method employed for gathering data for this study. The final section presents the data analysis procedures.

## Field Test and Development of Instruments and Media

In this section the field test will be described: this includes development of the teacher rating scale instrument, development of the videotaped lessons, and the field test and results.

### Initial Instrument

The Teacher Performance Rating Scale (Appendix A) used in this study was developed by this researcher to determine if anticipatory set bias exists in classroom observation. It was used by teacher evaluators to rate a lesson. The items on the scale are statements describing effective teaching behaviors based on the Hunter model (Hunter, 1984). The following four items directly relate to anticipatory set:

(1) Relates current lesson to previous learning.

(2) Provides focus for learning.

(3) Involves students in learning new objective.

(4) States instructional objective.

One initial item was provided for the teacher evaluator to rate the teacher's overall performance on the lesson.

The four point scale below was used for the pilot study:

| | |
|---|---|
| 1 = Must Improve | Performance jeopardizes continued employment in the district. |
| 2 = Needs Improvement | Performance is below the district expectations. |
| 3 = Meets Standard | Performance meets the expectations set by the district. |
| 4 = Exemplary | Performance exceeds district expectations. |

A total of twenty-three indicators were developed for the Teacher Performance Rating Scale. Items 4 through 7 directly evaluated anticipatory set (see Appendix A). Twelve indicators independent of anticipatory set were used to assess the effects of anticipatory set bias. Each of these teaching behaviors is a strategy teachers utilize which research has indicated is related to student achievement but is not related to the anticipatory set of the lesson. Ratings of these indicators should not be influenced by the anticipatory set. These items were developed to determine the existence of anticipatory set bias. If the evaluators were truly biased by an effective or ineffective anticipatory set, these independent teaching strategies would be influenced. For example,

indicator 12 states "Checks for student understanding." During the body of the lesson, an effective teacher would monitor student understanding of the lesson by asking questions assessing student reaction to the content of the lesson. This teaching strategy occurs after the anticipatory set and directly relates to the content being taught, not the effectiveness of the initial moments of the lesson.

Six indicators on the Teacher Performance Rating Scale were not used in data analysis. These indicators are 1, 2, 3, 16, 18, and 19 (see Appendix A). After careful examination it was determined that these items may not b e independent of the anticipatory set used by the teacher. They were not used in the data analysis. Appendix A shows all indicators. Those items not used in data analysis are starred.

A final item on the Teacher Performance Rating Scale was "Overall Rating." This item asked teacher evaluators to rate the teacher's overall performance in the lesson. It should be noted that overall performance is related to the anticipatory set therefore it is not independent.

A panel of four practitioners skilled in teacher evaluation helped to develop the Teacher Performance Rating Scale by providing feedback on the instrument. (Since this scale was later revised it will be referred to as the Initial Teacher Performance Rating Scale). The practitioners were selected based on an average of ten years administrative experience and 18.5 days training in teacher evaluation. All four were from the same rural/urban district. One

of the practitioners was a teacher evaluator who had completed her doctoral program in educational administration. The other three practitioners had also completed doctoral course work in the educational administration field. These four educators had a combined total of forty years of teacher evaluation experience and the equivalent of seventy-four days of teacher evaluation training (one day equaling six hours of training). They were asked to provide feedback on the following elements of the instrument: clarity of directions, clarity and specificity of criteria, and the validity of the criteria. The indicators were strategies or techniques commonly accepted as reflective of effective teaching and were derived from the research on effective teaching (Hunter, 1984). In the judgment of the panel, four of the twenty-two criteria were directly related to anticipatory set. One additional item was provided for the evaluators to make an overall rating of the lesson. After reviewing the instrument for clarity, specificity, and validity of the criteria, all four practitioners agreed that no changes needed to be made.

Lesson Selection

To conduct the study it was necessary to design a lesson and select a teacher and class for videotaping. A junior high language arts teacher and ninth grade class were selected for videotaping. Permission was then secured from the teacher's school district and her building principal, the teacher, and parents of the students (see Appendix B). The junior high is part of a small urban/rural

district of 4,000 students and 270 teachers. The school contains grades 7 through 9 and houses approximately 850 students. The teacher had been identified by her supervisors, peers, and students, as a master teacher through verbal recommendation. She had acquired a masters degree in English, taught part time at a local college, was in her 31st year of teaching in the public schools, and served as a department chairperson in her building. The teacher had received training in effective instruction, assisted others in her building with these skills, and had consistently and appropriately developed and utilized anticipatory sets for her lessons.

Several discussions with the teacher were held to review the plan for videotaping and develop the lesson and anticipatory set. The lesson was designed for a ninth grade language arts classroom. The lesson, on the novel Great Expectations, was a discussion and analysis of possible endings to the story. Two anticipatory sets were developed for the lesson. One anticipatory set was designed to reflect an effective anticipatory set utilizing the four criteria. The other anticipatory set was designed to reflect an ineffective anticipatory set--one that does not produce the desired effect reflected by the four indicators. The remainder of the lesson was designed to be a "typical lesson"--one typically rated a "good lesson." The ninth grade class selected for videotaping had previously been videotaped. Many students verbally reported that they were not uncomfortable with the presence of the technician and

the equipment. The district audio-visual director videotaped the lesson.

There were three segments for the two videotapes to be used: an effective anticipatory set, an ineffective anticipatory set, and the remainder of the lesson. The effective anticipatory set was taped, then the ineffective anticipatory set. The body of the lesson was then videotaped and used for both tapes. The effective anticipatory set was added to the body of the lesson making one tape--the "effective set" tape. The ineffective anticipatory set and the body of the lesson comprised the second tape--the "ineffective set" tape.

Each anticipatory set was three to five minutes in length followed by a twenty minute lesson. At the beginning of the lesson with the "effective set," directions were displayed on the overhead asking the students to be prepared to discuss their preferred ending for the novel Great Expectations. In the "ineffective" set the teacher took attendance, chatted with students, then told the students the activities that were planned for that class period. The lesson consisted of a discussion of the endings to the novel with students sharing their reasons for selecting their particular ending based on the novel's characters and plot.

The four panel members were asked to determine the teacher's performance level for each anticipatory set and the level of the teacher's performance during the remainder of the lesson using the Initial Teacher Performance Rating Scale. Each panel member was sent two tapes. All four panel members received a lesson absent either

anticipatory set. Two of the panel members each received a tape containing the "effective" anticipatory set and the other two panel members were each sent a tape containing the "ineffective" anticipatory set. The panel members were given two forms of the Initial Teacher Performance Rating Scale--one included the four statements relating to anticipatory set and the other instrument contained the remaining eighteen criteria and overall rating. They were requested to view the tapes separately and rate the lesson first using the instrument containing the eighteen criteria and overall rating. Then they were asked to rate the teacher's anticipatory set using the four indicators directly related to anticipatory set. Two of the panel members were asked to rate the "effective" anticipatory set while the other two were asked to rate the "ineffective" anticipatory set. The panel members then returned the tapes and the instruments to the researcher.

The lesson evaluation by all four panel members resulted in ratings of "Meets Standard" on eight of the twelve indicators independent of anticipatory set. Panel members viewing the effective anticipatory set rated three of the four anticipatory set items "Meets Standard" while the two panel members that viewed the ineffective anticipatory set rated three out of the four items "Needs Improvement." Each teacher evaluator rated the teacher's overall performance during the lesson. The overall performance in the lesson by the teacher was rated as "Meets Standard" by those viewing the

"effective set" tape and as "Needs Improvement" by those viewing the "ineffective set" tape.

During May 1987, a field test was completed. The purpose of this field test was to further develop and validate the rating instrument, and confirm the level of effectiveness of the videotaped lesson and anticipatory sets. Seventeen teachers and administrators from the urban/rural district in which the panel members were employed participated in a one day teacher evaluation workshop conducted by Dr. Jerry Valentine, Professor of Educational Administration from the University of Missouri at Columbia. Permission was received from the district to conduct a field test. All of the teachers (seven) had spent seven days on a committee developing a teacher evaluation process for the district. Six of the ten administrators had also served on a committee which had examined teacher evaluation criteria and rating instruments thus becoming knowledgeable about the appraisal process. Two others who participated in the workshop were building administrators and the remaining two were central office administrators.

The seventeen workshop participants had varying levels of education. Two had bachelor's degrees, seven held master's degrees, two had master's degrees plus thirty hours, five held specialist's degrees, and one held a Ph.D. in educational administration.

The participants had considerable training and experience in teacher evaluation including workshops, staff development activities, and course work. Five had spent ten hours or less, one had spent

from ten to twenty hours, seven had spent from twenty-one to thirty, and four had spent thirty or more hours in teacher evaluation training. Seven participants had less than one year experience evaluating teachers, six had from six to ten years, and four had from ten to fifteen years experience.

During the first part of the six hour workshop, the seventeen participants were provided information on the new district teacher evaluation process including a discussion of the criteria for effective teaching. After three hours, the participants were told that they were going to be randomly divided into two groups and asked to observe and rate a videotaped lesson. Group A would rate the lesson with the "effective" anticipatory set and Group B would rate the lesson with the "ineffective" anticipatory set. Groups A remained in the same room and Group B went to another room. One person from each group was selected to facilitate distribution of the materials necessary for the study. Prior to the participants viewing the lesson, the Information/Direction Sheet, Initial Explanation for Teacher Performance Rating Scale, Initial Teacher Performance Rating Scale, and the Evaluator Data Sheet were distributed to each participant (Appendix A). The participants were then asked to observe the videotape and rate the performance of the teacher using the Initial Teacher Performance Rating Scale.

There was no significant difference between Groups A and B in terms of gender, years experience as a teacher evaluator, and amount of teacher evaluation training. The "effective" anticipatory set was

rated significantly higher than the "ineffective" set tape on all four indicators by Group A. Using the t-test for unmatched pairs (Hinkle, Wiersma, & Jurs, 1979), each statement was significantly different at the .01 level. Group A rated the teacher higher on all indicators in comparison to Group B.

