

1982

# An analysis of the effectiveness of training teacher evaluators in specific steps in the processes of clinical supervision and teacher performance evaluation

Dorothy A. Faast  
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>

 Part of the [Educational Administration and Supervision Commons](#)

## Recommended Citation

Faast, Dorothy A., "An analysis of the effectiveness of training teacher evaluators in specific steps in the processes of clinical supervision and teacher performance evaluation " (1982). *Retrospective Theses and Dissertations*. 7536.  
<https://lib.dr.iastate.edu/rtd/7536>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

## INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.
2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

**University  
Microfilms  
International**

300 N. Zeeb Road  
Ann Arbor, MI 48106



8224328

**Faast, Dorothy A.**

**AN ANALYSIS OF THE EFFECTIVENESS OF TRAINING TEACHER  
EVALUATORS IN SPECIFIC STEPS IN THE PROCESSES OF CLINICAL  
SUPERVISION AND TEACHER PERFORMANCE EVALUATION**

*Iowa State University*

PH.D. 1982

**University  
Microfilms  
International** 300 N. Zeeb Road, Ann Arbor, MI 48106



**An analysis of the effectiveness of training teacher evaluators  
in specific steps in the processes of clinical supervision  
and teacher performance evaluation**

by

**Dorothy A. Faast**

**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**

**1982**

## TABLE OF CONTENTS

	Page
CHAPTER I. STATEMENT OF THE PROBLEM	1
Introduction	1
Statement of the Problem	2
The Hypotheses	3
Definitions	3
Delimitations	4
CHAPTER II. REVIEW OF LITERATURE	6
Teacher Effectiveness	6
Clinical Supervision and Teacher Performance Evaluation	8
Evaluator Training	11
Supervision and Evaluation Skills	14
Learning Style	17
Interrater Reliability and Validity	20
Summary	21
CHAPTER III. METHODS	24
Research Design	25
Training Sessions	26
Subjects	27
Materials Development	28
The Instruments	29
Statistical Treatment	35
CHAPTER IV. FINDINGS	37
Introduction	37

	Page
Preconference (Lesson Plan Analysis)	38
Classroom Observation	41
Supervisory Conference (Simulation)	45
Conference Climate (On-the-Job)	47
Perceptions of Supervision/Evaluation (On-the-Job)	48
Variability	52
Learning Style	57
<b>CHAPTER V. SUMMARY, CONCLUSIONS, LIMITATIONS, DISCUSSION, RECOMMENDATIONS</b>	<b>59</b>
Summary	59
Conclusions	61
Limitations	61
Discussion	63
Recommendations for Practice	70
Recommendations for Further Research	73
<b>BIBLIOGRAPHY</b>	<b>75</b>
<b>APPENDIX A: WORKSHOP PLANNERS</b>	<b>81</b>
<b>APPENDIX B: TRAINING OBJECTIVES</b>	<b>90</b>
<b>APPENDIX C: INSTRUMENTS</b>	<b>92</b>
<b>APPENDIX D: CORRELATION COEFFICIENTS OF CONFERENCE SIMULATION TRIAD</b>	<b>106</b>



## LIST OF TABLES

	Page
Table 1. Comparisons of evaluators' lesson plan analysis, using paired t-tests (N=50). . . . .	39
Table 2. Comparison of evaluators' lesson plan analysis and expert panel, using separate t-tests (N=evaluators - 50, experts - 8) . . . . .	40
Table 3. Frequencies of right and wrong responses before and after training regarding a teacher's use of anticipatory set in a lesson (N=55). . . . .	43
Table 4. Frequencies of right and wrong responses before and after training regarding a teacher's statement of objectives in a lesson (N=55). . . . .	43
Table 5. Frequencies of right and wrong responses before and after training on all seven steps in the teaching episode. . . . .	44
Table 6. Comparisons of observers' conference skills, using paired t-tests (N=21). . . . .	46
Table 7. Comparisons of teachers' perceptions of conference climate, using paired t-tests (N=23 triads). . . . .	49
Table 8. Comparisons of teachers' perceptions of supervision/evaluation, using paired t-tests (N=18). . . . .	51
Table 9. Changes in evaluators' ratings of teachers' 1981 and previous years, using the unrevised of the Des Moines evaluation instrument (N=17) . . . . .	53
Table 10. Comparison of evaluators' performance evaluation ratings, using t-tests (N=50). . . . .	54
Table 11. Comparison of evaluators' performance evaluation rating and panel of experts, using t-tests (N=50). . . . .	56
Table D.1. Pearson product-moment correlation coefficients: observer (O) with teacher (T), observer (O) with supervisor (S) and supervisor (S) with teacher (T) before and after training. . . . .	107

## CHAPTER I. STATEMENT OF THE PROBLEM

## Introduction

The demands for improved leadership and more responsibility in education today come from all segments of society. Administrators as well as students and teachers are being held accountable. Such change apparently cannot be made rapidly and in a revolutionary manner. Instead, Sergiovanni and Starrett indicate that "people in schools and on school committees are ready for some quiet and effective improvements" (63).

Teacher evaluation is an integral part of the educational system's attempt to be accountable and as such is a primary responsibility of administrators. While the approaches to the evaluation process are many and diverse (29), clinical supervision (CS) is a specific supervisory approach which may respond to today's educational needs. It has been characterized as "a structure by which instructional adequacy can be established" (41).

If clinical supervision is capable of serving as a method of educational improvement, then an effective way of training administrators in the process must be found. The research evidence indicates that the most powerful device for improving appraisals is rater training. Training has been shown to increase agreement among raters, reduce bias, increase rating accuracy, prevent inflation of ratings, and spread out the rating distribution (45). Training is needed in the importance of careful appraisals, the desirability of a wide range of rating scores, the

influence of recent events and the minimization of common rating errors such as central tendency, halo effect, and leniency (27). Lack of adequate training has been cited as the major problem of most appraisal systems (46).

The purpose of this study is to carefully scrutinize what effects training in CS has on evaluation as well as to pose questions for future investigations.

#### Statement of the Problem

Evaluating teacher performance is the responsibility of principals, assistant principals, and other supervisors in the public school setting. Because this task is both important and time consuming, evaluators need to be trained in the components necessary to make them both effective and efficient. This study will attempt to determine if evaluators trained in the components of CS and a more elaborate system which includes CS, teacher performance evaluation (TPE) (47), will have greater success as measured by variability in ratings of teachers on performance standards prescribed by a district. However, less variability is preferable among raters of the same teacher. The research will also measure teachers' perceptions of the effectiveness of supervision and evaluation before and after training, and a test of learning style will be given to evaluators to assess its effect on the effectiveness of the training.

### The Hypotheses

The present study will investigate the effectiveness of teacher evaluation before and after training evaluators in specific steps in clinical supervision and teacher performance evaluation. Effectiveness of the training will also be examined by measuring teachers' perceptions of evaluation. Learning style of evaluators will be examined as well to determine if it influences the effectiveness of training. The study can be more specifically defined by the following operational hypotheses:

1. Trained evaluators will analyze lesson plans more effectively.
2. Trained evaluators will capture data during the classroom observation more effectively.
3. Trained evaluators will be more effective in skills used in the supervisory conference.
4. Trained evaluators will be perceived by teachers as effective in the supervisory conference.
5. Trained evaluators will be perceived as effective by teachers.
6. Trained evaluators will rate teachers with greater variability.
7. The learning style of the evaluator will effect the effectiveness of the training.

### Definitions

Evaluator--the school official designated to evaluate, appraise, guide or direct the work of teachers.

Rater--that school official who evaluates performance on an instrument specified for that purpose.

Clinical Supervision (CS)--Clinical Supervision consists of five stages:

1) preobservation conference, 2) observation, 3) analysis and strategy, 4) supervision conference (postconference), 5) postconference analysis.

Teacher Performance Evaluation (TPE)--Teacher Performance Evaluation

has nine steps: 1) establish the rules of the game, 2) orient teachers, 3) analyze lesson plans, 4) preobservation conference, 5) observation(s), 6) postobservation conference, 7) synthesize data, 8) summative evaluation report, 9) set job improvement targets.

#### Delimitations

The subjects in this study were evaluators in the Des Moines Independent Schools involved in various professional improvement programs, i.e., workshops, conferences, or college classes during 1981-1982 and, therefore, may demonstrate a more positive disposition toward research.

The evaluator trainees were: elementary principals, secondary principals, assistant principals, central office supervisors and consultants.

All training modules in the investigation were conducted during the 1981-1982 school year. All except one day was release time paid for by the district. The typical trainee received approximately 20 hours of direct instruction and approximately 15 hours of independent practice. Trainees also had the option of six additional hours of training. The Des Moines School facilities were used for all required and volunteer workshops on evaluation.

The training of evaluators was in preobservation conferencing, classroom observations, analysis and strategy, and supervisory conferencing in the CS cycle. Training of evaluators was in analyzing lesson

plans, preobservation conferencing, classroom observations, postobservation conferencing, synthesizing data, writing a summative report, and setting job improvement targets in the TPE cycle.

## CHAPTER II. REVIEW OF LITERATURE

## Teacher Effectiveness

The ingredients of good teaching--resulting in heightened student achievement--have been pinpointed by many good research studies and validated in the classroom.

Teachers who create a "neutral" climate by balancing student self-direction and teacher-directed activities, find student achievement goes up (66). Those teachers who work to produce a healthy attitude toward school and self, who establish a rapport with students, and who know many techniques and cognitive skills experience a rise in student gains (50). Other studies found that time on task, high expectations, teaching to an objective that is at the correct level of difficulty, and teacher "withitness" (a term coined by Armstrong and Kounin at the University of Texas Research and Development Center) all promoted increased student achievement (19, 32, and 28). A University of Philadelphia study showed that the teacher/student link is also important and the time spent with students was shown to be directly proportional to learning (28).

Much of the knowledge on effective teaching has been summarized by Barak Rosenshine. In general, Rosenshine's work has delineated an approach often referred to as "direct instruction", which he defines as follows:

Direct instruction refers to academically focused, teacher-directed classrooms using sequenced and structured materials.

It refers to teaching activities where goals are clearly set for students, time allocated for instruction is sufficient and continuous, coverage of content is extensive, the performance of students is monitored, questions are at a low cognitive level so that students can produce many correct responses, and feedback to students is immediate and academically oriented. In direct instruction, the teacher controls instructional goals, chooses materials appropriate for the student's ability. Interaction is characterized as structured but not authoritarian. (59)

Rosenshine's review of hundreds of research studies concluded that the direct instruction approach is more likely than other approaches to produce gains in student learning.

Throughout the years of a child's school life, the teacher is the point of contact between him/her and the entire educational enterprise. Personnel costs to districts represent so large a share of the cost of education that we must realize that the best hope for improvement in cost effectiveness lies in improving the effectiveness of the teacher. One very important way to improve the effectiveness of teachers is by improving the way they are evaluated.

Everybody from John Goodlad, Dean of the UCLA Graduate School of Education, to newspaper columnists is telling principals that they are the key to improved classroom instruction and teacher competency. James Sweeney of Iowa State University found, in summarizing the literature on school effectiveness, strong evidence that school effectiveness is enhanced by principals who emphasize instruction (67). Principals feel the importance of this area as well. In two separate studies, 83 percent of a sampling of secondary principals said their most important task was working with teachers on instruction, and 86 percent of elementary principals similarly declared that their primary responsibility was



supervision and instructional improvement (28).

### Clinical Supervision and Teacher Performance Evaluation

The heart of the supervisory process is teacher evaluation. If principals are expected to properly evaluate teacher effectiveness, then they must have knowledge of a process. Clinical supervision (CS) is a supervisory process that may meet this need. The CS process has been around for two decades and yet too few principals in the United States have been trained in it or have even heard of it.

There is a theoretical base for presuming that evaluators trained in CS make a difference in teacher effectiveness. Goldhammer, Cogan, and Anderson (25) assert that clinical supervision and its five stages, the preconference, classroom observation, analysis and strategy, supervisory conference, and postconference analysis, are essential to changing teacher behavior. Boyan and Copeland found that evaluators trained in the model were able to make significant improvements in a variety of teaching behaviors (12). Data pointed to validation of the model (70) and indications that the model's tenets and processes are compatible with the desires of teachers and administrators as well (17, 1).

In the clinical supervision setting, changes in the teachers' classroom behavior occurred in directions designated as "desirable" (23, 35, 40). There was also evidence of teacher growth in self-confidence and self-direction (64). However, research related to the effectiveness of CS is sparse and that which does exist reflects a lack of rigor.

A variation of the CS process called Teacher Performance Evaluation

(TPE) was developed in 1977 by Richard Manatt of Iowa State University (47). While CS puts great stress on teacher behavior in the future, TPE stresses the here and now. The TPE cycle consists of nine steps:

1. Establish the rules of the game
2. Orient teachers
3. Analyze lesson plans
4. Conduct preobservation conference
5. Conduct classroom observation(s)
6. Conduct postobservation conference
7. Synthesize data
8. Write evaluation report
9. Set job improvement targets.

TPE is defined as judging the goodness of teaching (47). Although the TPE cycle is longer, it allows for individual and cyclical differences to enable the evaluator to use his/her time most effectively and efficiently. TPE is a tough-minded, quality assurance mechanism, a process performed by evaluators that compares one teacher to another and to the school organization's standards.

Madeline Hunter, principal of the University Elementary School at UCLA, has developed a model which is very useful in training teacher evaluators. The essential contribution of Hunter's work is teaching principals to identify seven components in the teaching process which are drawn from the psychology of learning. These seven components are: 1) anticipatory set, 2) statement of the objectives, 3) input, 4) modeling, 5) check for understanding, 6) guided practice, 7) independent

practice (32).

Hunter and her associates have trained hundreds of principals throughout North America to teach teachers how to use these components more effectively (33). Her research has had a major influence on Richard Manatt and his research team at Iowa State University as they developed the School Improvement Model (SIM).

The School Improvement Model Consortium is a group of five K-12 school organizations and Iowa State University's College of Education combining efforts in a massive research project to improve teacher and administrator performance with the ultimate goal of improved student learning. Financed in part by the Northwest Area Foundation (the balance of the funds coming from consortium organizations), the SIM project is a four-year, field-based effort centered in the public school districts of Edina, Minneapolis and Northfield, Minnesota; Spirit Lake, Iowa; and Breck, an independent school located in Minneapolis.

The consortium makes four important linkages. The linkages are: 1) teacher performance evaluation is described, appraised and related to student learning, 2) administrators' behavior, relationships to each other, to teachers and to students is described, measured, appraised and related to teacher performance, 3) functional classroom curriculum as well as testing techniques is matched to the goals and aims of the school "community", 4) interventions, in the form of training, changes in instructional strategy and improvement of leadership will be created and provided for each school community in amounts and ways judged necessary from the findings of link one, two and three in the particular

school organization (48).