Based on the findings for the other criteria, however, the researcher determined that the four point rating scale on the other thirteen items needed to be extended since the scores tended to group in the middle. A five point scale was adopted and the scale and response modes were changed to the following:

| | |
|---|---|
| 1 = Very Ineffective | Performance is highly unacceptable. Teacher does virtually nothing in this area that is of value to the lesson. |
| 2 = Ineffective | Performance is not at an acceptable level. Teacher has enough deficiencies in this area to be ineffective. |
| 3 = Effective | Performance is acceptable. Teacher demonstrates adequate skill in this area. |
| 4 = Very Effective | Performance is high quality. The teacher is above average in this area but not good enough to serve as a model for others. |
| 5 = Exemplary | Performance serves as a model for other teachers. The teacher demonstrates a high proficiency in this area. |

In order to assess the effectiveness of the revised teacher rating scale, the four panel members were asked to utilize the revised teacher rating scale to rate the same tapes they had rated previously. Each panel member received two tapes. Two panel members received a tape containing the "effective set" and the other two

panel members received a tape containing the "ineffective set." One tape received by all four panel members was the lesson absent either anticipatory set. Each panel member was asked to use the same procedure as when they viewed the tapes the first time.

The average rating for the lesson minus the set was "3," "effective" and the overall rating for the lesson was also a "3." The two panel members rating the "effective" anticipatory set rated each of the four indicators as a "4," "very effective," while the two panel members rating the "ineffective" anticipatory set rated each indicator a "2" indicating each was "ineffective."

Using the Pearson Product Moment Correlation the reliability of the instrument for the field test was calculated. A coefficient of .87 was obtained.

## The Sample

The respondents for this study were selected from teacher evaluation workshops and graduate classes in teacher supervision conducted or taught by Dr. Richard Manatt of Iowa State University. Participants for three teacher evaluation workshops and two graduate supervision classes participated. The workshops were conducted during July and August, 1987. The two graduate classes were taught during the fall of 1987.

### Population

One hundred and six subjects participated in this study. Further information for each site is provided below:

Fort Wayne, Indiana - Eighteen teachers and administrators participated at this site, eight were females and ten were males. The educators had from zero to seven years experience evaluating teachers, averaging .2 years. Their days of teacher evaluation training ranged from zero to fifteen, with an average of 2.4 days.

Erie, Pennsylvania - Forty-five teachers and administrators participated, thirty-seven were male and eight female. Their years of experience evaluating teachers ranged from zero to twenty-five, with an average of 7.0 years. Days spent in teacher evaluation training ranged from zero to thirty with an average of 5.5.

Independence, Kansas - Fourteen supervisors and administrators participated in this study at Independence. Two were female and twelve were male. Experience as teacher evaluators ranged from zero to twenty, with an average of 9.4 years. The group's training in teacher evaluation ranged from zero to thirty days, with an average of 7.8 days.

Ames, Iowa I - Twenty students from this class participated in the study. Gender distribution included ten males and ten females. Their teacher evaluation experience ranged from zero to six years, with an average of 1.1 years. Days spent in teacher evaluation training ranged from zero to seven days, with an average of .9 days.

Ames, Iowa II - This graduate level class contained nine students, two female and seven male. Years of experience as teacher evaluators ranged from zero to seventeen, with an average of 5.9

years. Days spent in teacher evaluation training ranged from six to twenty, with an average of 11.3 days.

## Data Collection

Data for this study were collected through teacher evaluation workshops conducted by Dr. Richard Manatt, and graduate level classes in Educational Administration at Iowa State University, Ames, Iowa. This section of this chapter will describe how data were collected. The final section describes how the data were analyzed.

### Training Design

Data were collected at three teacher evaluation training workshops and two graduate classes in teacher supervision conducted by Dr. Richard Manatt during the months of July, August, and September 1987. The workshop sessions ranged in length from three to five days lasting six and one-half to seven hours per day. For a complete schedule of the workshop refer to Appendix D.

After conducting the workshop for a minimum of one day, Dr. Manatt spent approximately twenty-five minutes reviewing the components of effective instruction using the Hunter model (1984). Anticipatory set was discussed for approximately two or three minutes with the following statement being displayed on an overhead projector screen:

Anticipatory Set

Bringing to a conscious level that which is to be learned and developing a mental readiness for new learning.

Following the discussion of effective instruction, participants were then given a blue or yellow Personal Data Card to complete (see Appendix C). Color coded cards were alternately distributed to participants. Participants were requested to complete the cards and place their social security number at the top of the card. The cards were collected. Participants were assigned to groups on the basis of the color of their cards (blue or yellow). Participants viewed the tapes in separate rooms.

Each participant was then given a Teacher Performance Rating Scale (Revised) and Explanation for Teacher Performance Rating Scale (Revised) and asked to provide their social security number and circle the letter of their group on the instrument (Appendix C). They were told they were going to view a videotape of a lesson. Each participant was asked to rate the teacher's performance, using the rating scale provided and then return the rating scale to the workshop facilitator. They then viewed the videotaped lesson and were asked to complete the rating scale. There was no discussion during or after viewing the videotape. Participants had whatever time was necessary to complete the instrument. It should be noted that all participants finished the instrument at about the same time. Group A observed and rated the "effective set" lesson while Group B observed and rated the "ineffective set" lesson.

The same procedure was used with the graduate classes as was used with the workshops. Following a review of the elements of effective instruction, anticipatory set was briefly discussed. The same

procedures for assigning participants to groups in the workshops were followed in the two graduate classes.

## Data Analysis Procedures

Data for statistical analysis were obtained from the Teacher Performance Rating Scale (Revised) and Personal Data Card. Descriptive statistics describe the sample and raters' responses on the Teacher Performance Rating Scale (Revised). The data were analyzed using StatView, a statistics program for use with the Macintosh personal computer (Feldman and Gagnon, 1985). Descriptive statistics, the t-test for unmatched pairs, and the correlation coefficient were used.

The t-test was used to determine if differences existed in number of years of teacher evaluation and training between Groups A and B. The t-test, an inferential statistic, is designed "to examine the difference between means" (Jendrek, 1985, p. 153).

The correlation coefficient was used to analyze two pieces of information obtained from the Personal Data Card and indicators on the Teacher Performance Rating Scale (Revised) not directly related to anticipatory set. Specifically, a correlation was calculated between: 1) the indicators and the amount of teacher evaluation training held by these evaluators and 2) the indicators and the amount of teacher evaluation experience held by the evaluators. "The correlation coefficient describes the extent to which two sets of data are related" (Hinkle, Wiersma, and Jurs, 1979, p. 71).

All 106 respondents were used in the data analysis. Individual t-tests were run comparing Group A and Group B on the twelve indicators and on the indicator measuring overall teacher performance. T-tests were also calculated comparing evaluators ratings of the four items directly related to anticipatory set.

# CHAPTER IV.

## ANALYSIS AND RESEARCH FINDINGS

The primary purpose of this study was to examine teacher evaluation ratings from five groups participating in teacher evaluation training to determine if anticipatory set bias influences teacher evaluation. Other purposes of the study included assessing the effect of teacher evaluation training and teacher evaluation experience on anticipatory set bias. This chapter is divided into two sections, (1) descriptive data and (2) hypothesis testing.

### Descriptive Data

Descriptive data, presented in Tables 1 through 2b, depict two important evaluator characteristics relevant to this study: days spent in teacher evaluation training and the number of years of teacher evaluation experience. Table 1 presents this information by site combining Group A, participants viewing the effective set tape, and Group B, participants viewing the ineffective set tape. Tables 2a and 2b display the data, reporting days of teacher evaluation training and teacher evaluation experience for the randomly assigned groups.

Table 1 shows the number of evaluators at each site. A total of 106 evaluators participated in this study with the largest group

(45) representing Erie, Pennsylvania and the smallest (9) coming from students enrolled in an advanced teacher evaluation course at Ames, Iowa. The average participant had 4.9 days of evaluation training, six hours being equal to one day teacher evaluation training. Each participant had been a teacher evaluator for an average of 5.0 years (see Table 1).

Table 1. Mean days teacher evaluation training and experience of evaluators (N=106 evaluators)

| Site | Number of Evaluators | $\bar{x}$ Days Teacher Evaluation Training | $\bar{x}$ Years Teacher Evaluation Exp. |
|---|---|---|---|
| Fort Wayne, Indiana | 18 | 2.4 | .2 |
| Erie, Pennsylvania | 45 | 5.5 | 7.0 |
| Independence, Kansas | 14 | 7.8 | 9.4 |
| Iowa State University | 20 | .9 | 1.1 |
| Iowa State University | 9 | 11.3 | 5.9 |
| | 106 | $\bar{x}$ = 4.90 | $\bar{x}$ = 5.0 |

The two Iowa State University sites had the most and least amount of teacher evaluation training. The students in the graduate level course in beginning teacher supervision averaged .9 days

training while the participants in the advanced course averaged 11.3 days.

Fort Wayne, Indiana had the lowest mean score in years experience evaluating teachers with .2 years. The highest score was reported by the site at Independence, Kansas with 9.4 years experience.

Table 2a provides the data regarding training and experience for the participants who viewed the effective set lesson. The combined average number of days of teacher evaluation training for all sites was 4.4 days and the average number of years as a teacher evaluator

Table 2a. Mean days teacher evaluation training and mean years teacher evaluation experience for effective set lesson evaluators (N=52)

| Site | Number of Evaluators | $\bar{x}$ Days Teacher Evaluation Training | $\bar{x}$ Years Teacher Evaluation Exp. |
|---|---|---|---|
| Fort Wayne, Indiana | 9 | 2.6 | .3 |
| Erie, Pennsylvania | 21 | 4.9 | 8.4 |
| Independence, Kansas | 7 | 6.9 | 8.4 |
| Iowa State University | 11 | 1.2 | 1.2 |
| Iowa State University | 4 | 10.8 | 2.8 |
| | N=52 | $\bar{x}$=4.4 | $\bar{x}$=5.0 |

was 5.0. The groups at Iowa State University had the lowest and highest mean score in days of teacher evaluation training with the beginning teacher supervision course reporting 1.2 days and the advanced site reporting 10.8 days training. Fort Wayne, Indiana had the lowest mean score in years of teacher evaluation experience with .3 while Erie, Pennsylvania and Independence, Kansas each obtained a mean score of 8.4 years experience.