Barak Rosenshine has created a very useful concept called "direct instruction", after many years of sifting through thousands of teacher-effect studies. Direct instruction refers to an instructional sequence comprised of demonstration-practice-feedback. His approach was used as the instructional sequence plan for evaluator training in the present investigation (59).

The question that remains is how can the potential of CS and TPE be realized. Training evaluators in the processes and measuring their effectiveness may be one way to gather information to help answer this question among others.

#### Evaluator Training

In a 1978 survey of 1,600 secondary principals by NASSP, the category most often mentioned as the one principals felt they were not handling very well was teacher competency. A 1978 poll by NASSP revealed that 53 percent of 2,500 principals felt their primary need in inservice education was to develop teacher evaluation skills (28).

Kowalski reports that many administrators feel inadequately prepared in their educational training programs and too long and too far removed from their own classroom experiences to do suitable teacher evaluation (39). George Redfern says most principals aren't hired on the basis of their potential as instructional leaders (57). John Goodlad agrees. "He/she hasn't been hired with that role in mind. He/she doesn't have the authority, leadership or opportunities for the sustained inservice

that almost all principals require" (28).

Administrator groups and state and local inservice experts, meeting at a Washington, D.C. workshop sponsored by the federal government, concurred that administrator training should be focused on the principal. Such training must have the full support of the superintendents and school boards. "What principals really need is training that will equip them with the intellectual and human relations skills necessary to manage improvement efforts in their schools" (22).

Training is defined as the organized procedure by which people learn knowledge and/or skills for a definite purpose, such as to aid in the achievement of organizational goals (4). Training is known to make the following contributions (4):

1. Reduce learning time to reach acceptable performance
2. Improve performance on the present job
3. Formulate attitude
4. Aid in solving operational problems
5. Fill manpower needs
6. Benefit employees.

Sergiovanni and Starrett explain training as a process of accumulating a series of programmed behaviors which can be applied with reliability to a series of highly predictable situations (63). But supervision is largely a varied, situational, and unpredictable discipline. It is difficult to generate universal laws and principles of procedure for supervision, and for this reason the supervisory setting is more suitable to education than training. Evaluators can be educated in a workshop

setting. Latham et al. reported that only an intensive educational workshop was effective in training raters. They prescribed sessions giving subjects a chance to practice observing and rating videotaped performances with immediate feedback regarding the accuracy of their ratings. After six months, trained raters functioned as error-free raters (42, 43).

One effort at a broader approach, at the inservice level, has been developed by Karolyn Snyder (65). Snyder had adapted Benjamin Bloom's conclusions on quality instruction (6) as a framework for training principals and supervisors in clinical supervision. As one of the programs within Administrators-For-Change-Training (ACT), this approach is designed to introduce supervisors to the nature of Bloom's four variables through observation sequences that involve looking at videotapes of actual instruction to discover the nature and use of each of the four variables which are: drive (motivation), cue (stimulation), response (participation), and reward (reinforcement). Videotapes were used in conjunction with the training in the present investigation. Training for this investigation also made use of lectures, handouts, and various simulation activities.

Schmuck, Runkel, Arends and Arends concur that long-lasting individual change can occur when a person internalizes new concepts and is encouraged to behave in new ways (61). Cognitive and normative change can be stimulated by means of information presented in the form of lectures or written handouts. Behavioral and structural change arises from experiential learning.

Since Kurt Lewin's research during the 1930s and 1940s showed the importance of active group participation in helping people to learn new skills and attitudes, much social-psychological research has strongly supported simulation and experiential learning.

What would the payoffs be of improved evaluation?

1. A higher level of teacher performance
2. Improved educational services for students
3. A more accurate record of the status of performance of staff members
4. A greater commitment to the importance of being "evaluation minded" in the performance of one's duties and responsibilities.

Robert Goldhammer states that it's clear that training and inservice programs in the 1980s will have to be even more elaborate and intense than those he called for in the 1960s. "We need a much larger vision of what the supervisor's role should be. Proper inservice programs for administrators can help alleviate problems in evaluation" (28).

#### Supervision and Evaluation Skills

Robert Katz has identified three basic skills upon which he believes successful supervision rests--technical, human, and conceptual (34). Technical skills assume ability to use knowledge, methods, and techniques to perform specific tasks. The mechanics associated with lesson plan analysis, data gathering, conferencing, and filling out an evaluation instrument might be examples of technical skills which can be taught in training.

Human skill refers to one's ability and judgment in working with people. This skill requires self-understanding as well as consideration for others. Its knowledge base includes understanding and facility for, instructional leadership, motivation, attitudinal development, and the development of human resources.

Conceptual skill refers to the supervisor's ability to view the school, the district, and the educational program as a whole.

### Lesson plan analysis

It seems sensible to expect that improved planning for instruction will improve the quality of teaching that ensues. What is remarkable is that there is so little research evidence to support such a sensible notion.  
(53)

The need to establish and instruct from predetermined objectives is well-supported in educational circles (55, 68, and 71); however, very little evidence can be found to support the notion that analysis of instructional planning by the principal, makes a difference in the evaluation of effective teaching. Frudden did find, however, that there were more accurate ratings of teachers by evaluators who had training in lesson plan analysis than those with no training (21).

### Gathering observational data

Evaluation experts agree that a prime need is for inservice training for principals in how to conduct classroom observations (28). Erline Minton has stated that evaluators trained in observation techniques can assess teaching without having a knowledge of the subject matter being taught (52).



Data may be gathered in the classroom by organizing around categories of pupil and teacher behavior (54), by notes recording classroom events verbatim (24), or by systematic observation using any combination of established techniques (20).

It is, therefore, imperative that we train evaluators in the skills necessary to effectively gather data in the classroom and investigate its worth with regard to summarizing teacher behavior for evaluation purposes.

### Conferencing

Training in conference technology can be helpful to educational supervisors. How the evaluator or supervisor conducts the culminating conference depends on knowledge of conference skills. This is as important as how the supervisory function is performed throughout the evaluation cycle.

A major evaluation problem is lack of time spent in conferencing. The typical principal spends less than one-half hour per teacher per year in end-of-the-year postevaluation conferences (47). Training in the skills necessary to conduct a conference could motivate principals to conference with teachers and thus evaluate them more effectively.

In such training, especially in role-playing situations, there is the omnipresent danger that despite changing characters the play may vary too little (25). In training evaluators, every effort must be made to vary the situations to discourage the same type of behavior occurring in each simulated activity.

## Learning Style

"Learning style consists of distinctive behaviors which serve as indicators of how a person learns from and adapts to his environment. It gives clues as to how a person's mind operates (26)."

Anthony Gregorc has done extensive work in the analysis of learning/teaching styles and the potent forces behind them. His study of environment tells us that: ". . . every environment places demands upon individuals for adaptation."

Applied to training, this means that when a trainer selects a method of presentation, such as a lecture, he/she is placing certain and limited adaptation demands on the minds of the trainees. Gregorc would say that the most successful trainees in a workshop just happen to have adaptive abilities that match hidden demands being placed on them by the training method.

Practitioners, researchers, and writers have been focusing their attention on the realized and potential benefits of aligning learning styles and teaching styles. As a result of their efforts, there are many materials out on the market today available to educators interested in the learning/teaching style alignment process.

Gregorc's research revealed four distinct learning patterns and teaching techniques associated with these learning patterns. These are as follows (26, pp. 20-22):

### Concrete Sequential (CS)

The concrete sequential learning preference is characterized by the propensity to derive information through direct hands-on experience. CS learners exhibit extraordinary development of

their five senses. They appreciate order and logical sequence of the if-then, premise-conclusion variety. They like touchable, concrete materials. In a biology class, a plaster model handled by the teacher would be insufficient for these learners. They want to have the real thing to take apart themselves.

#### Teaching techniques (CS)

1. Workbooks
2. Lectures accompanied with overhead transparencies, drawings, models, demonstration
3. Hands-on materials
4. Field trips.

#### Concrete Random (CR)

The concrete random learning preference is characterized by an experimental attitude and accompanying behavior. CR learners get the gist of ideas quickly and demonstrate the ability to make intuitive leaps in exploring unstructured problem-solving experiences. Sometimes they also have insights and make leaps in structured situations. Then they are chided for not showing their steps or for jumping to conclusions.

Concrete random learners utilize the trial-and-error approach in acquiring information. They do not like cut-and-dried procedures that deny them opportunities to find answers in their own ways. They do not respond well to teacher intervention in their dependent efforts. They work well independently or in small groups.

#### Teaching techniques (CR)

1. Games or simulations
2. Independent study
3. Brief mini-lectures
4. Problem solving activities.

#### Abstract Sequential (AS)

The abstract sequential learning preference is characterized by excellent decoding abilities with written, verbal, and image symbols. AS learners have a wealth of conceptual "pictures" in their minds against which they match what they read, hear, or see in graphic and pictorial form. They possess and like to use reading, listening, and visual translation skills. A symbol or picture is worth a thousand words to them.

These learners prefer a presentation that has substance, is

rational, and is sequential in nature. They are able to extract main ideas from a logical presentation. They learn well from authorities and like vicarious experiences.

#### Teaching techniques (AS)

1. Audio tapes
2. Extensive reading assignments
3. Slides
4. Lectures.

#### Abstract Random (AR)

Abstract random learners are distinguishable by their attention to human behavior and a capacity to sense and interpret "vibrations." They are attuned to nuances of atmosphere and mood. They associate the medium with the message and tie a speaker's manner, delivery, and personality to the message being conveyed. In doing so, they evaluate a learning experience as a whole.

Abstract random learners prefer to receive information in an unstructured manner and therefore like group discussions, activities which involve multi-sensory experiences, and busy environments.

They prefer freedom from rules and guidelines. They seem to gather information and delay reaction; they organize material through reflection to get what they want.

Additional studies have shown that other dualities such as deductive/inductive processing are manifest in behaviors like test taking question-and-answer patterns, self-directedness, and various forms of group work.

#### Teaching techniques (AR)

1. Movies
2. Television
3. Lecture with discussion of material presented
4. Short reading assignments.

The education profession has talked about learning style differences and the need for variety in curricula, instruction, and goals, but there is no clear-cut dedication to them as yet.

### Interrater Reliability and Validity

When you apply the reliability or the validity standard to the available literature on teacher evaluation, "mountains of documents melt away" (44).

Howsam, in reviewing research up to 1960, found little consistency in supervisor ratings of teachers. He concluded that ratings in general were biased and subjective. He found a good deal of early evidence which reflected little interrater reliability in supervisory ratings (30).

Research on performance appraisal shows that the closer the factors are to actual behavior and results, the more the raters will agree in their evaluations. "Getting raters to observe work-related behaviors more systematically and representatively is a potentially fruitful approach" (11).

Raters are seldom skilled in making systematic work-related behavior observations. They need to become adept at observing and recording relevant job behaviors so they may be better equipped with the information necessary for making accurate evaluations of teacher performance. Criteria on the evaluation instrument should be those that can be clearly described so that all raters will have the same kind of behavior in mind (3, 51). If raters can first be trained to observe work-related behaviors more competently, and second to use scales more accurately, it is possible that more error-free portrayals of performance can be made (11).

Every major writer in the field of teacher evaluation supports some kind of training for the evaluator. Dale Bolton has cited untrained administrators as one source of low reliability in evaluating teachers.

He contends that training can increase reliability (8).

Much work has been done by researchers in examining the effects of training on reducing rater errors (5, 8). Latham, Wexley, and Pursell, in reviewing literature on rater error in industry, listed many of the same concerns as mentioned by critics of teacher evaluation (43).

Latham's et al. work demonstrated that training reduced rater error and this reduction held up over time.

Kirchner and Reisberg found that trained supervisors were more discriminating in their rating and showed more "spread" and variation in their ratings. "Since a basic objective of rating is to differentiate performance, the trained supervisors are doing the more effective job" (38).

#### Summary

The burden of evidence regarding the skills of the teacher evaluator suggests that each appraiser should be able to identify effective teaching behaviors, be proficient at lesson plan analysis, gather descriptive data in the classroom observation, conduct an evaluation conference with coaching/feedback, fill out an evaluation report and accurately rate a teacher on a prescribed form.

A task analysis of these several skills suggests that the content of a teacher evaluation training program should include the following cognitive data: Madeline Hunter's (32) seven steps in the teaching episode and Barak Rosenshine's (59) approach referred to as "direct instruction".

Madeline Hunter's seven steps include: 1) anticipatory set, 2) statement of the objectives, 3) input, 4) modeling, 5) check for understanding, 6) guided practice, 7) independent practice.

Barak Rosenshine's "direct instruction" approach includes academically focused, teacher-directed classrooms using sequenced and structured materials. Activities are used in which goals are clear to students, sufficient time is allocated for instruction, coverage of content is extensive, the performance of students is monitored, questions produce many correct responses, and feedback to students is immediate and academically oriented.

Furthermore, cognitive and normative change from training can be stimulated by means of information presented in the form of lecturettes or written handouts. Behavioral and structural change arises from experiential learning.

The literature indicates that the most difficult outcomes to obtain in a CS/TPE training program are: 1) lesson analysis centered on effective teaching behaviors, 2) accurate classroom data capture, 3) interrater reliability regarding the summative evaluation, 4) giving feedback on the seven steps in the teaching process in a way that enables a teacher to change his/her instructional behavior permanently, 5) and finally, building a positive relationship between evaluator and evaluatee when the whole process is so often seen as a routine, bureaucratic chore.

While findings and indications can be summarized regarding the constituent elements of CS and TPE, no general conclusions can be drawn

regarding the most effective training methodology. Leading authorities and empirical research indicate that too many directions are being taken without looking at common goals of understanding and utilizing the processes that lend themselves to advantageous results for teachers and students. Said another way, most of the training is opinion-based, not research-based in origin.

For the scant experiments extant, a methodological inadequacy is that most of the cited studies, which are specific to CS, have very small samples, e.g., five to ten subjects, and therefore, have limited generalizability. Perhaps the greatest barrier to doing meaningful research in this area has been the political and emotional sensitivity involved in performance appraisal--not unlike medical research, human lives are at stake (or at least livelihoods!). If the researcher is wrong, teachers' and/or principals' reputations may be irreparably harmed.

It appears that future investigations must also overcome problems in design which characterize much of the available work. Care should be taken to insure that clinical procedures are followed and that input variables are controlled to insure that participants in studies are unaware of expected results. Further, because of the possibility of the Hawthorne effect, any differences in results must be interpreted carefully. These factors prevent studies in this area of interest from becoming a potent force in helping to create better teaching and learning situations in classrooms today.



## CHAPTER III. METHODS

This investigation had seven purposes, five of which generated an operational hypothesis which in turn suggested an empirical hypothesis. Two hypotheses are such that they may influence the first five but do not suggest empirical tests. Purposes and hypotheses are as follows:

1. to determine the effectiveness of training evaluators in lesson plan analysis; trained evaluators will analyze lesson plans more effectively; there will be no significant difference between mean scores of the evaluators' lesson plan analysis before and after training.

2. to determine the effectiveness of trained evaluators for the classroom observation; trained evaluators will capture data during the classroom observation more effectively; there will be no significant difference in percentage of appropriate data gathered by the evaluators before and after training.

3. to determine the effectiveness of training evaluators in supervisory conference skills; trained evaluators will be more effective in the supervisory conferences; there will be no significant differences between evaluators' mean scores of conference skills before and after training.

4. to determine the effectiveness of training evaluators on teachers' perceptions of them in the supervisory conference; trained evaluators will be perceived by teachers as more effective in the supervisory conference; there will be no significant difference between mean scores of teachers' perceptions of evaluators in the supervisory conference

before and after the evaluators were trained.

5. to determine the effectiveness of training evaluators on teachers' perceptions of them as evaluators and supervisors; trained evaluators will be perceived by teachers as more effective evaluators and supervisors; there will be no significant difference between mean scores of teachers' perceptions of evaluators as evaluators and supervisors before and after they were trained.

6. to determine if evaluators' learning style effected the effectiveness of training; the learning style of evaluators will effect the effectiveness of the training; the relationship between learning style of evaluators and effectiveness of the training will be discussed.

7. to determine if the training of evaluators effected the variability in ratings of teachers on specified standards; trained evaluators will rate teachers with greater variability on specified standards; it will be postulated that there will be no overall variation from previous years in ratings of teachers on specified standards.

#### Research Design

The present investigation used some materials and methods generated by the School Improvement Model Consortium (48) as well as videotape training modules developed at Iowa State University for the Association for Supervision and Curriculum Development (49). In order to measure acquisition and application of performance evaluation skills in the present study, the investigation was conducted with a pretest-posttest design. The steps in this particular approach are as follows: Step

One - administration of a pretest measuring the dependent variable, Step Two - application of the experimental treatment (independent variable) to the subjects, Step Three - administration of a posttest measuring the dependent variable again. Differences due to application of the experimental treatment are then determined by comparing the pretest and posttest scores (9). In this investigation, evaluators in Des Moines Schools were pretested on various skills, given training in those skills, and subsequently posttested. The results were gathered to determine the effects of training. These data were then subjected to the appropriate tests of significance to determine if the differences were greater than might have occurred by chance.

#### Training Sessions

Training of evaluators was held in the Des Moines School District facilities. All training sessions were held during the 1981-1982 school year. The district released trainees from their regular assignments for all training sessions, with the exception of one day, a Saturday, for which overtime was paid by the district. The schedule of training for evaluators was as follows:

Subject	Date	Sample and Time
TPE State of the Art	August 10-11, 1981	$\frac{1}{2}$ evaluators, 6 hours each day
Data Gathering	September 23-24, 1981	$\frac{1}{2}$ evaluators, 5 hours each day
Conference Skills	October 7-8, 1981	$\frac{1}{2}$ evaluators, 6 hours each day
Lesson Plan Analysis and Objectives for Improvement	November 13, 1981 November 14, 1981	All evaluators, 4 $\frac{1}{2}$ hours All evaluators, 4 $\frac{1}{2}$ hours

Note: Complete workshop plans are contained in Appendix A.

### Subjects

The sample for this study was comprised of 125 subjects, all with some responsibility for evaluation in the Des Moines Independent Community School District. The persons sampled attended all training sessions and data were collected at these sessions.

Tests in the skill areas: lesson plan analysis, data gathering, and conference skills were administered before training sessions in these skills and posttested after the training. Teachers' perceptions of the quality of appraisals were measured before and after the training as well. Learning style was tested after training as was variability of evaluator ratings on district standards. Dates and subject areas of tests as well as subjects tested were as follows:

Pretest	Date	Posttest	Date	Subjects
Lesson Plan Analysis	September 23	Lesson Plan Analysis	November 14	55 evaluators
Data Gathering	August 10	Data Gathering	October 7	55 evaluators
Conference Skills	September 24	Conference Skills	December 3	21 evaluators
Perceptions: Conference Climate	October 1	Perceptions: Conference Climate	December 1	40 teachers
Perceptions: Supervision and Evaluation	November 1	Perceptions: Supervision and Evaluation	January 1	40 teachers
Learning Style		Learning Style	January 15, 1982	57 evaluators
Variability in Rating		Variability in Rating	December 15	23 evaluators

Note: All instruments are on file in E005 Quadrangle, Iowa State University, Ames, Iowa

## Materials Development

Training of teacher evaluators is a highly complex task. For purposes of this investigation, cognitive, behavioral, and structural change was examined. Cognitive change can be stimulated by means of information presented in lectures and written materials. Behavioral and structural change arises from experiential learning. Input from established experts was important for cognitive change in this investigation and generation of written materials to accompany lectures was a major task. A broad approach to instructional materials was taken in training sessions to provide for behavioral and structural change as well. For example, videotapes were used with simulation activities to provide opportunities for experiential learning. (All videotapes and accompanying materials were field tested with appropriate groups before use (49, 14, 69).) Materials developed for training sessions are designated by subject matter, author and date here:

Materials	Subject	Author	Date
Videotape "Tracing Our Origins"	State of the Art	Northeastern University	August 10-11
TPE Workbook Videotape "Evaluating Teacher Performance" Kit "Evaluating Teacher Performance"	State of the Art	ASCD Richard Manatt Dorothy Faast	August 10-11
Mirrors for the Classroom	Data Gathering	Charles Mitsakos	September 23-24
Conference Skills Workbook	Conference Skills	James Sweeney	October 7-8

## Materials developed for training sessions (Continued)

Materials	Subject	Author	Date
Videotape "Tracing Our Origins" Conference Simulation	Conference Skills	Northeastern University Dorothy Faast	October 7-8
Videotape "Evaluating Teacher Performance" Kit "Evaluating Teacher Performance"	Conference Skills	ASCD Richard Manatt Dorothy Faast	October 7-8
Videotape "Intermediate Language Arts" "Intermediate Language Arts" Lesson Plan	Lesson Plan Analysis	Georgia Teacher Assessment Project	November 13-14

Note: Objectives for training workshops are contained in Appendix B, and all materials used in workshops are on file in E005 Quadrangle, Iowa State University, Ames, Iowa

## The Instruments

The instruments used were diverse. These instruments were used: Teacher Instructional Plans and Materials Assessment Scale (Frudden), Observational Data Assessment (Faast), Conference Skills Assessment Scale (Sweeney, Spencer), Impact Message Inventory (Kiesler), Teachers' Perceptions of Supervision and Evaluation Survey (Young, Heischberger), Transaction Analysis Inventory (Gregorc), Des Moines Teacher Evaluation Instrument (unrevised form), School Improvement Model Prototype: Summative Evaluation Instrument (Manatt, Stow).

### Teacher Instructional Plans and Materials Assessment Scale

This instrument was developed by Sarah J. Frudden, Assistant Professor of Special Education, University of Northern Iowa, for the measurement of the competencies relating to the preparation of instruction, selection of objectives, and choosing materials and equipment. A five-point scale with a descriptive response mode makes up this seven-item instrument. A check above the appropriate phrase constitutes a rating. A panel of 40 persons: 20 classroom teachers, 12 administrators, and eight college faculty members provided ratings to be used as a measure determining rating proficiency. The suggested ratings from the panel of experts were used in the present study as the standard for accuracy by which rater proficiency was measured.

### Observational Data Assessment

An important part of the process of evaluation, and included in both CS and TPE, is the classroom observation. During the observation, data are gathered and serve as a record of methods, techniques, and behaviors noted during the lesson. These data are formative (descriptive) rather than summative (judgmental).

As a measure of proficiency in gathering formative observational data in the present investigation, each trainee in the sample viewed a videotaped instructional scene and identified teacher behaviors based on Madeline Hunter's Seven Steps in the Teaching Episode (33). A panel consisting of Madeline Hunter, Principal, UCLA Laboratory School; Richard Manatt, Director SIM, Iowa State University; Shirley Stow, Co-Director

SIM, Iowa State University; Sarah Frudden, Assistant Professor, Special Education, University of Northern Iowa; Charles Mitsakos, Associate Superintendent, Andover Public Schools, Massachusetts; James Huges, Educational Consultants, Inc., and Elaine Jarchow, Associate Professor, Secondary Education, Iowa State University viewed the same videotaped instructional scene and recorded behaviors based on Hunter's Seven Steps in the Teaching Episode. This panel of experts also indicated the appropriateness of the behavior. The observed behaviors from the panel of experts were used in this study as the standard for accuracy by which observer proficiency was measured.

#### Conference Skills Assessment Scale

Improving instruction is strongly dependent on the extent to which supervisors communicate effectively in the conference. To maximize conference effectiveness, evaluators in the sample were trained in these specific skills: opening (setting the climate), body (analyzing observed behavior), and closure (summary and goal setting).

The instrument used to measure these skills before and after training was the Conference Skills Assessment Scale, developed by Professors of the Educational Administration Section at Iowa State University, Ames, Iowa. The eleven items on the instrument are congruent with the literature on conferencing, thus they have construct validity (7). The Conference Skills Assessment Scale has a seven-point graphic response response mode plus NA (not applicable). High scores indicate success.

The instrument was used in conjunction with a simulation activity. Simulated information about a teacher was given to a trainee acting as



the supervisor to read before the conference. The trainee acting as the teacher received limited information regarding the situation. A conference was then staged and a third trainee acting as process observer rated the supervisor in his/her group on the Conference Skills Assessment Scale. After the conference, the teacher in the triad rated the supervisor on the instrument and the supervisor did a rating on him/herself. Mean scores before and after training were analyzed for significant change.

#### Impact Message Inventory

The Impact Message Inventory (IMI), developed by Keisler at Virginia Commonwealth University, Richmond, Virginia, measures the affective, behavioral and cognitive reactions an individual experiences as a consequence of a just-completed interaction with another person. In the present investigation, the IMI was used as a measure of teachers' perceptions of the climate in a supervisory conference. The instrument was administered to a random sample of teachers immediately following a conference. The instrument was mailed directly to the researcher by the teacher. The same teachers' perceptions were measured in this manner before and after evaluators were trained.

The inventory includes 15 subscales of interpersonal style. For this investigation, six subscales were identified by a panel of experts, all of whom are professors in the College of Education at Iowa State University, Ames, Iowa, as those interpersonal styles most appropriate to assessing the climate of the supervisory conference. Those chosen were: dominant, mistrusting, hostile, nurturant, affiliative, and

agreeable. Six descriptors made up each subscale for a total of 36 items on the revised instrument. Keisler, using split-half facotr analysis reported the following coefficients of reliability: .954, .755, and .772 (36). Analysis of results in this study is reported by subscale to assess the overall climate of the conference.

#### Survey of Teachers' Perceptions of Supervision and Evaluation

In this age of accountability, there is an obvious need for school supervision and evaluation. One of the reasons supervision is a neglected task is that principals are well-aware that teachers may become anxious and unhappy at the prospect of being supervised (72). The present investigation examines this dilemma.

An instrument developed by James Young and Robert Heischberger at New York State University, Fredonia, entitled Teachers' Perceptions of Supervision and Evaluation (TPSE) was given to a random sample of teachers in the present study to measure how they perceived supervision and evaluation in their setting. They evaluated their supervisors before and after the supervisors were trained. The instrument was mailed directly to the researcher upon completion. The survey consists of questions regarding the need for supervision, participation in evaluation, amount of time spent in the processes, and the type of supervisor-teacher relationship that exists. The questions were based on concerns related to effective supervision and evaluation as registered by educators throughout the United States in recent educational publications. In addition, the instrument offers the opportunity to teachers to suggest ways to improve instruction and the supervisory/evaluation process.

### Transaction Analysis Inventory

"Learning style consists of distinctive behaviors which serve as indicators of how a person learns from and adapts to his environment. It also gives clues as to how a person's mind operates," says Anthony Gregorc, author of the Transaction Analysis Inventory (TAI). This instrument was developed for use in determining the subject's learning style. The instrument solicits the cataloging of overt behavior and analysis of this cataloging determines its underlying cause. From this, certain inferences are drawn that tell us about the nature of the learner.

Gregorc's analysis of overt behavior indicated that some people's mind operate best in concrete situations, others in abstract, and some in both. Some individuals have an ordering preference that is sequential, while others demonstrate nonsequential patterned preferences. Some use both. Some people process best through deduction, while others use forms of induction. And, again, others use both. Some individuals function best on their own, while others are most productive through group activity. Some do equally well in either situation (26). The TAI was then developed to reveal learning style by measuring these behaviors. In the present study, the TAI was used to understand the learning styles most prevalent in the sample and thus discern its impact on effectiveness of training.

Des Moines Teacher Evaluation Instruction (unrevised form)

An important component in the present investigation was to examine discrimination in rating teachers on specific standards before and after evaluator training. In order to get a reading of variability in ratings on performance standards, those evaluators in the same Des Moines buildings for at least three years were asked to rate one teacher, whom they had rated before the training, on the instrument previously used in the district. The instrument was an unrevised form used by the district before revision in the Fall of 1981. It consisted of 11 items (criteria) and a five-point response mode plus not applicable (performance standards). This year's ratings of the teacher was then compared to ratings collected by the district since 1977 to detect and measure any variability changes.

## Statistical Treatment

Data processing was conducted at the Iowa State University Computation Center using the Statistical Package for the Social Sciences whereby five hypotheses were subjected to statistical treatment.

The first hypothetical question was submitted to examination by a paired t, wherein mean scores of lesson plan analysis before training were compared to mean scores of lesson plan analysis after training.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} - 2r \left(\frac{s_1}{n_1}\right) \left(\frac{s_2}{n_2}\right)}} \quad (56)$$

Hypotheses three, four, and five were submitted to examination by a paired t as well. That is, mean scores of conference skills, teachers' perceptions of conference climate, and teachers' perceptions of supervision and evaluation were compared before and after training of evaluators.

To test proficiency in capturing observational data, hypothesis two, the McNemar test for the differences between two correlated portions was used. The formula for calculating difference in values is:

$$P_1 - P_2 = \frac{a - d}{n} \quad (18)$$

and the formula for the z test statistics with correction for continuity is:

$$Z = \frac{1d - a1 - 1}{\sqrt{d + a}} \quad (18)$$

This procedure was used to determine if there was a significant difference between right answers before and after the training due to fact that the same subjects were involved at both times.

## CHAPTER IV. FINDINGS

## Introduction

The basic problem for this investigation was to examine the effectiveness of training teacher evaluators in specific steps in clinical supervision and teacher performance evaluation. To accomplish this goal, the researcher collected data from subjects responsible for teacher evaluation in the Des Moines Independent Community Schools at several training workshops and randomly sampled teachers with regard to their perceptions of supervision and evaluation skills of their supervisor/evaluator.

Because the series of training workshops described herein became the whole management staff development program for 1981-1982 in Des Moines, attendance at the sessions frequently totaled 125. Such large numbers were desirable for a common base of understanding and to demonstrate support for the first-line supervisor (e.g., principal, assistant principal) by the central office supervisors, directors, assistant superintendents, and the superintendent.