Table 2b displays the number of evaluators, extent of teacher evaluation training and years teacher evaluation experience by site for the ineffective set group. The average participant had spent 5.4

Table 2b. Mean days teacher evaluation training and mean years teacher evaluation experience for ineffective set lesson evaluators (N=54)

| Site | Number of Evaluators | $\bar{x}$ Days Teacher Evaluation Training | $\bar{x}$ Years Teacher Evaluation Exp. |
|---|---|---|---|
| Fort Wayne, Indiana | 9 | 2.3 | .02 |
| Erie, Pennsylvania | 24 | 6.1 | 8.8 |
| Independence, Kansas | 7 | 8.7 | 10.4 |
| Iowa State University | 9 | .6 | 1.1 |
| Iowa State University | 5 | 11.8 | 9.0 |
| | N=54 | $\bar{x}$=5.4 | $\bar{x}$=6.3 |

days in evaluation training. The combined mean score in years for teacher evaluation experience was 6.3. The students enrolled in the beginning teacher supervision class at Iowa State University had the least evaluation training (.6 days) while those enrolled in the advanced supervision class had the most (11.8 days). Participants at Fort Wayne, Indiana had the least years experience evaluating teachers (.02) and Independence, Kansas reported the highest score (10.4).

It was necessary to determine if there was a difference in the amount of training and years of teacher evaluation experience between those who rated the effective set lesson and those who rated the ineffective set lesson (see Tables 3a and 3b). To assess the differences between groups a t-test was calculated. The difference was not significant at the .05 level for either days of teacher evaluation training or years of teacher evaluation experience.

Table 3a. T-test analysis for significance of difference in mean scores of days of teacher evaluation training between the effective set lesson evaluators and ineffective set lesson evaluators (N=106)

| df | Effective Set $\bar{x}$ score | Ineffective Set $\bar{x}$ score | t-value |
|---|---|---|---|
| 104 | 4.404 | 5.389 | -.788 |

Table 3b. T-test analysis for significance of difference in mean scores of years of teacher evaluation experience between the effective set and ineffective evaluators (N=106)

| df | Effective Set $\bar{x}$ score | Ineffective Set $\bar{x}$ score | t-value |
|---|---|---|---|
| 104 | 5.000 | 6.296 | -1.083 |

Table 4 shows the mean scores, standard deviation, and difference between means for the anticipatory set indicators, the twelve teaching strategies, and the overall teacher performance rating.

Anticipatory Set

The mean scores for the indicators for the "effective set lesson" were considerably higher than those for the "ineffective set lesson." The largest difference was on, "Involves students in learning new objective" with a difference of .812. The smallest difference was on indicator 1, "Relates current lesson to previous learning," mean difference of .606.

A summary mean score for anticipatory set was obtained by summing the responses for the four indicators and dividing by the number of respondents. Evaluators viewing the effective set lesson rated it 3.885 and those who observed the ineffective set lesson rated it 3.185, a mean difference of .7.

A t-test was calculated to determine the difference between the ratings by the effective set lesson evaluators and the ineffective

Table 4. Mean Scores, standard deviations, and mean differences by indicator on the Teacher Performance Rating Scale (N=106)

| INDICATORS | Effective Set Lesson N=52 | | Ineffective Set Lesson N=54 | | |
|---|---|---|---|---|---|
| | M | SD | M | SD | Diff |
| ANTICIPATORY SET | | | | | |
| Relates current lesson to previous learning. | 3.865 | .817 | 3.259 | .894 | .606 |
| Provides focus for new learning. | 3.885 | .900 | 3.167 | .818 | .718 |
| Involves students in learning new objective. | 3.923 | .813 | 3.111 | .818 | .812 |
| States instructional objective. | 3.865 | 1.085 | 3.204 | 1.035 | .661 |
| INDEPENDENT TEACHING STRATEGIES | | | | | |
| Provides a clear explanation of new material. | 3.769 | .831 | 3.185 | .803 | .584 |
| Provides clear directions. | 3.827 | .857 | 3.278 | .878 | .549 |
| Incorporates effective questioning techniques. | 4.000 | .816 | 3.222 | 1.093 | .778 |
| Uses demonstrations, examples, and anecdotes to teach the lesson. | 3.519 | .874 | 2.926 | .887 | .593 |
| Checks for student understanding. | 3.462 | .956 | 3.037 | 1.009 | .425 |
| Paces lesson appropriately and/or adjusts as needed. | 3.500 | .828 | 2.963 | .951 | .537 |
| Gives supportive and immediate feedback to students. | 3.808 | .793 | 3.259 | .935 | .549 |
| Provides opportunities for student participation. | 3.942 | .669 | 3.426 | .838 | .516 |

Table 4. - Continued

| INDICATORS | Effective Set Lesson N=52 | | Ineffective Set Lesson N=54 | | |
|---|---|---|---|---|---|
| | M | SD | M | SD | Diff |
| Displays a thorough knowledge of subject matter. | 4.500 | .754 | 4.148 | .920 | .352 |
| Maintains a high standard for student behavior. | 4.269 | .630 | 3.481 | .795 | .788 |
| Demonstrates sensitivity in relating to students. | 3.673 | .873 | 3.148 | .920 | .525 |
| Involves students in summary of lesson. | 3.538 | .939 | 2.852 | 1.106 | .686 |
| OVERALL LESSON PERFORMANCE RATING | | | | | |
| Rate overall performance using the following responses. | 3.962 | .625 | 2.98 | .687 | .982 |
| 1 = Very Ineffective<br>2 = Ineffective<br>3 = Effective<br>4 = Very Effective<br>5 = Exemplary | | | | | |

set lesson evaluators on ratings of the four indicators. A t-value of 15.895 was obtained. This was significant at the .0005 level (see Table 5).

Independent Teaching Strategies The rating of teaching strategies required by the indicators independent of anticipatory set but relevant to the remainder of the lesson were consistently higher by the effective set group than the ratings by the ineffective set

group. The largest difference in the ratings was on indicator 20, "Maintains a high standard for student behavior" (.788) while indicator 17, "Displays a thorough knowledge of subject matter," showed the least difference (.352).

Overall Performance Indicator 23 was provided for each evaluator to rate the teacher's overall performance. The effective set group rated the overall lesson 3.962. The ineffective set group rated the lesson 2.98 with a standard deviation of .687. The difference in overall ratings was .982.

Table 5. T-test analysis of difference between mean scores on anticipatory set indicators of effective set lesson and ineffective set lesson (N=106)

| df | Effective set $\bar{x}$ score | Ineffective set $\bar{x}$ score | t-value |
|---|---|---|---|
| 3 | 3.885 | 3.185 | 15.895* |

* Significant at the .0005 level.

After careful inspection of the indicators it was determined that the teaching strategies represented by the following six indicators may have some relevance to anticipatory set.

- Demonstrates effective personal organization skills.
- Organizes students for effective instruction.
- Provides the structure for learning.
- Models effective communication skills.

- Incorporates effective communication skills.

- Ensures student time on task.

Anticipatory set may influence the ratings on teaching strategies reflected in these six indicators. For example, some evaluators may believe the teacher is not using student contact time effectively if they have an ineffective lesson beginning. Thus, they may mark the teacher lower on "Ensures student time on task" than if the teacher demonstrated an effective anticipatory set. Evaluator ratings of these six indicators may have been justifiably influenced by the effective or ineffective anticipatory set. As a result of this possibility, these indicators were not included in the hypotheses testing.

## Hypothesis Testing

Each of the questions posed in this study resulted in a specific research hypotheses. All hypotheses were tested for significance at the .05 level with probabilities less than .05 also reported. Hypotheses are presented and discussed in the order of the questions posed by the study. Table 6 shows the t-value and level of significance for each indicator.

Hypothesis 1. There is no significant difference in lesson ratings of evaluators who observe a teaching lesson with an effective anticipatory set and those evaluators who observe a teaching lesson with an ineffective anticipatory set in their ratings of overall teacher performance.

This hypothesis was formulated to determine if anticipatory set bias affects evaluators' ratings of the teacher's overall performance. The Overall Teacher Performance Rating on the Teacher Performance Rating Scale was used to test this hypothesis. Using the t-test for unmatched pairs, a t-value of 7.677 was obtained. This t-value is significant at the .0005 level (Table 6).

On the basis of these tests, the null hypothesis was rejected. Evaluators who observed a teaching lesson with an effective anticipatory set rated overall performance significantly higher than those evaluators who observed a teaching lesson with an ineffective anticipatory set.

Hypothesis 2. There is no significant difference between evaluators who observe a teaching lesson with an effective anticipatory set and evaluators who observe a teaching lesson with an ineffective anticipatory set on ratings of eight of twelve teaching strategies independent of anticipatory set.

The hypothesis was formulated to determine if anticipatory set bias affects evaluators' ratings of other teaching strategies independent of anticipatory set but directly related to the body of the lesson. The t-test for unmatched pairs was computed for the effective set group and the ineffective set group on each indicator (see Table 6). All t-values were found to be significant at the .025 to .0005 level with seven indicators significant at the .0005 level.

On the basis of these results the null hypothesis was rejected. Evaluators who observe a teaching lesson with an effective anticipatory set rated teaching behaviors independent of the

Table 6. T-test analysis for significance of difference between mean scores on anticipatory set indicators and indicators relevant to the body of the lesson (N=106)

| Indicators | t-value |
|---|---|
| INDEPENDENT TEACHING STRATEGIES | |
| Provides a clear explanation of new material. | 3.680***** |
| Provides clear directions. | 3.258**** |
| Incorporates effective questioning techniques. | 4.138***** |
| Uses demonstrations, examples, and anecdotes to teach the lesson. | 3.467***** |
| Checks for student understanding. | 2.219** |
| Paces lesson appropriately and/or adjusts as needed. | 3.095**** |
| Gives supportive and immediate feedback to students. | 3.250**** |
| Provides opportunities for student participation. | 3.498***** |
| Displays a thorough knowledge of subject matter. | 2.150** |
| Maintains a high standard for student behavior. | 5.642***** |
| Demonstrates sensitivity in relating to students. | 3.048**** |
| Involves students in summary of lesson. | 3.440***** |
| OVERALL TEACHER PERFORMANCE RATING | |
| Rate overall performance using the following responses. | 7.677***** |

** Significant at <.025 level.
**** Significant at <.005 level.
***** Significant at <.0005 level.

anticipatory set higher than those evaluators who observed a teaching lesson with an ineffective anticipatory set.

Hypothesis 3. There is no significant positive relationship between mean evaluator ratings of eight out of twelve teaching strategies independent of anticipatory set and amount of teacher evaluation training of the evaluator.

This hypothesis was tested by utilizing information from the Personal Data Cards and ratings on the Teacher Performance Rating Scale. Evaluators reported their amount of teacher evaluation training in days, a single day being equal to six hours. The correlation coefficient was used to compare the amount of training held by each evaluator and their ratings of teacher performance indicators pairing each evaluator with their rating on the indicators. A correlation coefficient was calculated for the effective set group and the ineffective set group at the .05 level of significance. A coefficient of .273 or higher was considered significant.