Nonetheless, the specific training pool for experimental purposes was delimited to 81 administrators who had the specific assignment for teacher performance evaluation in the district. Furthermore, the bulk of the workshops was provided in mirror-image days (i.e., two days, back to back, half of the evaluators each day) which had identical content. This had the advantage of keeping the buildings staffed with 50 percent of the administrators on any given day.

This procedure, coupled with normal absences and a lack of compulsory attendance, resulted in a not unexpected shrink when pre- and post-test subjects were included in the statistical analysis. A further problem arose with data which were captured linking workshop simulations to "real world" data captured from faculty members. Consequently, the number of subjects' data sets for most analyses obtained was in the range of 20-60 pairs.

The data gathered to test the hypotheses proposed have been arrayed in series following the specific steps in the processes of clinical supervision and Teacher Performance Evaluation. This order, in general, also follows the sequence of training workshops. In each case, descriptive statistics are provided to show the direction and magnitude of change in the subjects' behavior. Also the appropriate statistical procedure was applied and the results tabled for display. Instruments used in data collection can be found in Appendix C.

#### Preconference (Lesson Plan Analysis)

In an attempt to determine if the subjects improved in their ability to analyze a teacher's lesson plan, they were asked to rate lesson plan two from the ASCD kit (49) before and after training.

A paired t-test on subjects' mean ratings was reported item by item in Table 1. A separate t-test comparing the subjects to a panel of expert evaluators was reported item by item in Table 2.

To understand the mean responses found in these tables, it is necessary to recall that the graphic response mode on the instrument

Table 1. Comparisons of evaluators' lesson plan analysis, using paired t-tests (N=50)

Group	Item						
	Objectives 1	Content 2	Procedures 3	Cognitive levels 4	Student differences 5	Rates of learning 6	Evaluation 7
Before ( $\bar{X}$ )	3.42	2.70	2.86	2.62	2.12	0.80	2.04
After ( $\bar{X}$ )	2.22	2.48	1.92	1.86	1.42	0.73	1.80
t-value	7.10**	1.12	5.61**	4.08**	4.69**	0.84	1.34
Probability	0.00	0.27	0.00	0.00	0.00	0.41	0.19

\*\*  
p < .01, 2.668.



Table 2. Comparison of evaluators' lesson plan analysis and expert panel, using separate t-tests (N=evaluators - 50, experts - 8)

Category	Item						
	Objectives 1	Content 2	Procedures 3	Cognitive levels 4	Student differences 5	Rates of learning 6	Evaluation 7
Evaluators (before)	3.42	2.70	2.86	2.62	2.12	1.74	2.04
Experts	2.50	1.88	2.00	2.63	1.38	1.25	1.88
Evaluators (after)	2.22	2.48	1.92	1.86	1.42	1.60	1.80
Experts	2.50	1.88	2.00	2.63	1.38	1.25	1.88
t-value (before)	2.59*	2.17	3.66**	-0.01	2.56*	1.78	0.39
Probability	0.03	0.06	0.00	0.99	0.03	0.11	0.71
t-value (after)	-0.79	1.66	-0.35	-1.90	0.16	1.29	-0.18
Probability	0.45	0.14	0.73	0.09	0.88	0.23	0.86

\*  $p < .05$ , 2.010.

\*\*  $p < .01$ , 2.668.

was specifically tailored to the criteria. As an example, criterion one was, "Learning is stated in terms of what the student will be able to do after mastery of the objective(s)." The subject, hereafter referred to as trainee, was asked to study the lesson plan, judge its adequacy and respond with one of the five numbers which represented the extent to which the teacher met that criteria. Generally, 3s, 4s, and 5s indicated "adequate = 3" to "excellent = 5" planning, 1s and 2s were associated with criterion one.

Analysis of Table 2 shows the trainees, prior to training, clustered around three (adequate planning). It should be pointed out that trainees rated significantly lower on all items after the training, the mean response clustering around two.

The trainees, when pretested in lesson plan analysis, differed significantly from the panel of experts on several items. Those items were numbers one, three, and five. However, when posttested, after training, there was no significant difference in mean responses of trainees and experts on any items.

#### Classroom Observation

Trainees were instructed to use the topical data capture method in pre- and posttests to measure their proficiency in recording behaviors observed in the classroom observation. Madeline Hunter's Seven Steps in the Teaching Episode served as a rubric under which to gather and classify data relating to teacher behaviors. Trainees were asked before and after training to view a videotaped teaching episode and record data

regarding the teacher's use of two of the seven steps and the appropriateness of their use.

Data gathered by trainees before and after training were compared to data gathered by experts in teacher performance appraisal. The experts' data constituted "right" answers. A McNemar test of correlated proportions was used to determine differences between right answers given by trainees before and after training.

The two steps selected most often by trainees were anticipatory set and statement of objectives. Therefore, the data from these two steps were subjected to statistical analysis and are reported in Tables 3 and 4, while only frequencies of right and wrong answers are reported for the other five steps. These are reported in Table 5.

Analysis of Table 3, regarding the teacher's use of anticipatory set in the lesson, indicates that the proportion of correct answers on the pretest was 0.38. Comparing this to a proportion of 0.73 correct answers on the posttest, we see there was a higher percentage of success after the training.

Table 4 similarly indicates a higher percentage of success after the training. The percentage of correct answers on the pretest regarding statement of objectives was 0.31 as compared to 0.60 on the posttest.

Another method of data capture, the timeline, was used in training evaluators as well. Trainees used this method when rating teacher performance on a summative report. Variability in ratings of a teacher on specific standards using timeline data gathered are reported in Tables

Table 3. Frequencies of right and wrong responses before and after training regarding a teacher's use of anticipatory set in a lesson (N=55)

	After		Total
	Wrong	Right	
Before			
Right	9	8	17
Wrong	<u>13</u>	<u>25</u>	<u>38</u>
Total	22	33	55

$P_1 - P_2 \leq -0.29^*$

Table 4. Frequencies of right and wrong responses before and after training regarding a teacher's statement of objectives in a lesson (N=55)

	After		Total
	Wrong	Right	
Before			
Right	8	13	21
Wrong	<u>7</u>	<u>27</u>	<u>34</u>
Total	15	40	55

$P_1 - P_2 \leq -0.35^*$

Table 5. Frequencies of right and wrong responses before and after training on all seven steps in the teaching episode

	Right/Right	Wrong/Wrong	Right/Wrong	Wrong/Right
ANTICIPATORY SET (N=55)				
Pretest				
Posttest	8	13	9	25
STATEMENT OF OBJECTIVES (N=55)				
Pretest				
Posttest	13	7	8	27
INPUT (N=9)				
Pretest				
Posttest	4	2	1	2
MODELING (N=11)				
Pretest				
Posttest	7	1	3	0
CHECK FOR COMPREHENSION (N=13)				
Pretest				
Posttest	6	1	4	2
GUIDED PRACTICE (N=17)				
Pretest				
Posttest	3	10	2	2
INDEPENDENT PRACTICE (N=5)				
Pretest				
Posttest	4	1	0	0

10 and 11.

### Supervisory Conference (Simulation)

Conference simulations were used to determine trainees' improvement in supervisory conference skills. Simulation exercises, involving the use of triads in which trainees rotated playing the roles of teacher, supervisor and process observer, were used before and after training. A scenario called "Jim and Mary" provided information for the pretest conference and one called "Jan Lyon" provided information for the posttest.

The process observer, in the triad, was asked to rate the supervisor in the triad, before and after training, on an eleven-item instrument. The instrument had a semantic differential response mode numbering 1 through 7. Pre- and postmean responses are revealed in Table 6. In order to fully understand the tabled responses, it is important to know that ratings of 1, 2, and 3 denote a skilled supervisor, a degree of adequacy is denoted by 4s and 5s and 6s and 7s identify the supervisor as unskilled.

Criteria one and two related to the opening of the supervisory conference. Criteria two through eight dealt with the body and criteria nine through eleven relate to the close of the conference.

Analysis of Table 6 reveals that the average response from observers, prior to training, was nearest to three, indicating the supervisor was somewhat skilled in the supervisory conference. Observers' responses were clustered around 2 after training indicating supervisors were rated

Table 6. Comparisons of observers' conference skills, using paired t-tests (N=21)

Group	Item										
	1	2	3	4	5	6	7	8	9	10	11
Before ( $\bar{X}$ )	3.48	3.14	3.24	3.35	4.29	3.48	3.52	3.26	3.76	3.33	2.42
After ( $\bar{X}$ )	2.43	2.38	2.00	2.05	2.90	2.14	3.29	3.95	3.57	2.90	2.74
t-value	1.83	1.49	3.40**	2.87**	2.29*	3.25**	0.32	-0.88	0.26	0.81	-0.60
Probability	0.081	0.152	0.003	0.010	0.033	0.004	0.752	0.392	0.795	0.427	0.559

\*  $p < .05$ , 2.086.

\*\*  $p < .01$ , 2.845.

as more highly skilled than before training. There was significant change, however, on only four items, numbers 3, 4, 5 and 6, all in the section denoted as the body of the conference. This is desirable due to the fact that changing teacher behavior, the real goal of the training, will only occur as a result of what happens in the body of the conference.

In addition to these findings, an attempt was made to determine the amount of agreement among the teacher, supervisor and observer, in the triad, regarding the supervisor's skills before and after training. Following the simulated conference the teacher was asked to rate the supervisor and the supervisor was asked to rate him/herself. A Pearson Product-Moment correlation coefficient was computed for these data. The statistical results from this procedure are recorded in Appendix D. These results reveal low correlations among members of the triad before training, except in a few scattered incidents, and high correlations on many items after the training.

#### Conference Climate (On-the-Job)

Data analyzed in the previous section regarding proficiency in conference skills were obtained via workshop simulations. However, it was desirable to measure these skills in the "real world" as well. This was accomplished by asking for teachers' perceptions of the climate of a supervisory conference immediately following that interaction (a regular evaluation conference) with their supervisor/evaluator. The inventory used had six subscales pertaining to conference climate, they were



dominant, hostile, mistrusting, agreeable, nurturant, and affiliative. Descriptors within the subscales can be found in Appendix C. A four-point, Likert-type scale was used with responses ranging from 1--not at all to 4--very much so.

Teachers were asked to complete the inventory after a conference with their supervisor/evaluator before the supervisor/evaluator had training and again after training. Inventories were mailed directly to the researcher upon completion. Data were analyzed with a paired t and results of the statistical treatment appear in Table 7.

Information contained in Table 7 reveals that teachers' mean responses on the first three subscales, dominant, hostile and mistrusting, were generally lower after training. This indicated a decrease in those behaviors by the supervisor. Mean responses by teachers on the last three subscales, agreeable, nurturant and affiliative, were higher after the training, indicating these behaviors were perceived to a greater degree. However, the only significant differences, were found in subscales dominant and hostile (decreasing) and agreeable and nurturant (increasing).

#### Perceptions of Supervision/Evaluation (On-the-Job)

In another attempt to measure evaluators' supervision and evaluation effectiveness in actual practice, teachers were given a questionnaire which asked for their perceptions of supervision/evaluation in their school. Teachers were to answer yes, no or neutral to ten statements.

A paired t on yes responses only, yes being the desired response

Table 7. Comparisons of teachers' perceptions of conference climate, using paired t-tests  
(N=23 triads)

Group	Subscale					
	Dominant	Hostile	Mistrusting	Agreeable	Nurturant	Affiliative
Before ( $\bar{X}$ )	1.76	1.34	1.64	3.28	3.18	2.84
After ( $\bar{X}$ )	1.49	1.07	1.29	3.62	3.52	2.99
t-value	2.73	2.43*	2.01	-2.77*	-2.67*	-1.01
Probability	0.012	0.024	0.057	0.011	0.014	0.324

\*  $p < .05$ , 2.074.

on all items, revealed no significant difference in teachers' perceptions of supervision/evaluation in their setting before and after their supervisor/evaluator was trained in specific steps in CS and TPE. The results are displayed in Table 8.

By examining Table 8, it is evident that only 33 percent of the teachers sampled responded "yes" to the 10 statements on the questionnaire. This indicates some deficiency in supervision/evaluation.

In addition to these 10 items, teachers were asked to answer two open-ended questions.

1. Name one humanistic characteristic pertaining to a supervisor that would cause you to hire him/her.

Answers given most often were: objectivity, honesty, and understanding.

2. What would decrease the fear of being supervised?

Most responses indicated that improved communications, and teachers' knowledge of expectations would do most to decrease the fear of being supervised.

Two forced-choice questions were included as well.

1. What kind of relationship would you like with your supervisor?

Choices included: collegueship, helping, evaluator, counselor, teacher.

Sixty-seven percent of the teachers chose a helping relationship and thirty-three percent chose collegueship. No one chose any of the others.

2. What is the main role you now see your supervisor/evaluator playing?

Choices included: instructional leader, public relations, administrative

Table 8. Comparisons of teachers' perceptions of supervision/evaluation, using paired t-tests (N=18)

Group	Those answering yes to items 1-10 <sup>a</sup>
Before ( $\bar{X}$ )	6.78
After ( $\bar{X}$ )	6.94
t-value	-0.61
Probability	0.55

<sup>a</sup>1) There is a definite need for supervision and evaluation of teachers in the public school. 2) The supervisor is quite often seen as potentially dangerous to a teacher. 3) Teachers should take part in developing or selecting evaluation instruments, so that they know the criteria against which they are being judged. 4) Evaluation should be used to diagnose teachers' performance so they can strengthen their weaknesses through in-service education. 5) Evaluation should be something in which teachers have a part along with students, parents, and administrators. 6) One way out of the evaluation dilemma is to put the focus on the learner, not the teacher, and to involve everyone in the business of assessing or supervising everything we do all the time. 7) It is important for the supervisor to have some understanding of the teacher's educational philosophy and how the teacher views his own profession. 8) Instead of focusing major attention on providing the basis for dismissal or continued employment, evaluation programs should focus attention on improving the performance of the employee presently serving on the job. 9) The building principal should spend at least 35% of his/her time in supervising. 10) My building principal spends at least 25% of his/her time in supervising.

leader, passive leader.

Sixty-one percent of the teachers chose administrative leader, 28 percent chose instructional leader, eleven percent chose passive leader and no one chose public relations.

### Variability

In an attempt to examine variability in rating teachers on specific standards before and after training, evaluators, three years or more in the same buildings, were asked to rate a teacher on the unrevised form of the Des Moines evaluation instrument (1980). This instrument consisted of 11 items and a five-point response mode. Standards on this forced-choice response mode were: does not meet, operates below, exceeds, meets district's standard. These ratings were compared to ratings of the same teacher on his/her previous evaluation report.

Scrutiny of the two reports reveals the changes reflected in Table 9. Only the composite rating was considered, as this was the criterion for retaining or dismissing a teacher.

Table 9 reveals that no one rated a teacher below district standards, either before or after the training. And it appears that more teachers were rated above district standards after the training. Since these results were inconclusive with regard to variability, an additional attempt to determine change in variability in ratings on specific standards was made.

Trainees were asked to watch a videotape and to evaluate a teacher's performance before and after their training. Vignette two of the ASCD kit (49) was chosen for this test. Trainees were taught to capture classroom observational data via a timeline on a legal pad. They were instructed to assume that they had performed all of the necessary steps in formative teacher evaluation for a one-year cycle. Additionally, they

Table 9. Changes in evaluators' ratings of teachers' 1981 and previous years, using the unrevised of the Des Moines evaluation instrument (N=17)

	Does not meet	Operates below	Exceeds	Far exceeds	Meets
1981	0	0	6	0	11
Previous report	0	0	7	0	10
<u>Change</u>					
				From exceeds to meets	6
				From meets to exceeds	7
				Remained meets	4
				Remained exceeds	0

were directed to assume that everytime they obtained data, the teacher observed (Darlene Frazier teaching Advanced Art) performed at exactly that level. With those directives in mind, they were asked to complete the summative evaluation report contained in the ASCD kit. The pretest was held on August 10 before any training. The posttest was conducted on December 3.

The results of the pre- and posttest are outlined in Table 10. Note that the summative instrument does not provide a composite rating, instead each of the 10 criteria is treated separately.

To fully understand the mean responses tabled, it is necessary to

Table 10. Comparison of evaluators' performance evaluation ratings, using t-tests (N=50)

Group	Item									
	Communi- cation 1	Organ- ization 2	Content 3	Capabil- ities 4	Feed- back 5	Setting 6	Personal organ- ization 7	Organ- izes students 8	Student work habits 9	Student self- discipline 10
Before ( $\bar{X}$ )	2.00	2.75	1.85	1.75	1.81	2.42	2.19	2.33	2.00	2.33
After ( $\bar{X}$ )	2.04	2.00	1.58	1.92	1.77	2.65	2.42	2.52	2.25	2.79
t-value	-0.53	4.64**	1.38	-1.31	0.32	-1.42	-2.04*	-1.50	-1.70	-2.81**
Prob- ability	0.60	0.00	0.18	0.20	0.75	0.16	0.05	0.14	0.10	0.01

\*  $p < .05$ , 2.010.

\*\*  $p < .01$ , 2.668.

remember that the numbers represent a graphic response mode which has been specifically tailored to the performance criteria. For example, criteria number one was "teacher communicates effectively with students." The trainee was to judge the level of performance from observing the videotape and studying the timeline notes obtained and respond with one of five numbers which represented the extent to which the teacher met the standard of performance. Generally speaking, twos and threes were associated with substandard performance, a four was standard performance, and a one was used when the trainee did not observe that behavior.

Analysis of Table 10 reveals that average responses from trainees, prior to training, clustered around two (below standard performance). It should be pointed out that this low rating was desirable for two reasons. First, the performance really was poor! Second, there is a natural drift toward higher ratings when the evaluator returns to the actual appraisal in his/her building and overrates the teacher to avoid hostility in the supervisory conference.

Returning to the table, it should be noted that the subjects' responses changed significantly after training on only three items. The items were two (instruction organized around objectives), seven (personal organization), and ten (promotes self-discipline). Number two was rated lower after training. The other two items were rated higher.

More than interrater reliability and change in individual perception are involved in accurate teacher performance evaluation. To determine if trainees were evaluating teacher performance accurately, an expert jury of clinical supervisors/TPE Trainers was utilized. Table 11



Table 11. Comparison of evaluators' performance evaluation rating and panel of experts, using t-tests (N=50)

Category	Item									
	Communi- cation 1	Organ- ization 2	Content 3	Capabil- ities 4	Feed- back 5	Setting 6	Personal organ- ization 7	Organ- izes students 8	Student work habits 9	Student self- discipline 10
Evaluators (before)	2.00	2.76	1.92	1.76	1.80	2.40	2.20	2.34	2.00	2.34
Experts	3.13	2.63	1.63	2.00	3.25	2.13	3.25	2.88	1.75	1.50
Evaluators (after)	2.04	2.04	1.58	1.92	1.78	2.64	2.40	2.50	2.24	2.76
Experts	3.13	2.63	1.63	2.00	3.25	2.13	3.25	2.88	1.75	1.50
t-value (before)	-4.74**	0.48	0.65	-1.12	-7.42**	0.66	-3.95**	-1.02	0.94	2.84**
Prob- ability	0.00	0.64	0.53	0.29	0.00	0.53	0.003	0.34	0.37	0.02
t-value (after)	-4.72**	-2.47*	-0.10	-0.39	-7.66**	1.25	-3.27**	-0.72	1.84	4.34**
Prob- ability	0.00	0.03	0.92	0.705	0.00	0.25	0.01	0.50	0.10	0.00

\*  $p < .05$ , 2.010.

\*\*  $p < .01$ , 2.668.

compares their ratings of Darlene Frazier's performance with trainees judgments before and after training. Generally speaking, the Des Moines trainees gave lower ratings than the experts, both before and after training. The responses of interest are those obtained after training and they varied significantly from the experts on items one, two, five, seven, and ten. In all instances, except number ten, they rated the teacher lower than did the experts.

### Learning Style

Learning styles of trainees were measured using the Transaction Analysis Inventory. The name of this instrument (by Gregorc, 26) is designed to mask its real intent, that of measuring individual learning style. The word transaction refers to an individual's adaptation to his/her environment, not the interaction between people as in transactional analysis.

An analysis of the results of this inventory revealed that 60 percent of the trainees (N=60) were concrete sequential learners. Sixteen percent were abstract random, 16 percent were concrete random, and 8 percent were abstract sequential.

The concrete sequential learner is able to derive information best through hands-on experience. He/she appreciates order and logical sequence and likes touchable, concrete materials. Teaching techniques that work best are lectures, accompanied by overhead/transparencies, drawings, models, demonstrations. Workbooks and hands-on materials are the best supplements to the lecture.

It is apparent from recent surveys by Gregorc and Manatt (26, 47) that most school administrators are concrete sequential learners. The Des Moines evaluators do not vary significantly from the norm.

Many of the techniques used in training evaluators in Des Moines were those that work best with concrete sequential learners. Therefore, effectiveness training should not have been hampered by the learning style of the majority of the trainees. Furthermore, the lecture/demonstration approach appears to be a good technique to use with all learning styles except concrete random.

CHAPTER V. SUMMARY, CONCLUSIONS, LIMITATIONS,  
DISCUSSION, RECOMMENDATIONS

Summary

During the 1981-82 school year, the Research Institute for Studies in Education at Iowa State University, in conjunction with the Des Moines Independent Community School District, developed a teacher performance evaluation system to be implemented in the Des Moines Schools that same year. Development of such a system involved instrumentation, procedural design, and training of teacher evaluators as major tasks. This investigation measured the effect of training teacher evaluators in specific steps in clinical supervision and teacher performance evaluation.

The instruments, tests, and analysis indicated:

1. Evaluators rated teachers' lesson plans lower after training, indicating an effect of training. Evaluators differed significantly from a panel of experts in lesson plan analysis before the training; however, after training, there was no significant difference, on any items between evaluators and experts.
2. Evaluators had greater success in gathering data from a classroom observation after training, as indicated by higher frequencies of answers identified as "right" by a panel of experts.
3. Evaluators became more proficient in supervisory conference skills after training. There was more agreement among evaluators in identification of these skill levels as well.
4. Evaluators were perceived, by teachers, as less dominant and

hostile and more agreeable and nurturant in the supervisory conference, after training.

5. Supervision/evaluation in the school setting, as perceived by teachers, did not improve after evaluators had been trained. Teachers expressed the desire for more objectivity, honesty, and understanding from supervisors/evaluators and indicated that a helping and collegial relationship with their supervisor/evaluator would be best. Most teachers felt their evaluator spent a major portion of his/her time being an administrative leader.

6. Evaluators rated a teacher, in a videotape simulation, below standard on all items on a summative evaluation report, before and after training. Their ratings changed significantly after training on only three items, one of which decreased and two of which increased. When compared to a panel of experts, they rated lower on all items before and after training (save one item after the training). On-the-job, evaluators rated no teachers below standards, either before or after training and more teachers above standards after the training, when using the unrevised form. Using the new instrument this year, however, there was increased variability in ratings on specific standards.

7. Evaluators in the Des Moines Independent Community Schools are primarily concrete/sequential learners.

### Conclusions

Considering the data collected and analyses made in this investigation, the following conclusions are offered regarding the effectiveness of training teacher evaluators in specific steps in clinical supervision and teacher performance evaluation:

1. The training program, as constituted, is effective.
2. Trained evaluators analyze lesson plans more effectively.
3. Trained evaluators capture data during the classroom observation more effectively.
4. Trained evaluators are more effective in recognizing and using supervisory conference skills.
5. Trained evaluators are perceived by teachers as effective in the supervisory conference.
6. Trained evaluators are not perceived, by teachers, as effective supervisors.
7. Trained evaluators do not rate teachers with more variability, on-the-job, however, there is some rating variability in simulation activities.
8. The learning style of the evaluator does not effect the effectiveness of the training.

### Limitations

It must be noted that certain limitations were imposed due to the nature of the design of this study.

1. Due to lack of compulsory attendance at training workshops,

absenteeism was often high. This caused a breaking of pairs for t-test analysis of pre- and posttests, thus a shrink in the desired N.

2. Since training workshops were staff development programs for all administrators in the Des Moines Schools, many participants were not first-line supervisors. This may have caused these participants to be less interested in some of the subject matter covered by the trainers.

3. Training workshops were held in a variety of settings, some good, some bad, all in school facilities. It is possible that a hotel function room, university setting, or other neutral ground, would have been more desirable, physically and psychologically.

4. Instruments used in the investigation were nonstandardized. These offered no norms to consider when analyzing data.

5. Another problem relating to instrumentation resulted from the response mode on the performance rating scale used to measure variability of rating on specific standards. A "one" was used to designate "not observed". If it had been a "zero", it would not have had an empirical influence on the resulting rating scores. The instrument was used because it was a component of current classroom simulation materials, a group norm on the items had been established, and results of expert jury's ratings on the items were available.

6. Since pretest data solicited from teachers had to be collected by the time evaluators were trained, any data returned late could not be used. In some cases, there was a very limited amount of time for return, and lack of return reduced the N significantly.

7. Results of the posttest on conference climate may have been

skewed because that conference may have been summative in nature, involving judgments of teacher behaviors. In contrast, the pretest data may have resulted from a formative conference, involving description of behaviors and a lack of judgment.

8. The study contained only a limited analysis of performance evaluation reports in the district. This was due to the fact that they were official business records used this year by the Personnel Office for promotion, tenure, transfer, and, in some instances, dismissal.

9. Teacher performance evaluation instruments and methods developed in Des Moines, as a result of this project, became the official procedures in the district during the 1981-1982 school year. Perhaps a year of test and try would have enhanced results.

10. Because extensive use was made of simulation activities in this investigation, application on-the-job may not have been optimum. An assessment center approach may have been useful.

#### Discussion

The first area of examination in this investigation sought to determine if training teacher evaluators in lesson plan analysis would make them more effective in that task. Upon analysis of data, it was found that trained evaluators were more effective in lesson plan analysis. Not only were the evaluators rating teachers lower after training, which, in this case, was desirable, since the lesson plan was inadequate, but also evaluators did not differ significantly on any items, from experts in lesson plan analysis after training. The success of the preconference



in the evaluation process often depends upon the ability of the evaluator to analyze a teacher's plans. Increased proficiency in this task should help to insure more productive preconferences.

The second area of investigation dealt with analyzing evaluators' competence in gathering data from a classroom observation. Analysis of the data gathered by evaluators indicated that they could better identify use and appropriateness of some teaching techniques observed in a lesson after training. Evaluators selected two of Hunter's (33) seven steps in the teaching episode to describe in a videotaped instructional segment. They recorded use and appropriateness of the steps as they observed them. Most of the evaluators chose anticipatory set and statement of objectives.

Statistical analysis on these items showed a high proportion of wrong answers before and right answers after training. The training appeared to have made them more effective. An insufficient number of evaluators chose the other five steps, therefore no formal analysis was done. However, studying frequencies of right and wrong answers on the other five steps shows that more were right before and right after training in all cases. This does not indicate that training had any effect.

This investigation also attempted to determine the effectiveness of evaluators in a supervisory conference. Analysis of data, captured in simulation exercises, indicated that evaluators were more highly skilled after training in conferencing techniques associated with the body of the conference (i.e., coaching and counseling). There were no significant

differences, however, on items relating to the opening and closing of the conference. Since changes in teacher behaviors most often occur as a result of what happens in the body of the conference, these results are desirable. This is not to say the opening and closing are unimportant or that evaluators were unskilled in these areas. All scores, including those in the opening and closing, fell in the range designated as "somewhat skilled" to "skilled" after the training. However, the significant changes to "highly skilled" occurred only in the body. It seems appropriate, at this time, to note with satisfaction the correlation in ratings among evaluators playing the roles of supervisor, teacher, and observer in the conference simulations. It appears there is significant interrater reliability on many items as a result of training workshops on conferencing. This is indicated by more highly correlated responses on the posttest.

Actual conference climate as perceived by teachers was measured in the investigation as well. Evaluators were rated by teachers as less dominant and hostile and more agreeable and nurturant after training. These results tell us something about how successfully skills gained in training workshops were applied on-the-job. In this case, these data indicate a fair amount of success in transfer of conference skills. Evaluators were given a lot of practice in conferencing techniques, both as part of the regular workshops and in optional guided practice sessions. This may help account for the significant changes.

An analysis of data, regarding teachers' perceptions of supervision/evaluation, revealed no significant differences before and after their

supervisor/evaluator was trained. Items 1-10 on a questionnaire they were given dealt with general aspects of a school supervision and evaluation program and although important to effectiveness do not relate directly, in most cases, to administrator behavior resulting from training. Therefore, we should not view the results of the questionnaire as a direct reflection on training effectiveness. We should, however, concern ourselves with those answers, given by teachers, to the four questions accompanying the questionnaire. When asked what humanistic characteristic would cause them to hire a supervisor, they most often answered: objectivity, honesty, and understanding. This information can be considered in administrator selection processes and must be considered when planning and designing staff development programs, but also it must be viewed as an indication of teachers' needs. The words objectivity, honesty, and understanding imply a fairness teachers not only want but that they deserve. The question, "What would decrease the fear of being supervised?" produced these answers: improved communications and knowledge of expectations. It is apparent that administrators need training in communication skills and this can be offered by the district, but informing teachers of what is expected of them is the responsibility of each administrator in his/her own situation. Students expect this of teachers, and it is reasonable for teachers to want the same from their supervisor.