Table 7a shows the results of the calculated correlation between ratings of those who observed the effective set lesson and the twelve indicators. None of the indicators was significant at the .05 level. The results for those who rated the ineffective set lesson are displayed in Table 7b. Again, no indicator was significant at the .05 level.

On the basis of these tests, hypothesis 3 was not rejected. There was no significant positive relationship between evaluator

Table 7a. Correlation between the amount of teacher evaluation training of evaluators in the effective set group and ratings of indicators independent of anticipatory set (N=52)

| Indicators | Covariance | Correlation | R-squared |
|---|---|---|---|
| Provides focus for new learning | .734 | .124 | .015 |
| Provides clear directions | -.399 | -.071 | .005 |
| Incorporates effective questioning techniques | .255 | .047 | .002 |
| Uses demonstrations, examples and anecdotes to teach the lesson | .041 | .007 | .00005108 |
| Checks for student understanding | -1.151 | -.182 | .033 |
| Paces lesson appropriately and/or adjusts as needed | -.578 | -.106 | .011 |
| Gives supportive and immediate feedback to students | .452 | .087 | .007 |
| Provides opportunities for student participation | -.427 | -.097 | .009 |
| Displays a thorough knowledge of subject matter | .382 | .077 | .006 |
| Maintains a high standard for student behavior | .144 | .035 | .001 |
| Demonstrates sensitivity in relating to students | -.630 | -.112 | .012 |
| Involves students in summary lesson | -.241 | - .039 | .002 |

ratings of teaching strategies independent of anticipatory set and the amount of teacher evaluation training of the evaluator.

Hypothesis 4. There is no significant positive relationship between mean teacher evaluator ratings of eight out of twelve teaching strategies independent of anticipatory set and amount of teacher evaluation experience of the evaluator.

A similar comparison approach was used to evaluate this hypothesis as was used with hypothesis 3. Data were obtained from the Personal Data Card and the Teacher Performance Rating Scale. Correlations were calculated comparing mean teacher evaluator ratings on indicators in the effective set group and in the ineffective set group pairing evaluator ratings on each indicator with the number of years experience held by the evaluator.

Table 8a displays the results of the correlation coefficient computed for the effective set group. None of the indicators displayed significance at the .05 level. Table 8b displays the results for the ineffective set group. No indicator was significant at the .05 level.

Based on these results, the null hypothesis was not rejected at the .05 level. Thus there is no significant positive correlation between teaching strategies independent of anticipatory set and teacher evaluation experience of the evaluator.

Table 7b. Correlation between the amount of teacher evaluation training of evaluators in the ineffective set group and ratings of indicators independent of anticipatory set (N=54)

| Indicators | Covariance | Correlation | R-squared |
|---|---|---|---|
| Provides focus for new learning | -2.085 | -.405 | .164 |
| Provides clear directions | -.431 | -.078 | .006 |
| Incorporates effective questioning techniques | -.616 | -.09 | .008 |
| Uses demonstrations, examples, and anecdotes to teach the lesson | -.725 | -.13 | .017 |
| Checks for student understanding | -.807 | -.127 | .016 |
| Paces lesson appropriately and/or adjusts as needed | -.797 | -.133 | .018 |
| Gives supportive and immediate feedback to students | -.801 | -.136 | .019 |
| Provides opportunities for student participation | -.376 | -.07 | .005 |
| Displays a thorough knowledge of subject matter | -.889 | -.154 | .024 |
| Maintains a high standard for student behavior | -.191 | -.038 | .001 |
| Demonstrates sensitivity in relating to students | -.776 | -.134 | .018 |
| Involves students in summary of lesson | -.658 | -.095 | .009 |

Table 8a. Correlation between the amount of teacher evaluation experience of the evaluators in the effective set group and ratings of indicators independent of the anticipatory set (N = 52)

| Indicator | Covariance | Correlation | R-squared |
|---|---|---|---|
| Provides focus for new learning | 1.373 | .241 | .058 |
| Provides clear directions | .725 | .134 | .018 |
| Incorporates effective questioning techniques | .941 | .182 | .033 |
| Uses demonstrations, examples, and anecdotes to teach the lesson | .804 | .145 | .021 |
| Checks for student understanding | -.02 | -.003 | .00001042 |
| Paces lesson appropriately and/or adjusts as needed | .471 | .09 | .008 |
| Gives supportive and immediate feedback to students | .412 | -.082 | .007 |
| Provides opportunities for student participation | .686 | .162 | .026 |
| Displays a thorough knowledge of subject matter | .549 | .115 | .013 |
| Maintains a high standard for student behavior | .784 | .197 | .039 |
| Demonstrates sensitivity in relating to students | .588 | .108 | .012 |
| Involves students in summary of lesson | .059 | .01 | .00009792 |

Table 8b. Correlation between the amount of teacher evaluation experience of the evaluators in the ineffective set group and ratings of indicators independent of anticipatory set (N = 54)

| Indicator | Covariance | Correlation | R-squared |
|---|---|---|---|
| Provides focus for new learning | .138 | .028 | .001 |
| Provides clear directions | .803 | .153 | .023 |
| Incorporates effective questioning techniques | 1.405 | .215 | .046 |
| Uses demonstrations, examples, and anecdotes to teach the lesson | .626 | .118 | .014 |
| Checks for student understanding | 1.479 | .245 | .06 |
| Paces lesson appropriately and/or adjusts as needed | .955 | .168 | .028 |
| Gives supportive and immediate feedback to students | .79 | .141 | .02 |
| Provides opportunities for student participation | .909 | .181 | .033 |
| Displays a thorough knowledge of subject matter | .352 | .064 | .004 |
| Maintains a high standard for student behavior | 1.119 | .235 | .055 |
| Demonstrates sensitivity in relating to students | 1.238 | .225 | .051 |
| Involves students in summary of lesson | .233 | .035 | .001 |

# CHAPTER V.

# SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

## Summary

The primary purposes of this study were to (1) examine teacher evaluation ratings from five groups participating in teacher evaluation training in an effort to assess the level of anticipatory set bias, and (2) to assess the effect of teacher evaluation training and teacher evaluation experience on anticipatory set bias. In essence, the study attempted to establish the degree to which an effective or ineffective set influences evaluator ratings of teacher overall performance and teaching strategies independent of anticipatory set.

A summary of the findings based on data gathered in the summer of 1987 from those participating in teacher evaluation training follows.

## Conclusions

### Findings

This study has very important implications for teachers, evaluators, and those who train teachers, for teacher evaluators, for those who train both groups and even for student learning. If the research indicated in this study is supported by further research,

one might conclude that without remedy, the evaluation that results from a lesson observation is often invalid, overly influenced by teacher performance at the beginning of the lesson. This, of course, would have a tremendous effect on performance based pay, on tenure, on the identification of job targets, and on a myriad of other issues related to instruction.

One hundred six educators involved in teacher evaluation training in Erie, Pennsylvania; Fort Wayne, Indiana; Independence, Kansas; and Ames, Iowa provided data for the study. During the summer of 1987, they participated in an activity designed to assess the extent of anticipatory set bias. The findings follow:

1. The level of efficacy of teacher anticipatory set significantly influenced evaluators' ratings of the teacher's overall performance.

2. The level of efficacy of teacher anticipatory set significantly influenced evaluators' rating of teaching strategies which were independent of anticipatory set.

3. The amount of teacher evaluation training did not affect the evaluator's tendency to be influenced by anticipatory set.

4. The extent of teacher evaluator experience did not influence the evaluator's tendency to be influenced by anticipatory set.

The implications of these findings are significant. Some have powerful implications and others are more subtle.

Two highly significant findings resulted from this study: 1) anticipatory set bias influences the overall rating of teacher performance given by an evaluator and 2) anticipatory set bias influences evaluator ratings of independent teaching strategies.

Evaluators who viewed the effective set tape were significantly more likely to rate the teacher's overall performance and teaching strategies independent of anticipatory set higher than those who viewed the ineffective set tape despite the fact that the performance and strategies rated by both groups were exactly the same. It appears that the first few moments of a teaching lesson have a powerful impact in evaluator's perception and may invalidate evaluator ratings of the lesson. Evaluators who are overly influenced by the initial impression made by the teacher apparently generalize that feeling to other teaching behaviors or strategies. Or perhaps as in business, they decide teacher performance based on their first impression (Weitzul, 1983).

It follows then that logical outcomes of lesson observation such as the post conference and growth plans are also influenced. The impact of the anticipatory set gets in the way of the "real performance" that occurred in the classroom and distorts the coaching process. The evaluator is unable to separate the effectiveness of each independent teaching strategy or behavior, thus developing a growth plan with the teacher that does not accurately reflect upon the teacher's performance.

These may be minor errors when one compares their impact to the significance of set bias in granting teachers tenure or in terminating tenured teachers. Teachers who have an ineffective anticipatory set may get low ratings on many teaching strategies. By the same token teachers who set the stage well but do not teach the

body of the lesson well may still receive high ratings in all areas. If the evaluator is not aware of this bias, the teacher who does not set the stage well may not be granted tenure or be in peril of termination.

Pay for performance may suffer similar problems. A teacher who demonstrates effective anticipatory set during lesson observations may receive inflated ratings on other teaching strategies. Conversely, the teacher who begins the lesson poorly but demonstrates the effective use of teaching strategies during the lesson may receive deflated ratings. These ratings frequently provide the basis for compensation. A teacher may be granted or denied compensation based on invalid ratings of performance (Cornett, 1985).

Two very distinct events occur after viewing a lesson with an effective or ineffective set. Evaluators rating teaching behaviors after viewing a lesson with an effective anticipatory set escalated teacher performance ratings while those who viewed a lesson with an ineffective set de-escalated their ratings of the same teaching behaviors. These factors will now be referred to as the Initial Perception Escalator or IPE and the Initial Perception De-escalator or IPD.

If the anticipatory set has such a significant effect on the evaluator, is it possible that the impact on students in the classroom may be just as powerful? If we make decisions whether or not to view a movie in its entirety on the initial moments (or set) of the film, is it possible that tuning into a lesson may be

determined by the students initial perception of the lesson? If the stage is set well, the excitement or need for learning and focus of the lesson are established, do students learn more from what follows? If so, improving anticipatory set would be an excellent goal for all teachers and trainers of teachers.