Teachers were asked what kind of relationship they wanted with their supervisor and their answers most often were, "a helping and collegial relationship." All of these answers tell us that teachers are

reaching out to administrators, asking for more communication and interaction. As supervisors/evaluators, administrators must particularly answer that call in order to be successful. The last statement asked teachers for a description of their supervisor/evaluator's main role. Most answered, administrative leader. They chose this answer most often over instructional leader, perhaps because the supervisor spends more time on office tasks than on people tasks. One can conclude from this that these teachers are not getting what they need from supervision/evaluation at this time.

Variability in rating teachers on specific standards is a desirable outcome of teacher performance evaluation training. Teachers constitute a heterogeneous group and it is not desirable to have homogeneous ratings of such a group. Variability in ratings offer discrimination power to the evaluation process. In reviewing data from past years' and this year's unrevised evaluation forms completed by principals, it was found that no teacher was rated below district standards and more teachers were rated above standards this year. This reveals very little variability in ratings and almost no discrimination.

In a discussion with Mr. Earl Bridgewater, Executive Director of Personnel in the Des Moines Schools, a different picture of variability in performance rating by evaluators emerged. Mr. Bridgewater pointed out several changes he perceived as he conferenced with principals regarding teachers they evaluated this year, using the new evaluation instrument. Unlike the unrevised instrument, used for purposes of this investigation, the new instrument has only three standards: meets, does

not meet and needs improvement.

Administrators have completed their formal teacher performance evaluations for the 1981-1982 school year. Positive outcomes of the recent training were obvious. The evaluation instruments revealed a more in-depth analysis of the classroom teacher performance, which included the writing of objectives for remediation as well as for self-improvement.

In addition to the stress on professional growth, principals seemed much more confident as instructional leaders, and developed greater evaluation consistency throughout the District. Twenty-two teachers were evaluated as either "needs improvement" or "does not meet district standards"--that's a 300% increase over last year.

This is encouraging. Training in various evaluation skills and use of the new instrument during some of the training seems to have had an effect on rating variability after all. However, an additional attempt was made to determine improved variability in ratings using a simulation. The data from pre- and posttests before and after training, was statistically analyzed and, in this case, results showed evaluators rating teachers below standards both times. Two things may account for this. The teacher in the videotaped instructional sequence was directed to give a poor performance, so it was desirable for evaluators to rate her lower. Also, no face-to-face conference with the teacher was possible in the simulated activity, therefore evaluators didn't have to face the negative feedback from the teacher on-the-job.

The simulation results show significant changes in ratings after training on only three of ten items. Therefore, there remains much work to be done to increase the ability to discriminate among teachers' performance. When responses after training were compared to those of a panel of experts, the evaluators varied significantly from the experts

on half of the items and in all cases but one, they were lower than the experts. Generally speaking, evaluators were tougher in simulation ratings than on-the-job. This is natural and we can expect a drift toward higher ratings when the evaluator returns to the actual appraisal in his/her building again. The typical evaluator will overrate the teacher to avoid hostility in the supervisory conference.

A learning style inventory administered to evaluators revealed that most in the Des Moines Schools are concrete sequential learners. This seems to reflect the general population of educators. Perhaps concrete sequential learners are drawn to teaching and administration occupations naturally because of their linear organizational ability, their comfort in dealing with bottom lines and deadlines, and their orientation toward material well-being (salaries, prestige, control of others' behavior). This may be why many of them are managers, not leaders. They tend to conserve and protect status quo and improve what is. They are typically not risk takers.

With regard to training, it is important to adapt methods and procedures to meet the learning styles of the trainee, but it is equally important that the trainee be adaptive to the environment as well. This mutual adaptation can insure a high measure of success in an instructional setting.

As previously mentioned, this investigation was part of an Iowa State University/Des Moines Public Schools project for performance evaluation systems development. A total of \$24,000.00 was expended for this year's portion of the two-year project. Year two will deal with

development of an administrator evaluation system. The project and this investigation have produced enough significant gains to be termed a success. A cadre of teacher performance evaluation steering committee members will function as a training maintenance group in year two. Practice in evaluation skills will be offered to trained evaluators as will orientation be offered to new administrators. This provides a perfect opportunity to investigate some areas of training effectiveness next year.

### Recommendations for Practice

#### Trainers

This study was an effort to analyze the effectiveness of training teacher evaluators in specific steps in clinical supervision and teacher performance evaluation. Results of this analysis point to several suggestions for trainers.

1. The training, in this instance, addressed itself to immediate application of skills. The Des Moines School board made this system of teacher performance evaluation the "official" system the same year it was developed. One full year of training before direct application is preferable.
2. In order to more fully understand and evaluate the effectiveness of the training, an in-depth analysis should be done of the summative reports of evaluations done this year.
3. A series of appropriate instruments should be assembled, each of which truly measures effectiveness in evaluation skills taught in the

training. Instruments should be accompanied by an adequate set of directions for uniform administration.

4. Trainers need to find ways to measure the kinds of things that can't be observed. Perhaps multiple data sources should be used (teacher self-report, peer ratings and student ratings).

5. Development of input (lectures and instructional materials) relating to subject areas in which weaknesses were found would be helpful for future training. These areas include effective teaching behaviors, communication of expectations to teachers, completion of a summative evaluation report, and gathering observational data.

6. In testing for proficiency in topical data gathering, identification of all seven steps in the teaching episode should be required.

7. Input and practice regarding conference skills should concentrate on all three skill areas (open, body, and close) equally to encourage good climate and goal setting, as well as coaching and counseling.

8. The indication that conference climate skills carried over from simulation to the job setting recommends that practice sessions be used, both mandatory and optional, for other skills as well.

9. Videotaped simulations were used successfully in the present investigation, but lack of ideal models of teaching in the videotaped segments suggest more work needs to be done in designing media for use in training. Longer segments would be helpful as well.

10. The setting for training workshops should be neutral ground. A hotel function room or university space would be recommended.



### Evaluators

Results of data analysis in this study lead to recommendations for evaluators as well as trainers. Some recommendations for evaluators are:

1. Evaluators should take advantage of practice sessions offered by the maintenance cadre in the district in year two for maintenance of skill level.
2. Areas of special concern to evaluators indicated by no significant gains after training are: open and close of the conference, identification of effective teaching behaviors and completion of a summative evaluation report. Review and improvement of workshop materials on these subjects is recommended.
3. Teachers' responses to queries about the kind of supervisors/evaluators they wanted must be studied by evaluators. The responses to these items indicate not only what teachers want, but also what they deserve, in a supervisor/evaluator.
4. Evaluators should take responsibility for helping new administrators to gain evaluation skills and offer support to these people during the orientation process.
5. Evaluators must make a continued effort to work at interrater reliability in the district. This will insure equal treatment of teachers across buildings concerning evaluation and subsequent decisions regarding transfer, tenure, promotion and retention.
6. It is important to consider whether evaluation progress in the district should take place in the future as a function of policy or as a

result of continued training. Evaluators' input in this matter is essential.

#### Recommendations for Further Research

Looking at this study and interpreting the findings leads to thoughts of further research. This "macro-study" touched upon many aspects of teacher performance evaluation in its investigation of the effects of training evaluators. Some of the specific steps in clinical supervision and teacher performance evaluation suggest topics for dissertations and/or staff research. For example, more work needs to be done in conferencing techniques and the related area of interpersonal communications. Techniques need to be developed to measure proficiency in establishing good conference climate, as well as basic conferencing skills. New methods for practice are essential as many administrators do not react well to role playing. Perhaps more work in self-evaluation in the conference, by use of audio or videotape, using valid and reliable criteria is necessary.

A good area of study might also be in the evaluator's use of formative data to complete a summative evaluation instrument. Analyzing judgments is no small task and maintaining objectivity and freedom from bias may be virtually impossible. However, any findings from a study would be helpful to move in that direction.

The responses teachers gave to items pertaining to supervision/evaluation in this study suggest several areas of concern, most centering around leadership and administrative style of the principal. More research studies would be helpful to relate effective leadership

characteristics to positive climate and effective schooling.

Another suggestion for consideration might be the replication of the present study with a heavy emphasis on measuring skills' application on-the-job. This investigation dealt heavily with simulation in major areas of evaluation. Future studies should work to develop instruments to measure effectiveness in the job setting and to provide more experiential training. It may be important to measure gains, in some cases, during training, but studies then must deal with the question of how to move these gains to the field.

Development of a teacher performance evaluation system is only one step in the quest for better instruction. Validation of the system must come from student gain scores. Input-process-output research could make the link of student achievement and teacher performance evaluation. Such research projects could help to validate teacher performance evaluation models and improve instructional quality as well. These projects are far beyond dissertations or staff research. Money for projects of this magnitude must be procured from public or private foundations.

## BIBLIOGRAPHY

1. Arbucci, R. P. A study of the relationship between a clinical supervision innovation and the attitude of a professional staff toward instructional supervision in a suburban school district. Unpublished dissertation. Columbia University Teachers College, 1978.
2. Barrett, R. S. Performance rating. Chicago: Science Research Association, 1966.
3. Bayroff, A. C., H. R. Haggerty, and E. A. Rundquist. Validity of ratings as related to rating techniques and conditions. *Personnel Psychology* 7, No. 1 (Spring 1954): 93-113.
4. Beach, Dale S. *Personnel: The management of people at work*. New York: Macmillan and Company, Inc., 1980.
5. Bernadin, H. John. Effects of rater training on leniency and halo errors in students' rating of instructors. *Journal of Applied Psychology* 63, No. 3 (1978): 301-308.
6. Bloom, B. S. *Human characteristics and school learning*. New York: McGraw-Hill, 1976.
7. Blumberg, Arthur. Supervisor-teacher relationships: A look at the supervisory conference. *Administrators Notebook* 19, No. 1 (September 1970): 1-4.
8. Bolton, Dale L. *Selection and evaluation of teachers*. Berkeley, California: McCutcheon, 1973.
9. Borg, Walter R., and Meredith D. Gall. *Educational research*. New York: Loneman Publishing Company, 1979.
10. Borman, Walter C. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology* 60, No. 5 (1975): 556-560.
11. Borman, W. C., and Marvin D. Dunnette. Behavior-based vs. trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology* 60, No. 5 (1975): 561-563.
12. Boyan, Norman J., and Willis D. Copeland. A training program for supervisors: Anatomy of an educational development. *Journal of Educational Research* 68, No. 3 (1974): 100-115.

13. Campbell, J. P., M. D. Dunnette, E. E. Lawler, and K. E. Weick. Managerial behaviors, performance and effectiveness. New York: McGraw-Hill, 1970.
14. Coffin, Gregg. Tracing our origins. Boston, Massachusetts: Urban Schools Collaborative, Northeastern University, 1980. (Videotaped)
15. Dunn, Rita, and Kenneth Dunn. Teaching students through their individual learning styles. Englewood Cliffs, New Jersey: Prentice-Hall, 1978.
16. Dunn, Rita, and Kenneth Dunn. Teaching to learning styles works. Education U.S.A. (November 30, 1981): 108.
17. Eaker, R. E. An analysis of the clinical supervision process as perceived by selected teachers and administrators. Unpublished dissertation. University of Tennessee, 1972.
18. Edwards, Allen L. Experimental design in psychological research. New York: Holt, Rhinehart and Winston, 1960.
19. Fisher, C. W., N. N. Filby, R. Marliave, L. S. Cahen, M. M. Dishaw, J. E. Moore, and D. C. Berliner. Teaching behaviors, academic learning time and student achievement: Final report of phase III-B, beginning teacher evaluation study technical report series (Tech. Rep. V-1). San Francisco, California: Far West Laboratory, 1978.
20. Flanders, N. W. Interaction analysis and clinical supervision. Journal of Research and Development in Education 9, No. 2 (1976): 36-46.
21. Frudden, Sarah J. A comparison of school administrators' ratings of teaching performance utilizing varied instructional materials in the assessment process. Unpublished dissertation. Iowa State University, 1980.
22. Gappert, Gary. Research for Better Schools, Inc. Report to Administrator Groups, Washington, D.C., 1980.
23. Garman, N. B. A study of clinical supervision as a resource of college teachers of English. Unpublished dissertation. University of Pittsburgh, 1971.
24. Goldhammer, R., R. Anderson, and R. Krajewski. Clinical supervision, special methods for the supervision of teachers. New York: Holt, Rinehart, and Winston, 1980.
25. Goldhammer, R., M. Cogan, and R. Anderson. Clinical supervision. Boston: Houghton-Mifflin Co., 1973.

26. Gregorc, Anthony. Learning/teaching styles: Potent forces behind them. *Educational Leadership* 36, No. 4 (January 1979): 234.
27. Gruenfeld, L. W., and P. Weissenberg. Supervisory characteristics and attitudes toward performance appraisals. *Personnel Psychology* 19, No. 2 (Summer 1966): 143-152.
28. Gudridge, B. M. Teacher competency: Problems and solutions. Sacramento, California: Education News Service for AASA, 1980.
29. Haeefele, Donald L. How to evaluate the teacher--let me count the ways. *Phi Delta Kappan* 61, No. 5 (January 1980): 349-352.
30. Howsam, Robert. Who's a good teacher? Burlingame, California: California Teachers' Association, 1960.
31. Hunt, David. Matching models in education. Ontario, Canada: Ontario Institute for Studies in Education, 1971.
32. Hunter, Madeline. Prescription for improved instruction. El Sagundo, California: TIP Publications, 1976.
33. Hunter, Madeline. Diagnostic teaching. *The Elementary School Journal* 80, No. 1 (1979): 41-46.
34. Katz, Robert L. Skills of an effective administrator. *Harvard Business Review* 33, No. 1 (1955): 33-42.
35. Kerr, B. J. An investigation of the process of using feedback data within the clinical supervision cycle to facilitate teachers' individualization of instruction. Unpublished dissertation. University of Pittsburgh, 1976.
36. Kiesler, D. J., J. C. Anchin, M. J. Perkins, B. M. Chirico, E. M. Kyle, and E. J. Federman. The impact message inventory. Richmond: Virginia Commonwealth University, 1975, 1976.
37. Kirby, Patricia. Cognitive style, learning style, and transfer skill acquisition. Columbus, Ohio: National Center for Studies in Education, 1971.
38. Kirchner, Wayne K., and Donald J. Reisberg. Difference between better and less effective supervisors in appraisal of subordinates. *Personnel Psychology* 15, No. 3 (1962): 295-302.
39. Kowalski, Theodore J. The principal's role in curriculum development: What are the barriers? *Contemporary Education* 50, No. 3 (Spring 1979): 159-161.

40. Krajewski, R. J. Clinical supervision to facilitate teacher self-improvement. *Journal of Research and Development in Education* 9, No. 2 (Winter 1976): 58-66.
41. Krey, Robert D. Assumptions supporting structure in clinical supervision. *Contemporary Education* 49, No. 1 (Fall 1977): 16-23.
42. Latham, Gary P., Kenneth L. Palmer, and Everett Hildebaugh. Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology* 60, No. 5 (1975): 550-555.
43. Latham, Gary P., Kenneth N. Wexley, and Elliot D. Pursell. Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology* 60, No. 5 (1975): 555-560.
44. Levin, Benjy. Teacher evaluation--A review of research. *Educational Leadership* 37, No. 3 (December 1979): 240.
45. Levine, J., and J. Butler. Lecture vs. group discussion in changing behavior. *Journal of Applied Psychology* 4, No. 8 (February 1952): 29-33.
46. Lopez, F. M. Evaluating employee performance. Chicago: Public Personnel Association, 1968.
47. Manatt, Richard P. Evaluating and improving teacher performance. Englewood, Colorado: Educational Consulting Associates, 1977.
48. Manatt, Richard P., and Dianne Blackmer. The school improvement model consortium. Ames, Iowa: School Improvement Model Project, N229 Quadrangle, Iowa State University, December, 1980.
49. Manatt, Richard P., and Dorothy A. Faast. Evaluating teacher performance. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1981.
50. Medley, D. M. Teacher competence and teacher effectiveness: A review of process-product research. Washington, D.C.: American Association of Colleges for Teacher Education, 1977.
51. Miner, John B. Management appraisal: A capsule review and current references. *Business Horizons* 11, No. 5 (October 1968): 83.
52. Minton, Erlene. Clinical supervision. Englewood, Colorado: Educational Consulting Associates, 1979.
53. Morine, Greta. Planning skills: Paradox and parodies. *Journal of Teacher Education* 24, No. 2 (1973): 135.

54. Mosher, R. L., and D. E. Purpel. *Supervision: The reluctant profession*. Boston: Houghton-Mifflin Company, 1972.
55. Popham, W. James, and Eva I. Baker. *Systematic instruction*. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.
56. Popham, W. James, and Kenneth A. Sirotnik. *Educational statistics: Use and interpretation*. New York: Harper and Row,
57. Redfern, George B. *How to appraise teaching performance*. Columbus, Ohio: School Management Institute, Inc., 1963.
58. Redfern, George B. *Evaluating teachers and administrators: A performance objectives approach*. Boulder, Colorado: Westview Press, Inc., 1980.
59. Rosenshine, Barak. *Content, time, and direct instruction*. In *Research on teaching*. Edited by Penelope L. Peterson and Herbert J. Walberg. Berkeley, California: McCutchan, 1979.
60. Rucker, Dennis. *A study of principals' teaching style preference and teachers' teaching style as a source of bias in teacher evaluation*. Unpublished dissertation. Iowa State University, 1981.
61. Schmuck, Richard, Philip Runkel, Jane Arends, and Richard Arends. *The second handbook of organization of development in schools*. Palo Alto, California: Mayfield Publishing Co., 1977.
62. Sealy, Ronald. *Evaluation of educational personnel and the law*. Boston: Law and Education Institute, 1980.
63. Sergiovanni, Thomas J., and Robert J. Starrett. *Supervision: Human perspectives*. New York: McGraw-Hill, 1979.
64. Shuma, K. Y. *Changes effectuated by a clinical supervisory relationship which emphasizes a helping relationship and a conference format made congruent with the establishment and maintenance of this helping relationship*. Unpublished dissertation. University of Pittsburgh, 1973.
65. Snyder, Karolyn. *ACT: Administrator-for-change-training, module six*. Lubbock, Texas: Pedamorphosis, Inc., 1978.
66. Soar, R. S., and R. M. Soar. *Classroom behavior, pupil characteristics, and pupil growth for the school year and summer*. Gainesville, Florida: College of Education, University of Florida, December, 1973.



67. Sweeney, James. Does the principal make a difference? *Educational Leadership* 39, No. 5 (February 1982): 346-352.
68. Taba, Hilda. *Curriculum development, theory and practice*. New York: Harcourt Brace and World, 1962.
69. Teacher Assessment Project. *Intermediate language arts*. College of Education, University of Georgia, Athens, Georgia, 1978.
70. Turner, H. M. *The implementation and critical documentation of a model of clinical supervision: A case study*. Unpublished dissertation. University of California, 1976.
71. Tyler, Ralph W. *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press, 1950.
72. Young, James M., and Robert L. Heischberger. Teachers' perceptions of an effective school supervision and evaluation program. *Education* 96, No. 1 (Fall 1975): 10-19.

**APPENDIX A: WORKSHOP PLANNERS**

GROUP OR  
 SCHOOL: Des Moines Administrators and Supervisors

Page # 1 WORKSHOP PLANNER

DATE(S): August 10, 1981

of 2 Des Moines Public Schools Project

Director: Richard P. Manatt, Ph.D.

ATTENDING: All evaluators

Coordinator: Dorothy A. Faast

TIME	TOPIC	PRESENTER	MODE	VISUALS	HANDOUTS
8:30	Teacher Performance Eval.: State of the Art	Manatt	Lecture	Trans- parencies	
9:45	Break				
10:00	Zero Warning Walk-in	Manatt Sweeney	Large Group	Videotape	
11:30	Lunch				
12:30	Lesson Plan Analysis Frazier I Data Capture by Timeline Frazier I	Manatt Stow	Large Group	Videotape	Lesson Plan Lesson Analysis Instrument
2:00	Break				
2:15	Continue Frazier I	Manatt Stow			
3:00	Question and Answer Preview Day 2	Manatt			
3:30	Dismissal				

GROUP OR  
 SCHOOL: Des Moines Administrators and Supervisors

Page # 2 WORKSHOP PLANNER

DATE(S): August 11, 1981

of 2 Des Moines Public Schools Project

ATTENDING: All evaluators

Director: Richard P. Manatt, Ph.D.

Coordinator: Dorothy A. Faast

TIME	TOPIC	PRESENTER	MODE	VISUALS	HANDOUTS
8:30	Review/Preview	Stow	Large Group	Trans- parencies	
8:45	Conferencing Techniques	Manatt	Large Group	Trans- parencies Videotape	
11:30	Lunch				
12:30	Objectives for Improving Instruction	Stow	Large Group	Trans- parencies	Worksheets
1:00	Lesson Analysis Data Capture  Frazier II Lindholm	Stow Sweeney Faast	Elem./Sec. Breakout	Videotape	Lesson Analysis Instrument Lesson Plan
3:15	Question and Answer	Stow Sweeney Faast			
3:30	Dismissal				

GROUP OR  
 SCHOOL: Des Moines Administrators and Supervisors

Page # 1 WORKSHOP PLANNER

DATE(S): September 23 - 24

of 2 Des Moines Public Schools Project

ATTENDING: ½ evaluators each day

Director: Richard P. Manatt, Ph.D.

Coordinator: Dorothy A. Faast

TIME	TOPIC	PRESENTER	MODE	VISUALS	HANDOUTS
8:30	August Workshop Recap	Manatt	Large Group	Trans- parencies	
9:00	"Mirrors on the Classroom" Overview/Assumptions Techniques: Academically Engaged Time Classroom Interaction Teacher Talk (questions - feedback)	Mitsakos	Large Group		Workbook
9:30	Guided Practice: Fleming Data Gathering	Mitsakos	Large Group	Videotape	ASCD kit
10:00	Postobservation Analysis	Mitsakos	Small Groups (Elem.-Sec.)		
	Postobservation Conference	Mitsakos			DSM Evaluation Instrument
11:00	Debriefing	Mitsakos	Large Group		
11:30	Lunch				

GROUP OR

SCHOOL: Des Moines Administrators and Supervisors

DATE(S): September 23 - 24

ATTENDING: ½ evaluators each day

Page # 2 WORKSHOP PLANNER

of 2 Des Moines Public Schools Project

Director: Richard P. Manatt, Ph.D.

Coordinator: Dorothy A. Faast

TIME	TOPIC	PRESENTER	MODE	VISUALS	HANDOUTS
1:00	"Mirrors for the Classroom" Review Assumptions Techniques: Teacher Talk Giving Directions Seeking Control Cognitive Clarity	Mitsakos	Large Group		
2:30	Guided Practice; Frazier III Data Capture Postobservation Analysis	Mitsakos	Small Groups Subgroups	Videotape	DSM Evaluation Instrument
3:00	Debriefing  Summary	Mitsakos  Mitsakos/ Manatt	Large Group		

GROUP OR  
SCHOOL: Des Moines Administrators

Page # 1 WORKSHOP PLANNER  
of 1 Des Moines Public Schools Project

DATE(S): October 7 - 8

Director: Richard P. Manatt, Ph.D.  
Coordinator: Dorothy A. Faast

ATTENDING: ½ evaluators each day

TIME	TOPIC	PRESENTER	MODE	VISUALS	HANDOUTS
8:30	Review/preview Capturing Topical Data	Manatt	Large Group	Videotape	
9:15	The Postobservation Conference (Determining Entry Level Skills)	Manatt	Large Group	Trans- parencies	
10:00	Coffee Break				
10:15	The Postobservation Conference (Guided Practice)	Manatt	Triads		Situation data sheets
11:00	Types of Postobservation Conferences	Stow	Large Group	Trans- parencies	
11:30	Lunch				
12:30	The Conference Process	Sweeney	Large Group	Trans- parencies	
2:00	Coffee Break				
2:15	Guided Practice	Sweeney	Triads	Videotape	Situation data sheets
3:15	Summary	Sweeney/ Manatt/ Stow	Large Group		
3:30	Dismissal				

86

GROUP OR  
SCHOOL: Des Moines Evaluators

Page # 1 WORKSHOP PLANNER

DATE(S): Friday - November 13

of 3 Des Moines Public Schools Project

ATTENDING: ½ evaluators a.m. ½ evaluators p.m.

Director: Richard P. Manatt, Ph.D.

Coordinator: Dorothy A. Faast

TIME	TOPIC	PRESENTER	MODE	VISUALS	HANDOUTS
7:30	Introduction	Bridgewater	Large Group		
7:45	Lesson Analysis	Frudden	Large Group	Videotape Workbook	Workbook Mr. Sims
9:00	Break				
9:15	Lesson Analysis Guided Practice Grabau Yoho Frazier I	Frudden	Large Group	Lesson Plans	Lesson Plans
10:00	The Summative Conference	Manatt Brdigewater	Large Group	DSM Eval. Instrument	DSM Eval. Instrument
12:00	Dismissal				



GROUP OR  
 SCHOOL: Des Moines Evaluators

Page # 2 WORKSHOP PLANNER  
 of 3 Des Moines Public Schools Project  
 Director: Richard P. Manatt, Ph.D.  
 Coordinator: Dorothy A. Faast

DATE(S): Friday - November 13

ATTENDING: ½ evaluators a.m. ½ evaluators p.m.

TIME	TOPIC	PRESENTER	MODE	VISUALS	HANDOUTS
1:00	Introduction	Bridgewater	Large Group		
1:15	Lesson Analysis	Frudden	Large Group	Videotape Workbook	Workbook Mr. Sims
2:30	Break				
2:45	Lesson Analysis Guided Practice Grabau Yoho Frazier I	Frudden	Large Group	Lesson Plans	Lesson Plans  DSM Eval. Instrument
3:30	The Summative Conference	Manatt Bridgewater	Large Group	DSM Eval. Instrument	
5:30	Dismissal				

GROUP OR  
SCHOOL: Des Moines Evaluators

Page # 3 WORKSHOP PLANNER

DATE(S): Saturday, November 14, 1981

of 3 Des Moines Public Schools Project

ATTENDING: All evaluators ½ day

Director: Richard P. Manatt, Ph.D.

Coordinator: Dorothy A. Faast

TIME	TOPIC	PRESENTER	MODE	VISUALS	HANDOUTS
8:00	Introduction	Bridgewater	Large Group		
8:00	Conference Practice	Sweeney	Triads	Workbook	Scenario DSM Evaluation Instrument
9:30	Break				
9:45	Objectives for Improvement	Stow DSM Cadre	Small Groups		
12:30	Dismissal				

**APPENDIX B: TRAINING OBJECTIVES**

PERFORMANCE EVALUATION OF THE  
EDUCATIONAL PROFESSIONS

objectives for skills building workshops  
for educational supervisors and administrators

At the completion of the workshops, given the extensive evaluation and supervisory tools provided in the written materials and utilizing videotaped classroom and management vignettes, each participant will be able to:

- (1) Independently produce an evaluation system/cycle centering on productive teaching, counseling and administrative behaviors.
- (2) Discuss critically the various alternatives for increasing teacher effect, viz., criterion-referenced gain scores, PBI, job improvement targets, motivation and reinforcement.
- (3) Improve the processes and criteria used by his/her building/district for preconferences, classroom observations and postconferences.
- (4) Identify what a teacher is doing that enhances learning (in the classroom setting) as well as what behaviors get in the way of learning.
- (5) Coach and counsel teachers on the improvement of teaching and learning productivity with concomitant concern for the teacher's time and well-being.
- (6) Utilize the Dimensional Management Strategies (DMS) to enhance school productivity of administrators, counselors and teachers.
- (7) Value the role of the first-line supervisor in the improvement of instruction.

**APPENDIX C: INSTRUMENTS**

INSTRUCTIONAL PLANS AND MATERIALS ASSESSMENT SCALE

**DIRECTIONS:** For each of the items, select the number above the statement that best describes the teacher's lesson planning and/or materials. Record your choice by circling the number on the answer sheet.

1. The learning is stated in terms of what the student will be able to do after mastery of the educational objective(s).

1	2	3	4	5
No objectives are stated; are in general terms, are teacher behaviors; are inappropriate.		Objectives usually are appropriate to topic and student; are usually specifically stated and measurable.		All objectives are appropriate, specifically stated and measurable.

2. Content, materials and media selected are appropriate vehicles for teaching the objectives of the lesson.

1	2	3	4	5
Content, materials and media are not appropriate.		Content, materials and media usually are appropriate; they usually are objective-specific; selection of resources is adequate.		Content, materials and media are all appropriate and objective-specific; selection is wide and imaginative.

3. The designated instructional procedures are appropriate to accomplish lesson objectives.

1	2	3	4	5
Procedures are not specified or are inappropriate for objectives and/or students.		Procedures are usually specified and usually appropriate for students.		Procedures are specific, varied, and appropriate for students.

4. The lesson plans outline instruction that is planned to accommodate a variety of cognitive levels.

1	2	3	4	5
Students are required to acquire factual information, explain, or summarize.		Students are required to apply information, clarify parts of complex ideas, or synthesize information.		Students are required to evaluate, judge, or value ideas and information.

5. Differences in student capabilities are evidenced in the planning of instruction.

1	2	3	4	5
No provision for individual student capabilities is planned.		Plans intermittently provide for individual student capabilities.		Plans consistently provide for individual student capabilities, i.e., remedial, maintenance and enrichment activities are specified.

6. Individual rates of learning are provided for within the instructional planning.

1	2	3	4	5
All students are expected to work on the same objectives at the same time.		Students work at their own rate on some of the objectives, some of the time.		Students work at their own rate on objectives with assistance from teacher or peer; may work on enrichment objectives, as appropriate.