There are other interesting findings. These revolve around differences in how discrete teacher behaviors were assessed, i.e. the apparent difference in how each teaching strategy was influenced by set. Two indicators on the Teacher Performance Rating Scale deserve particular attention. They are "Maintains a high standard for student behavior" and "Displays thorough knowledge of subject matter." Neither had any connection with the anticipatory set, yet, manages student behavior was the teaching strategy which appeared to be most influenced by the set. The first impression of organization, control, and energy given by a teacher in the classroom significantly impacts the evaluator's perception of student behavior during the rest of the lesson (Valentine, 1984). Since evaluators, particularly principals, place a high premium on student management this may have a particularly powerful effect on the overall rating.

It is also interesting to note the area least influenced by bias was "Displays thorough knowledge of subject matter." Either evaluators tend not to connect teaching strategies with content knowledge or they may make little effort to determine the teacher's level of expertise in the content area. If the latter is the case, is it fair to ask evaluators to rate teachers in other areas? Or

should knowledge of subject matter be evaluated by subject matter specialists?

The results of this study failed to show a positive relationship between evaluator training or evaluator experience and their ability to overcome anticipatory set bias. It is hardly surprising since currently little training related to anticipatory set bias seems to be happening. Evaluators who are unaware of this bias will continue to exhibit this tendency (Goldhammer, 1969). Experience is not a known deterrent to averting bias; apparently teacher evaluators tend to repeat past practices unless they receive training otherwise.

Lesson analysis and ratings are key elements in the teacher evaluation process and appear to be inappropriately influenced by anticipatory set (Brophy, 1979). Typically post conferences are based on ratings made by the evaluator and inferences based on those ratings. When these ratings are skewed, they become incorrectly reflected in the post conference. And since anticipatory set bias may have a major impact on observation and ratings, the entire performance appraisal process may benefit from this information.

Teachers are the ones who may benefit greatly from this study if the findings are disseminated and adjustments made. Staff development activities should emphasize the development of an effective anticipatory set and the role it plays in instruction and evaluation. Currently little is done to address this important need. Teacher preparation programs should also stress the significance that

anticipatory set may have in teacher evaluation and student achievement and train new teachers accordingly.

## Recommendations for Practitioners

Below are two recommendations which could make a difference in our schools.

1. If anticipatory set bias influences evaluator ratings, it may also influence the learner. Teachers should be aware of its potential impact. Staff development activities should focus on the development of effective anticipatory set and the role it plays in student learning and evaluation.

2. Teacher evaluators should be made aware of anticipatory set bias and the Initial Perception effect. Discussion and awareness training of the Initial Perception effect and its effect on evaluator ratings, professional growth targets, financial compensation, and student achievement should be included in evaluator training.

## Recommendations for Further Research

Below are suggestions or recommendations for further research. This study was an initial foray into an unexplored, ambiguous area. Replication and more work are needed to confirm findings and analyze the phenomenon more specifically.

1. This study should be replicated in other settings with other lessons, teachers, and evaluators.

2. The effect of anticipatory set bias on the evaluator and teacher have been addressed in this study. Additional research on

the effect of set on students and student bias toward instruction is suggested.

3. Further research efforts should address the effect of a neutral anticipatory set and/or lack of anticipatory set in a teaching lesson.

4. Performance bias can be reduced through awareness and training. A study of the effect of evaluator training on teacher ratings would be helpful.

5. An effective anticipatory set has a positive impact on teacher ratings. Studying evaluator ratings and student perceptions before and after training teachers on effective sets may aid in assessing the impact of this bias.

6. Anticipatory set bias may not be the only bias affecting teacher ratings. Similar studies should be conducted assessing the effects of other possible instruction biases affecting student achievement.

# BIBLIOGRAPHY

Allport, G. W. The Nature of Prejudice. Reading, MA: Addison-Wesley Publishing Co., Inc., 1954.

Barnes, J. L. "Age as a Perceptual Cue in the Performance Appraisal Process: The Relationship Between Perceived Age and Performance Ratings of Exempt Employees." Ph.D. dissertation, The Pennsylvania State University, 1980.

Barnes, J. L., and Landy, F. J. Scaling Behavioral Anchors. Applied Psychological Measurement 3 (1979): 193-200.

Bass, B. M. "Reducing Leniency in Merit Ratings." Personnel Psychology 9 (1956): 359-369.

Beebe, R. "Developing Sound Performance Appraisal Procedures." National Association of Secondary School Principal's Bulletin 71, 497 (1987): 96-101.

Berman, J. S., and Kenny, D. A. "Correlational Bias in Observer Ratings." Journal of Personality and Social Psychology 34, 2 (1976): 263-273.

Bernardin, H. J. "Effects of Rater Training on Leniency and Halo Errors in Student Ratings of Instructors." Journal of Applied Psychology 63 (1978): 301-308.

Bernardin, H. J. "Rater Training to Enhance Accuracy: A look at the Big Picture." Paper presented at Performance Appraisal Conference, Old Dominion University, Norfolk, VA, April 1980.

Bernardin, H. J., and Smith, P. C. "A Clarification of Some Issues Regarding the Development and Use of Behaviorly Anchored Rating Scales (BARS)." Journal of Applied Psychology 66 (1981): 458-463.

Blackmer, D., et al. School Improvement Model Teacher Performance Criteria with Response Modes and Standards Publication No. 81-2. Ames, Iowa: School Improvement Model Project, Iowa State University, 1981.

Bolton, D. Selection and Evaluation of Teachers. Berkeley, CA: McCutchan Publishing, 1980.

Borich, G. The Appraisal of Teaching: Concepts and Process. Reading, MA: Addison-Wesley Publishing Co., 1977.

Borman, W. C. "Effect of Instructions to Avoid Halo Error on Reliability and Validity of Performance Evaluation Ratings." Journal of Applied Psychology 60 (1975): 556-560.

Borman, W. C. "Format and Training Effects on Rating Accuracy and Rater Errors." Journal of Applied Psychology 64 (1979): 410-421.

Brandt, R. "On Teacher Evaluation: A Conversation with Tom McGreal." Educational Leadership 44, 7 (April 1987): 20-24.

Brophy, J. E. Using Observation to Improve Your Teaching. East Lansing, MI: Institute for Research on Teaching, Michigan State University, Fall 1979.

Brown, E. M. "Influence of Training, Method, and Relationship on the Halo Effect." Journal of Applied Psychology 52 (1968): 195-199.

Buckner, D. N. "The Predictability of Ratings as a Function of Interrater Agreement." Journal of Applied Psychology 43 (1959): 60-64.

Burnaska, R. F., and Hollman, T. D. "Characteristics of Effective Employee Performance Review and Development Interviews." Journal of Applied Psychology 59 (1974): 307-312.

Buttram, J. L. and Wilson, B. L. "Promising Trends in Teacher Evaluation." Educational Leadership 44, 7 (April 1987): 4-6.

Cahen, L. "An Experimental Manipulation of the Halo Effect." Ph.D. dissertation, Stanford University, 1966.

Campbell, D. T., and Stanley, J. C. "Experimental and Quasi-Experimental Designs for Research on Teaching." Handbook on Research on Teaching, edited by N.L. Gage. Chicago, IL: Rand McNally and Co., 1966.

Cascio, W. F., and Valenzi, E. R. "Behaviorally Anchored Rating Scales: Effects of Education and Job Experience of Raters and Rates." Journal of Applied Psychology 62 (1977): 278-282.

Christner, C. A. "The J.R. Syndrome: Administrator Bias in Teacher Evaluation." Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA, 13-17 April 1981.

Cooper, W. H. "Conceptual Similarity as a Source of Illusory Halo in Job Performance Ratings." Journal of Applied Psychology 66 (1981): 302-307.

Cordaro, L., and Ison, J. R. "Psychology of the Scientist: X-Observer Bias in Classical Conditioning of the Planarian." Psychological Reports 13 (1963): 787-789.

Cornett, L. M. "Trends and Emerging Issues in Career Ladder Plans." Educational Leadership 43, 3 (November 1985): 6-10.

Crooks, L. A., ed. An Investigation of Sources of Bias in the Prediction of Job Performance. Princeton, N.J.: Educational Testing Service, 1972.

Dickenson, T. L., and Zellinger, P. M. "A Comparison of the Behaviorally Anchored Rating and Mixed Standard Scale Formats." Journal of Applied Psychology 65 (1980): 147-154.

Downie, N. M., and Heath, R. W. Basic Statistical Methods. 4th ed. New York, NY: Harper and Rowe, Publishers, 1974.

Duke, D., and Stiggins, R. Teacher Evaluation: Five Keys to Growth. Washington, D.C.: National Education Association, 1986.

Edmonds, R., and Fredericksen, J. Search for Effective Schools. Brooklyn, NY: Board of Education, Office of the Chancellor, 1979.

Elashoff, J. and Snow R. Pygmalion Reconsidered. Worthing, Ohio: Charles A. Jones Publishing Company, 1971.

Elmore, P. B., and LaPointe, K. A. "Effect of Teacher Sex, Student Race, and Teacher Warmth on the Evaluation of College Instructors." Journal of Educational Psychology 67 (1975): 368-374.

Feldman, D., and Gagnon, J. Statview: The Graphics, Statistics Utility for Macintosh. Calavasas, CA: Brain Power, 1985.

Fiske, D. W., and Cox, J. A., Jr. "The Consistency of Ratings by Peers." Journal of Applied Psychology 44 (1960): 11-17.

Fletcher, C., and Williams, R. Performance Appraisal and Career Development. Brookfield, VT: Brookfield Publishing Company, 1985.

Frudden, S. A Comparison of School Administrators' Ratings of Teacher Performance Utilizing Varied Instructional Materials in the Assessment Process. Ames, IA: Iowa State University, 1982.

Frudden, S., and Manatt, R. P. Lesson Analysis--The Neglected Key to Teacher Performance Evaluation. Ames, IA: Iowa State University, 1982.

The Georgia Assessment Project. The Teacher Assessment Project. Athens, GA: The University of Georgia, 1980.

Geosits, M. S. "An Investigation of Bias in Appraising Teachers Using an Open Approach to Instruction." Ph.D. dissertation, Lehigh University, 1978.

Goldhammer, R. Clinical Supervision: Special Methods for the Supervision of Teachers. New York: Holt, Rinehart and Winston, Inc., 1969.

Gorden, M. E. "The Effect of the Correctness of the Behavior Observed on the Accuracy of Ratings." Organizational Behavior and Human Performance 5 (1970): 366-377.

Haberman, M. "The Relationship of Bogus Expectation to Success in Student Teaching (or Pygmalion's Illegitimate Son)." Journal of Teacher Education 21 (1970): 69-72.

Harrington, L. "The Association of Administrators' Learning Styles, Background Differences, and Teacher Appraisal Judgments." Ph.D. dissertation, Iowa State University, Ames, IA, 1984.

Hastie, R. "Memory for Behavioral Information That Confirms or Contradicts a Personality Impression." R. Hastie, T. Olstrom, E. Ebbesen, R. Wyer, D. Hamilton, and D. Carlston (Eds.), Person Memory: The Cognitive Basis of Social Perception. Hillsdale, N.J.: Lawrence Erlbaum, 1980.

Hinkle, D. E.; Wiersma, W.; and Jurs, S. G. Applied Statistics for the Behavioral Sciences. Boston, MA: Houghton Mifflin Co., 1979.

Holdzkom, D. "Appraising Teacher Performance in North Carolina." Educational Leadership 44, 7 (April 1987): 40-44.

Hollander, E. P. "Validity of Peer Nominations in Predicting a Distant Performance Criterion." Journal of Applied Psychology 49 (1969): 434-438.

Hosford, P. L., ed. Using What We Know About Teaching. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1984.

Hunter, M. Improved Instruction. El Segundo, CA: Theory-Into-Practice Publications, 1976.

Hunter, M. "Knowing, Teaching, and Supervising." P. L. Hosford. Using What We Know about Teaching. Alexandria, VA: Association for Supervision and Curriculum Development, 1984.

James, L. R. "Criterion Models and Construct Validity for Criteria." Psychological Bulletin 80 (1973): 75-83.

Jendrek, M. P. Through the Maze: Statistics with Computer Applications. Belmont, California: Wadsworth Publishing, 1985.

Joyce, B., and Werl, M. Models of Teaching. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1972.

Kingstrom, P. O., and Bass, A. R. "A Critical Analysis of Studies Comparing Behaviorally Anchored Rating Scales (BARS) and Other Rating Formats." Personnel Psychology 34 (1981): 263-289.

Klieger, W. A., and Mosel, J. N. "The Effect of Opportunity to Observe and Rater Status on the Reliability of Performance Ratings." Personnel Psychology 6 (1953): 57-64.

Klores, M. S. "Rater Bias in Forced-Distribution Ratings." Personnel Psychology 19 (1966): 411-421.

Kolman, A. S. Definition of the Situation and Observer Bias. (ERIC ED 118 600).

Kowalski, J. P. Evaluating Teacher Performance. Arlington, VA: Research Service Inc., 1978.

Landy, F. J., and Farr, J. L. The Measurement of Work Performance: Methods, Theory and Applications. New York, NY: Academic Press, 1983.

Landy, F. J., and Farr, J. L. "Performance Rating." Psychological Bulletin 87 (January 1980): 72-107.

Latham, G. T.; Wexley, K. N.; and Pursell, E. D. "Training Managers to Minimize Rating Errors in the Observation of Behavior." Journal of Applied Psychology 60, 5 (1975): 550-555.

Latham, G. P., and Wexley, K. N. "Behavioral Observation Scales for Performance Appraisal Purposes." Personnel Psychology 30 (1977): 255-268.

Leedy, P. D. Practical Research: Planning and Design. New York: Macmillan Publishers, 1974.

Levin, T., and Long, R. Effective Instruction. Alexandria, VA: Association for Supervision and Curriculum Development, 1981.

Mai-Dalton, R. R.; Feldman-Summers, S.; and Mitchell, T. R. "Effect of Employee Gender and Behavioral Style on the Evaluations of Male and Female Banking Executives." Journal of Applied Psychology 64 (1979): 221-226.

Maier, N. R. F. The Appraisal Interview: Three Basic Approaches. LaJolla, CA: University Associates, 1976.

Manatt, R. P. Evaluating Teacher Performance. Part I and Part II, 60 minute videotapes. Alexandria, VA: Association for Supervision and Curriculum Development, 1981.

Manatt, R. P., and Blackmer, D. The School Improvement Model Consortium. Ames, IA: School Improvement Model Project, Iowa State University, December 1980.

Manatt, R. P.; Palmer, K.; and Hidlebaugh, E. "Evaluating Teacher Performance with Improved Rating Scales." NAASP Bulletin 60, 401 (1976): 21-24.

Marston, P. T. and Borchardt, L. M. Review of Literature and Research Bias in Teacher Observation Research and Development Center for Teacher Education, The University of Texas at Austin, R & D Report No. 5079, September 1979.

Martin, J. Mastering Instruction. Newton, Mass.: Allyn and Bacon, 1983.

McNeil, J. D., and Popham, W. J. "Administrator and Peer Ratings." G. D. Borich. The Appraisal of Teaching: Concepts and Processes. Reading, MA: Addison-Wesley, 1977, p. 268.

Medley, D. M.; Coker, H.; Coker, J. G.; Lorentz, J. R.; Soar, R. S.; and Spaulding, R. L. "Assessing Teacher Performance from Observed Competency Indicators Defined by Classroom Teachers." Journal of Educational Research 74 (1981): 197-216.

Medley, D. M.; Coker, H.; and Soar, R. S. Measurement Based Evaluation of Teacher Performance. An Empirical Approach. New York, NY: Longman, Inc., 1984.

Meyer, H. H. Performance Review Discussion: Making It Constructive. Paper presented at the Annual Convention of the American Psychological Association, Washington, D.C., August, 1986.

Millman, J. Handbook of Teacher Evaluation. Beverly Hills, CA: SAGE Publications, 1981.

Mitaskos, C. L. Mirrors for the Classroom--A Guide to Observation Techniques for Teachers and Supervisors. Andover, MA: Andover Public School System, 1980.

Mobley, W. "Supervisor and Employee Race and Sex Effects on Performance Appraisals: A Field Study of Adverse Impact and Generalizability." Academy of Management Journal 25, 3 (1982): 598-606.

Mullins, C. J., and Force, R. C. "Rater Accuracy as a Generalized Ability." Journal of Applied Psychology 46 (1962): 191-193.

Multicultural Workshop Video Tape Series. Gregory Coffin, Director. Boston, Mass.: Urban Schools Collaborative, Northeastern University, 1980.

National Commission on Excellence in Education. A Nation at Risk: The Imperative for Education and Reform. Washington, D.C.: U.S. Government Printing Office, 1983.

Natriello, G., and Dornbusch, S. M. "Pitfalls in the Evaluation of Teachers by Principals." Administrator's Notebook 29 (1980): 1-4.

Olson, L. "Performance-Based Pay Systems for Teachers Are Being Re-examined." Education Week 6, 29 (1987): 1-13.

Perrot, E. Effective Teaching: A Practical Guide to Improving Your Teaching. New York, NY: Longman Group Limited, 1982.

Peters, D. L., and McCormick, E. J. "Comparative Reliability of Numerically Anchored Versus Job-Task Anchored Rating Scales." Journal of Applied Psychology 50 (1966): 92-96.

Peterson, P. and Walberg, H. Research on Teaching, Concepts, Findings and Implications. Berkeley, CA: McCutchan Publishing, 1979.

Pursell, E. D.; Dorsett, D. L.; and Latham, G. P. "Obtaining Valid Predictors by Minimizing Rating Errors in the Criterion." Personnel Psychology 33 (1980): 91-96.

Reavis, C. A. and Shine, S. C. "A Study of the Effects of Prefatory Remarks on Teacher Evaluation." Educational Research Service, U.S. Department of H.E.W, Washington, D.C., 1977.

Redfern, G. Successful Teacher Evaluation. Santa Barbara, CA: Rex Fleming Productions, 1973.

Retzer, M. J. "The Analysis of Bias in Teacher Evaluation." Ph.D. dissertation, University of Pittsburgh, 1980.

Rice, B. "Performance Review: The Job Nobody Likes." Psychology Today, 19 (September 1985): 30-36.

Robinson, J. P., and Shaver, R. Measures of Social Psychological Attitudes. Ann Arbor, MI: Institute for Social Research, 1978.

Romberg, T. A. SMPL: The Systematic Monitoring of the Program of Learning. Berea, OH: Berea City School District, 1981.

Rose, G. L. "Sex Effects on Effort Attributions in Managerial Performance Evaluation." Organizational Behavior and Human Performance 21 (1978): 367-378.

Rosenshine, B. "Direct Instruction." In Instructional Leadership Handbook. Edited by J. W. Keefe and J. M. Jenkins. Reston, VA: National Association of Secondary School Principals, 1984.

Rosenthal, R., and Fode, K. "The Effects of Experimenter Bias on the Performance of the Albino Rat." Behavioral Science 19, (1963): 15-18.

Rucker, D. W. "A Study of Principal's Teaching Style Preference and Teachers' Teaching Style as a Source of Bias in Teacher Evaluation." Ph.D. dissertation, Iowa State University, Ames, IA, 1981.

Sapone, C. V. "Appraisal and Evaluation Systems: Perceptions of Administrators, Teachers." NASSP Bulletin 65 (1981): 25-30.

Schmidt, F. L., and Johnson, R. H. "Effect of Race on Peer Ratings in an Industrial Setting." Journal of Applied Psychology 62, (1973): 261-264.

Scriven, M. "The Evaluation of Teachers and Teaching." In Borich, G. D. (Ed.), The Appraisal of Teaching: Concepts and Processes. Reading, MA: Addison-Wesley, 1977.

Scriven, M. "Evaluation Ideologies." In Madaus, G. F; Scriven, M.; and Stufflebeam, D. L. (Ed.), Evaluation Model: Viewpoints in Educational and Human Services Evaluation. Hingham, MA: Kliuver Academic Publishers, 1983.

Sharon, A. T. "Eliminating Bias from Student Ratings of College Instructors." Journal of Applied Psychology 54 (1970): 278-281.

Sharon, A. T., and Bartlett, C. J. "Effect of Instructional Conditions in Producing Leniency on Two Types of Rating Scales." Personnel Psychology 22 (1969): 251-263.

Sherif, M. "A Study of Some Social Factors In Perception." Archives of Psychology 27 (1935): 187.

Simon A. and Boyer, G. Mirrors for Behavior: An Anthology of Classroom Observation Instruments. Philadelphia, PA: Research for Better Schools, Inc., 1970.

Simon, W. "Expectancy Effects in Scoring of Vocal Items: A Study of Scorer Bias." Journal of Educational Measurements 6 (1969): 183-189.

Skindrud, K. D. "An Evaluation of Observer Bias in Experimental - Field Studies of Social Interaction." Ph.D. dissertation, University of Oregon, 1972.

Smith, B.; Peterson, D.; and Micceri, T. "Evaluation and Professional Improvement Aspects of the Florida Performance Measurement System." Educational Leadership 44, 7 (1987): 16-18.

Soar, R. S. and Others. "Teacher Evaluation: A Critique of Currently Used Methods." Phi Delta Kappan 65 (1983): 239-246.

Spool, M. D. "Training Programs for Observers of Behavior: A Review." Personnel Psychology 31 (1978): 853-888.

Stallings, J. Madeline Hunter's Model: A Study of Implementation and Student Effects. NIE Grant No. R-81-0024. Presented to the Annual Convention of the American Educational Research Association, New Orleans, LA, April 1984.

Stow, S. "Using Effectiveness Research in Teacher Evaluation." Educational Leadership 37, 7 (1979): 55-58.

Stow, S. and Sweeney, J. "Developing A Teacher Performance Evaluation System." Educational Leadership 38, 7 (1981): 538-541.

Sweeney, J., and Stow S. "Performance Improvement: A People Program." Education 101, 3 (1981): 267-269.

Tesser, A. and Rosen, S. Bias in Interpersonal Communication. ERIC ED 119 253, 1975.

Thomas, M. D. Performance Evaluation of Educational Personnel. Bloomington, IN: Delta Kappa Foundation, 1979.

Thornton, G. C., and Zorich, S. "Training to Improve Observer Accuracy." Journal of Applied Psychology 65 (1980): 351-354.

Tuckman, B., and Others. Teacher Behavior is in the Eye of the Beholder: The Perceptions of Principals. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY, 4-8 April 1977.

Tversky, A., and Kahneman, D. "Judgement Under Uncertainty: Heuristics and Biases." Science 185 (1974): 1124-1131.

Valentine, J. "Applying the Steps in the Teaching Process." Missouri Schools 51, 1 (September 1985): 18-20.

Valentine, J. "Instructional Strategy." In Instructional Leadership Handbook, Sec. 3. Edited by J. W. Keefe and J. M. Jenkins. Reston, VA: National Association of Secondary School Principals, 1984.

Wahlberg, H. J. "What Works in a Nation Still at Risk." Educational Leadership 44, 1 (September 1986): 7-10.

Weitzul, J. B. "Can You Trust First Impressions". Life Association News 78, 9 (September 1983): 187-190.

Wells, W. D. and Smith, G. "Four Semantic Rating Scales Compared." Journal of Applied Psychology 44 (1960): 393-397.

Wexley, K. N. and Pulakos, E. D. "Sex Effects on Performance Ratings in Manager-Subordinate Dyads: A Field Study." Journal of Applied Psychology 67 (1980): 433-439.

Wexley, K. N. and Pulakos, E. D. "The Effects of Perceptual Congruence and Sex on Subordinates' Performance Appraisals of Their Managers." Academy of Management Journal 26, 4 (1983): 666-676.

Wherry, R. J. The Control of Bias in Rating: Survey of the Literature (Personnel Research Board Report 922). Washington, D.C.: Department of the Army, Personnel Research Section, [1952].

Wiersma, W., and Gibney, T. "Observation as an Approach to Measuring Teacher Competency." The Journal of the Association of Teacher Educators 7 (Spring-Summer 1985): 59-67.

Wise, J. E.; Nordberg, R. B.; and Rietz, D. J. Methods of Research in Education. Boston, MA: D. C. Heath and Co., 1967.

Wolf, P. R. Implementation of the Hunter Instructional Model: A Staff Development Study. Paper Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 1984.

Zavala, A. "Development of the Forced-Choice Rating Scale Technique." Psychological Bulletin 63 (1965): 117-124.

Zedeck, S.; Imparato, N.; Krausz, M.; and Oleno, T. "Development of Behaviorally Anchored Rating Scales as a Function of Organizational Level." Journal of Applied Psychology 59 (1974): 249-252.

Zunin, L. and Zunin, N. Contact: The First Four Minutes. New York: Ballantine Books, 1987.

ACKNOWLEDGEMENTS

Many thanks are due to those whose advice, expertise, assistance, and support made this study possible. A special thanks is expressed to my major professor, Dr. Jim Sweeney and his never ending encouragement and guidance.

I am also grateful to Liberty Public Schools in allowing me the opportunity to field test this study and Dr. Richard Manatt for the final data collection.

Appreciation is also extended to committee members Dr. Mary Huba, Dr. Fred Borgen, and Dr. Ross Engel. Their initial insights and support helped shape this study. Thanks also to Dr. Karl Koenig for his special assistance with analysis of the study data.

Special thanks are extended to my many colleagues. Their support of this pursuit provided the much needed encouragement to help me through these last few months. Thanks also goes to my typist, Penny Owens, whose technical assistance was greatly appreciated.

I would like to acknowledge the love and support of my family, especially my husband, Steve. With his caring and concern, this project was seen to completion. He always knew the right words to say.

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.

# APPENDIX A. INITIAL INSTRUMENT

GROUP A B S.S.#_____

INITIAL TEACHER PERFORMANCE RATING SCALE

DIRECTIONS: After viewing the videotape, please circle the number beside each indicator showing what you believe to be the most appropriate level of performance for that indicator.

1 = Must Improve
2 = Needs Improvement
3 = Meets Standard
4 = Exemplary

| INDICATORS | LOW | | | HIGH |
|---|---|---|---|---|
| * 1. Demonstrates effective personal organizational skills. | 1 | 2 | 3 | 4 |
| * 2. Organizes students for effective instruction. | 1 | 2 | 3 | 4 |
| * 3. Provides the structure for learning. | 1 | 2 | 3 | 4 |
| 4. Relates current lesson to previous learning. | 1 | 2 | 3 | 4 |
| 5. Provides focus for new learning. | 1 | 2 | 3 | 4 |
| 6. Involves students in learning new objective. | 1 | 2 | 3 | 4 |
| 7. States instructional objective. | 1 | 2 | 3 | 4 |
| 8. Provides a clear explanation of new material. | 1 | 2 | 3 | 4 |
| 9. Provides clear directions. | 1 | 2 | 3 | 4 |
| 10. Incorporates effective questioning techniques. | 1 | 2 | 3 | 4 |
| 11. Uses demonstrations, examples, and anecdotes to teach the lesson. | 1 | 2 | 3 | 4 |
| 12. Checks for student understanding. | 1 | 2 | 3 | 4 |
| 13. Paces lesson appropriately and/or adjusts as needed. | 1 | 2 | 3 | 4 |
| 14. Gives supportive and immediate feedback to students. | 1 | 2 | 3 | 4 |
| 15. Provides opportunities for student participation. | 1 | 2 | 3 | 4 |
| *16. Models effective communication skills. | 1 | 2 | 3 | 4 |
| 17. Displays a thorough knowledge of subject matter. | 1 | 2 | 3 | 4 |
| *18. Incorporates techniques to motivate students. | 1 | 2 | 3 | 4 |
| *19. Ensures student time on task. | 1 | 2 | 3 | 4 |
| 20. Maintains a high standard for student behavior. | 1 | 2 | 3 | 4 |
| 21. Demonstrates sensitivity in relating to students. | 1 | 2 | 3 | 4 |
| 22. Involves students in summary of lesson. | 1 | 2 | 3 | 4 |
| OVERALL RATING | | | | |
| Rate overall performance using the following responses: 1 = Must Improve 3 = Meets Standard 2 = Needs Improvement 4 = Exemplary | 1 | 2 | 3 | 4 |

*Indicators not used in data analysis.

Explanation for Teacher Performance Rating Scale

| | |
|---|---|
| Must Improve | Performance jeopardizes continued employment in the district. |
| Needs Improvement | Performance is below the district expectations. |
| Meets Standard | Performance meets the expectations set by the district. |
| Exemplary | Performance exceeds district expectations. |

# INFORMATION/DIRECTION SHEET

The purpose of this activity is to determine the accuracy of evaluators' ratings of teacher observations. Because teacher evaluation is mandated by nearly every state in the nation, it has become a vital component in improving instruction in the classroom. However, teacher evaluation is often biased by the evaluators' experiences and training in this area. It is the purpose of this activity to examine bias in teacher observation and to determine if and how it affects evaluator ratings of teacher performance on a given criterion.

Directions:

After receiving explanation, you will:

1. View a videotape.
2. Rate the performance of the teacher on "Effectiveness of Instruction" following the directions you receive. You should be using the format provided you.
3. Complete the "Evaluator Data" sheet.
4. Return all materials to the workshop coordinator.

Evaluator's I.D. #

EVALUATOR DATA

Please circle appropriate response.

1. Age of Evaluator.

   20 - 29  30 - 39  40 - 49  50 - 59  60+

2. Gender.

   Male  Female

3. Race.

   White  Black  Hispanic  American Indian  Other

4. Current level of education.

   Bachelor's Degree  Master's Degree  Master's + 30

   Specialist's Degree  Ph.D or Ed.D

5. Number of years teaching experience.

   Less than 1  1 - 5  6 - 10  11 - 15  16 - 20  20+

6. Number of years experience in educational administration.

   Less than 1  1 - 5  6 - 10  11 - 15  16 - 20  20+

7. Number of hours spent in teacher evaluation training. (Includes workshops, staff development, coursework, etc.)

   Less than 10  10 - 20  21 - 30  30+

8. Number of years experience in each of the following. (Includes both teaching and administration)

   Grades K - 6:  0 - 5  6 - 10  11 - 15  16 - 20  20+

   Grades 7 - 12:  0 - 5  6 - 10  11 - 15  16 - 20  20+

9. Number of years experience evaluating teachers.

   Less than 1  1 - 5  6 - 10  11 - 15  16 - 20  20+

## APPENDIX B. STATEMENTS OF CONSENT

TO: Diana Bourisaw
Research Investigator

FROM: Grundy Newton, Principal
Liberty Junior High School

RE: Permission to Videotape

DATE: May 13, 1987

The Liberty Public Schools supports the advancement of education through human subjects research. Therefore you have permission to videotape two lessons in Juarenne Hester's ninth grade language arts classroom during the month of May to further the purpose of your dissertation research.

This consent is given on the condition that the teacher and parents of the students involved also consent to the videotaping.

I realize after the tapes have been used for research purposes, they will be catalogued in the Liberty Public School Professional Library. These tapes will be used to train Liberty teachers in classroom observation skills therefore it is not anticipated that they will be erased.

Sincerely,

Grundy Newton

TO: Juarene Hester, Teacher
Liberty Public Schools

FROM: Diana Bourisaw
Research Investigator

RE: Production Release

DATE: May 13, 1987

You and your class will be involved in an Iowa State University research project. Harold McGuire, audio visual specialist for Liberty Public Schools, will be videotaping two lessons in your classroom. These videotapes will be used to conduct a research project on bias in teacher observation. These tapes will be viewed by Educational Administration students, administrators or supervisors of teachers.

After the tapes have been used for research purposes, they will be catalogued in the Liberty Public School Professional Library. These tapes will be used to train Liberty teachers in classroom observation skills. Therefore I do not anticipate that they will be erased at anytime.

I agree to the following:

1. I consent to my appearance in this production.

2. The producer and research investigator are released from any liability for claims by me or anyone else arising from my participation or appearance in this production.

3. My appearance or participation in this production confers upon them or me no ownership rights whatsoever.

_____

Signature of Teacher

CONSENT STATEMENT

During the month of May two lessons will be filmed in your child's language arts class. These videotapes will be used in a study on teacher evaluation. The taping will not significantly distract from your child's regular language arts instruction. Confidentiality will be guarded, first names only will be used during the lesson. The study will not focus on your child's performance, but the performance of the teacher.

The Department of Professional Studies in the College of Education at Iowa State University and Liberty Public Schools supports the practice of human subjects participating in research. This information has been provided so that you can decide whether you wish to allow your child to participate in this study. You should be aware that your child's participation is strictly voluntary; that is, your child is not required to participate. However, by allowing your child to participate you will help provide important information for the advancement of teaching.

After the tapes have been used for research purposes, they will be catalogued in the Liberty Public School Professional Library. These tapes will be used to train Liberty teachers in classroom observation skills. Therefore, I do not anticipate that they will be erased at anytime.

Do not hesitate to ask any questions about the study. You may contact me at the telephone number listed below.

Thank you for your cooperation.

Sincerely,

Diana Bourisaw
Principal Investigator
School phone: 781-4540

_____
Student's Name

_____
Parent's Signature

## APPENDIX C. REVISED INSTRUMENTS

PERSONAL DATA CARD

_____

Name:_____ _____ Female _____ Male

Mailing Address:

_____

_____

_____

_____

Teaching Major?_____

Years Experience as teacher evaluator?_____

Number of days teacher evaluation training?_____

Explanation for Teacher Performance Rating Scale

| | |
|---|---|
| Exemplary | Performance serves as a model for other teachers. The teacher demonstrates a high proficiency in this area. |
| Very Effective | Performance is high quality. The teacher is above average in this area but not good enough to serve as a model for others. |
| Effective | Performance is acceptable. Teacher demonstrates adequate skill in this area. |
| Ineffective | Performance is not at an acceptable level. Teacher has enough deficiences in this area to be ineffective. |
| Very Ineffective | Performance is highly unacceptable. Teacher does virtually nothing in this area that is of value to the lesson. |

GROUP A B S.S.#_____

TEACHER PERFORMANCE RATING SCALE (REVISED)

DIRECTIONS: After viewing the videotape, please circle the number beside each indicator showing what you believe to be the most appropriate level of performance for that indicator.

1 = Very Ineffective 3 = Effective 5 = Exemplary
2 = Ineffective 4 = Very Effective

| INDICATORS | LOW | | | | HIGH |
|---|---|---|---|---|---|
| 1. Demonstrates effective personal organizational skills. | 1 | 2 | 3 | 4 | 5 |
| 2. Organizes students for effective instruction. | 1 | 2 | 3 | 4 | 5 |
| 3. Provides the structure for learning. | 1 | 2 | 3 | 4 | 5 |
| 4. Relates current lesson to previous learning. | 1 | 2 | 3 | 4 | 5 |
| 5. Provides focus for new learning. | 1 | 2 | 3 | 4 | 5 |
| 6. Involves students in learning new objective. | 1 | 2 | 3 | 4 | 5 |
| 7. States instructional objective. | 1 | 2 | 3 | 4 | 5 |
| 8. Provides a clear explanation of new material. | 1 | 2 | 3 | 4 | 5 |
| 9. Provides clear directions. | 1 | 2 | 3 | 4 | 5 |
| 10. Incorporates effective questioning techniques. | 1 | 2 | 3 | 4 | 5 |
| 11. Uses demonstrations, examples, and anecdotes to teach the lesson. | 1 | 2 | 3 | 4 | 5 |
| 12. Checks for student understanding. | 1 | 2 | 3 | 4 | 5 |
| 13. Paces lesson appropriately and/or adjusts as needed. | 1 | 2 | 3 | 4 | 5 |
| 14. Gives supportive and immediate feedback to students. | 1 | 2 | 3 | 4 | 5 |
| 15. Provides opportunities for student participation. | 1 | 2 | 3 | 4 | 5 |
| 16. Models effective communication skills. | 1 | 2 | 3 | 4 | 5 |
| 17. Displays a thorough knowledge of subject matter. | 1 | 2 | 3 | 4 | 5 |
| 18. Incorporates techniques to motivate students. | 1 | 2 | 3 | 4 | 5 |
| 19. Ensures student time on task. | 1 | 2 | 3 | 4 | 5 |
| 20. Maintains a high standard for student behavior. | 1 | 2 | 3 | 4 | 5 |
| 21. Demonstrates sensitivity in relating to students. | 1 | 2 | 3 | 4 | 5 |
| 22. Involves students in summary of lesson. | 1 | 2 | 3 | 4 | 5 |

OVERALL RATING

Rate overall performance using the following responses:

1 = Very Ineffective 3 = Effective 5 = Exemplary
2 = Ineffective 4 = Very Effective 1 2 3 4 5

# APPENDIX D. TRAINING FORMAT

# INSTRUCTIONAL PLAN

Title Improving and Evaluating Teacher Performance Page # 1

Group or School School Leaders (Kansas) of 3

Date(s) August 17-19 (M-W), 1987

Attending Superintendents, Central Office Administrators, Principals

Presenting Consultant(s)

Dick Manatt

Associated with:
Richard P. Manatt, Educational Consultant
2926 Monroe Drive, Ames, IA 50010

| TIME | TOPIC | PRESENTER | MODE | VISUALS | HANDOUTS | REMARKS |
|---|---|---|---|---|---|---|
| Day One | | | | | | |
| 9:00 | Overview of Performance Appraisal | Manatt | LGI | O/H | Key Questions | -- |
| 10:15 | --Break-- | OYO | -- | -- | -- | -- |
| 10:30 | Pretest | Manatt | IS | -- | Skills of the Effective Evaluator | -- |
| 12:00 | --Lunch-- | OYO | -- | -- | -- | -- |
| 1:00 | Classroom Observation Training (Benchmark) | Manatt | SGI | Video Bob Johnson | Research on Teaching | Group 2 |
| | | | | | Observation Handbook | Group 1 |
| 2:00 | --Break-- | OYO | -- | -- | -- | -- |
| 2:15 | The Teacher Performance Evaluation Components | Manatt | LGI | O/H | Workbook Mod 1 | -- |
| 3:00 | Questions and Answers Plus Review | Manatt | Q & A | -- | -- | -- |
| 3:30 | Dismissal | | | | | |

# INSTRUCTIONAL PLAN

Title Improving and Evaluating Teacher Performance Page # 2

Group or School School Leaders (Kansas) of 3

Date(s) August 17019 (M-W), 1987

Attending Superintendents, Central Office Administrators, Principals

Presenting Consultant(s)

Dick Manatt

Associated with:
Richard P. Manatt, Educational Consultant
2926 Monroe Drive, Ames, IA 50010

| TIME | TOPIC | PRESENTER | MODE | VISUALS | HANDOUTS | REMARKS |
|---|---|---|---|---|---|---|
| Day Two | | | | | | |
| 9:00 | Classroom Observation Training (Drop By) | Manatt | SGI | O/H/Video Gerry Page 1 | Formative Scan Form Timeline | Group 1 |
| 10:00 | --Break-- | OYO | -- | -- | -- | -- |
| 10:15 | Analyzing Lesson Design | Manatt | LGI | O/H | -- | Artifacts Stu. Data |
| 12:00 | --Lunch-- | OYO | -- | -- | -- | -- |
| 1:00 | Classroom Observation Training (Announced Visit) | Manatt | SGI | O/H/Video Gerry Page 2 | Formative Scan Form | Group 1 |
| 2:00 | --Break-- | OYO | -- | -- | -- | -- |
| 2:10 | Summative Evaluation Report | Manatt | IS | -- | Summative Evaluation Scan Form | Both Groups |
| 3:00 | Questions and Answers Plus Review | Manatt | Q & A | Video: Drop By | Rating Scale | -- |
| 3:30 | Dismissal | | | | | |

# INSTRUCTIONAL PLAN

Title Improving and Evaluating Teacher Performance Page # 3

Group or School School Leaders (Kansas) of 3

Date(s) August 17-19 (M-W), 1987

Attending Superintendents, Central Office Administrators, Principals

Presenting Consultant(s)

Dick Manatt

Associated with:
Richard P. Manatt, Educational Consultant
2926 Monroe Drive, Ames, IA 50010

| TIME | TOPIC | PRESENTER | MODE | VISUALS | HANDOUTS | REMARKS |
|---|---|---|---|---|---|---|
| Day Three | | | | | | |
| 9:00 | Research on Effective Teaching | Manatt | IS | O/H | Teacher's Job Description | Research on Teaching and Observation Handbook |
| 10:00 | --Break-- | OYO | -- | -- | -- | -- |
| 10:15 | Supervising the Marginal Teacher | Manatt | LGI | O/H | Mod 1 SMT | -- |
| 11:00 | Due Process Supervision | Manatt | LGI | O/H | Mod 2 SMT | -- |
| 12:00 | --Lunch-- | OYO | -- | -- | -- | -- |
| 1:00 | SMT In Action | Manatt | LGI | O/H | Mod 3 SMT | -- |
| 2:00 | --Break-- | OYO | -- | -- | -- | -- |
| 2:10 | Activities of the Dismissal Team | Manatt | LGI | O/H/Video | Mod 4 SMT | -- |
| 3:00 | Winning and Workshop Evaluation | Manatt | LGI | O/H | Legal Aspects | -- |
| 3:30 | Dismissal | | | | | |