7. Evaluation of student progress on the objectives is indicated.

1	2	3	4	5
No evaluation of student progress is indicated.		Only general, summative evaluation of student progress is indicated.		Specific, measurable student progress is indicated on each objective.

FEEDBACK ANALYSIS INVENTORYID Number

Supervisor \_\_\_\_\_

Teacher \_\_\_\_\_

Observer \_\_\_\_\_

Below are 11 descriptors to help you analyze feedback techniques and process in the conference you observed. Carefully examine the descriptors and circle the number which best characterizes what occurred in the conference.

										<u>Comments</u>
1.	Took appropriate action to set the teacher at ease	1	2	3	4	5	6	7	Supervisory behavior caused tenseness, or anxiety	
2.	Clearly stated conference purpose	1	2	3	4	5	6	7	No purpose given	
3.	Clearly presented the data in specific terms	1	2	3	4	5	6	7	Vague and confusing presentation of data	
4.	Appropriate amount of information discussed	1	2	3	4	5	6	7	Overload of information	
5.	Evaluative	1	2	3	4	5	6	7	Descriptive	
6.	Supervisor shared information	1	2	3	4	5	6	7	Supervisor gave advice	
7.	Well-timed	1	2	3	4	5	6	7	Poorly timed	
8.	Discussed alternatives	1	2	3	4	5	6	7	Emphasized the "one best path"	
9.	Effectively checked for the teacher's willingness to change	1	2	3	4	5	6	7	Was oblivious to teacher's attitude toward change	



Feedback Analysis Inventory - Page 2

									<u>Comments</u>
10.	Effectively summarized areas of agreement and/or disagree- ment	1	2	3	4	5	6	7	Made no attempt to summarize areas of agree- ment and/or dis- agreement
11.	Summarized main points of con- ference	1	2	3	4	5	6	7	Made no effort to summarize main points of conference

SCALE:                   1--Not at all                   3--Moderately so  
                           2--Somewhat                   4--Very much so

WHEN I AM WITH THIS PERSON, HE MAKES ME FEEL . . .

- |   |   |
|---|---|
| D 1. <input type="checkbox"/> bossed around.        | H 7. <input type="checkbox"/> annoyed.                              |
| M 2. <input type="checkbox"/> distant from him.     | N 8. <input type="checkbox"/> loved.                                |
| AFF 3. <input type="checkbox"/> entertained.        | AGR 9. <input type="checkbox"/> welcome with him.                   |
| AGR 4. <input type="checkbox"/> appreciated by him. | D 10. <input type="checkbox"/> like an impersonal audience.         |
| H 5. <input type="checkbox"/> cold.                 | M 11. <input type="checkbox"/> uneasy.                              |
| N 6. <input type="checkbox"/> complimented.         | AFF 12. <input type="checkbox"/> like I'm just one of many friends. |
- 

Using the same scale, consider this statement:

WHEN I AM WITH THIS PERSON, HE MAKES ME FEEL THAT . . .

- |   |   |
|---|---|
| D 13. <input type="checkbox"/> I want to tell him to give someone else a chance to make a decision. | H 19. <input type="checkbox"/> I want to get away from him.         |
| M 14. <input type="checkbox"/> I should be cautious about what I say or do around him.              | N 20. <input type="checkbox"/> I shouldn't hesitate to call on him. |
| AFF 15. <input type="checkbox"/> I could lean on him for support.                                   | AGR 21. <input type="checkbox"/> I could ask him to do anything.    |
| AGR 16. <input type="checkbox"/> I can ask him to carry his share of the load.                      | D 22. <input type="checkbox"/> I want to protect myself.            |
| H 17. <input type="checkbox"/> I want to stay away from him.  | M 23. <input type="checkbox"/> I should leave him alone.            |
| N 18. <input type="checkbox"/> I can join in the activities.  | AFF 24. <input type="checkbox"/> I should like him.                 |
-

SCALE:                   1--Not at all                   3--Moderately so  
                           2--Somewhat                   4--Very much so

WHEN I AM WITH THIS PERSON, IT APPEARS TO ME THAT . . .

- D 25.  he wants to be the center of attention.                   H 31.  he's carrying a grudge
- M 26.  he doesn't want to get involved with me.                   N 32.  he trusts me.
- AFF 27.  he carries his share of the load.                   AGR 33.  he's genuinely interested in me.
- AGR 28.  his time is mine if I need it.                   D 34.  he thinks he's always in control of things.
- H 29.  he thinks it's every man for himself.                   M 35.  as far as he's concerned, I could just as easily be someone else.
- N 30.  he wants to be helpful.                   AFF 36.  he enjoys being with people.

D = Dominant

AGR = Agreeable

M = Mistrusting

H = Hostile

AFF = Affiliative

N = Nurturant

S U R V E Y   Q U E S T I O N N A I R E  
S C H O O L   S U P E R V I S I O N   A N D   E V A L U A T I O N   P R O G R A M

Part 1

Please check the appropriate box in reacting to the following statements:

1. There is a definite need for supervision and evaluation of teachers in the public school.  Yes  No  Neutral
2. The supervisor is quite often seen as potentially dangerous to a teacher.  Yes  No  Neutral
3. Teachers should take part in developing or selecting evaluation instruments, so that they know the criteria against which they are being judged.  Yes  No  Neutral
4. Evaluation should be used to diagnose teachers' performance so they can strengthen their weaknesses through inservice education.  Yes  No  Neutral
5. Evaluation should be something in which teachers have a part along with students, parents, and administrators.  Yes  No  Neutral
6. One way out of the evaluation dilemma is to put the focus on the learner, not the teacher, and to involve everyone in the business of assessing or supervising everything we do all the time.  Yes  No  Neutral
7. It is important for the supervisor to have some understanding of the teacher's educational philosophy and how the teacher views his own profession.  Yes  No  Neutral
8. Instead of focusing major attention on providing the basis for dismissal or continued employment, evaluation programs should focus attention on improving the performance of the employee presently serving on the job.  Yes  No  Neutral
9. The building principal should spend at least 35% of his/her time in supervising.  Yes  No  Neutral
10. My building principal spends at least 25% of his/her time in supervising.  Yes  No  Neutral

**TEACHER PERFORMANCE EVALUATION ASSESSMENT SCALE  
(ASCD NORMING)**

\_\_\_\_\_  
(Teacher's Name)

\_\_\_\_\_  
(Years Experience  
in District)

\_\_\_\_\_  
(Building)

\_\_\_\_\_  
(Evaluatee's Signature)

\_\_\_\_\_  
(Date)

\_\_\_\_\_  
(Evaluator's Signature)

\_\_\_\_\_  
(Date)

**Directions:** For each criterion, please select the number representing the phrase which best describes the appraisee's performance on that item. Enter that number (by circling) on the separate answer sheet.

**A. PRODUCTIVE TEACHING TECHNIQUES**

**CRITERIA**

**LEVELS OF PERFORMANCE**

The teacher . . .

1. Communicates effectively with students.

1  
Not Observed

2  
Communications from the teacher are frequently unclear; students often appear confused.

3  
Communications from the teacher are usually clear but student input is not encouraged.

STANDARD  
4  
Communications from the teacher are clear. Relevant dialog is encouraged.

5  
In addition to meeting the standard, the teacher is extremely skillful in using a variety of verbal and nonverbal communications.

2. Organizes instruction around objectives.

1  
Not Observed

2  
Instruction does not relate to the stated objectives.

3  
Instruction marginally relates to the stated objectives.

4  
The teacher clearly organizes instruction around the stated objectives.

5  
In addition to meeting the standard, all objectives are appropriate, specifically stated and measurable.

3. Demonstrates ability to select appropriate learning content.

1  
Not Observed

2  
Learning content does not relate to approved curriculum guide(s).

3  
Learning content is marginally related to the approved curriculum guide(s).

4  
Learning content is related to the approved curriculum guide(s).

5  
In addition to meeting the standard, the teacher shows initiative and leadership in review and development of curriculum.

**TEACHER PERFORMANCE EVALUATION ASSESSMENT SCALE  
(ASCD NORMINO)**

**A. PRODUCTIVE TEACHING TECHNIQUES (continued)**

CRITERIA	LEVELS OF PERFORMANCE				
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
4. Identifies capabilities of students.	Not Observed	The teacher does not provide for individual capabilities of students.	The teacher intermittently provides for individual capabilities of students.	<b>STANDARD</b> The teacher identifies and teaches to student capabilities.	In addition to meeting the standard, the teacher provides remedial and/or enrichment activities.
5. Provides students with specific oral, evaluative feedback.	Not Observed	The teacher gives no oral evaluative feedback.	The teacher is inconsistent in giving oral evaluative feedback.	The teacher gives specific oral evaluative feedback.	In addition to meeting the standard, the teacher gives oral feedback with reinforcement and encouragement.

COMMENTS:

TEACHER PERFORMANCE EVALUATION ASSESSMENT SCALE  
(ASCD NORMING)

B. ORGANIZED, STRUCTURED CLASS MANAGEMENT

CRITERIA	LEVELS OF PERFORMANCE				
The teacher . . .	<u>STANDARD</u>				
6. Organizes the educational setting.	<u>1</u> Not Observed	<u>2</u> The teacher displays little or no skill in organizing the educational setting.	<u>3</u> The educational setting is ineffectively managed, <u>i.e.</u> , the teacher chooses to use activities which are inappropriate for the physical environment.	<u>4</u> A functional classroom environment is maintained. Appropriate activities are selected.	<u>5</u> In addition to meeting the standard, the teacher assesses and adjusts the educational setting to provide for a variety of learning styles.
7. Demonstrates evidence of personal organization.	<u>1</u> Not Observed	<u>2</u> The teacher is disorganized in lesson preparation and organization.	<u>3</u> The teacher intermittently presents materials in an organized manner.	<u>4</u> Appropriate lesson preparation and organization of work is evident, <u>i.e.</u> , materials are available; presentations progress logically.	<u>5</u> In addition to meeting the standard, the teacher shows evidence of long-range planning.
8. Organizes students for effective instruction.	<u>1</u> Not Observed	<u>2</u> There is little or no evidence of established classroom routine; students appear to be confused.	<u>3</u> The teacher is inconsistent and/or ineffective in maintaining classroom routine.	<u>4</u> The teacher has established an effective classroom routine which students clearly understand.	<u>5</u> In addition to meeting the standard, the students participate in carrying out the classroom routine.

COMMENTS:

TEACHER PERFORMANCE EVALUATION ASSESSMENT SCALE  
(ASCD NORMING)

C. INTELLECTUAL STIMULATION

CRITERIA	LEVELS OF PERFORMANCE				
The teacher . . . 9. Helps students develop efficient learning skills and work habits.	<u>1</u> Not Observed	<u>2</u> The teacher makes no effort to help students develop efficient learning skills and work habits.	<u>3</u> The teacher makes an effort to help students develop efficient learning skills and work habits but occasionally models them incorrectly.	<u>STANDARD</u> <u>4</u> The teacher is able to model and reinforce efficient learning skills and work habits.	<u>5</u> In addition to meeting the standard, the teacher stimulates students to assume responsibility in a wide variety of settings.
COMMENTS:					

D. POSITIVE INTERPERSONAL RELATIONS

CRITERIA	LEVELS OF PERFORMANCE				
The teacher . . . 10. Promotes self-discipline and responsibility	<u>1</u> Not Observed	<u>2</u> The teacher dissuades students from being responsible and self-disciplined through constant exposure to activities requiring dependency.	<u>3</u> The teacher inconsistently provides opportunities for students to develop responsibility/self-discipline.	<u>STANDARD</u> <u>4</u> The teacher provides opportunities for students to demonstrate responsible behaviors.	<u>5</u> In addition to meeting the standard, the teacher encourages all students to demonstrate responsible behaviors in a wide variety of settings; e.g., through independent study, assignments, and group leadership roles.
COMMENTS:					



TEACHER PERFORMANCE EVALUATION  
(unrevised form)

Name \_\_\_\_\_ Date \_\_\_\_\_ Time \_\_\_\_\_

School \_\_\_\_\_ Grades \_\_\_\_\_ Subjects \_\_\_\_\_ Team Member \_\_\_\_\_ Non-Team Member \_\_\_\_\_

Circle type of activity observed: teacher lecture, general class discussion, question and answer format, small group, library work, demonstration, test, laboratory, film, T.V., pupil reports, rehearsals, (other \_\_\_\_\_)

**I. INTERPERSONAL RELATIONS**

**COMMENTS:**

N/A    Does                    Far  
or    Not    Oper.    Ex-    Ex-  
Obs.   Meet   below   ceeds   ceeds   Meets  
**PERFORMANCE STANDARDS**

- 1. Students
- 2. Staff
- 3. Parents


Objectives for Improved Interpersonal Relations: (Activities and time lines where applicable)

**II. PROFESSIONAL QUALITIES**

**COMMENTS:**

N/A    Does                    Far  
or    Not    Oper.    Ex-    Ex-  
Obs.   Meet   below   ceeds   ceeds   Meets  
**PERFORMANCE STANDARDS**

- 1. Professional Interest and Growth
- 2. Cooperation
- 3. Dependability


Objectives for Improved Professional Qualities: (Activities and time lines where applicable)

III. THE LEARNING SITUATION

COMMENTS:

N/A Does Far  
 or Not Oper. Ex- Ex-  
 Obs. Meet below ceeds ceeds Meets  
 PERFORMANCE STANDARDS

1. Planning and preparation
2. Knowledge of subject matter
3. Instructional techniques
4. Class management and control
5. Evaluation techniques


Objectives for Improving the Learning Situation: (Activities and time lines where applicable)

N/A Does Far  
 or Not Oper. Ex- Ex-  
 Obs. Meet below ceeds ceeds Meets  
 PERFORMANCE STANDARDS

Performance Rating 

--	--	--	--	--	--

Evaluator's Signature \_\_\_\_\_

Teacher's Signature \_\_\_\_\_ Date \_\_\_\_\_

**APPENDIX D: CORRELATION COEFFICIENTS OF CONFERENCE  
SIMULATION TRIAD**

Table D.1. Pearson product-moment correlation coefficients: observer (O) with teacher (T), observer (O) with supervisor (S) and supervisor (S) with teacher (T) before and after training

Item	Before training			After training		
	O with T	O with S	S with T	O with T	O with S	S with T
1	.2967	.2786	.2743	.6171**	.8073**	.5903**
2	-.0590	.3265	.2287	.0717	.2563	.3423
3	-.0084	-.2152	-.2578	.0951	.6213**	.1788
4	-.3572	-.3796	.1442	.5295*	.6535**	.2353
5	.4455*	.6147**	.5815**	.3880	.6952**	.6290**
6	.3695	.4482*	.5339*	.6184**	.7241**	.6084**
7	-.0410	.0129	.1689	.3662	.1899	.2848
8	.7789**	.5986**	.6050**	.6025**	.6086**	.6856**
9	.2181	.0251	.3191	.5622**	.7078**	.7307**
10	-.1367	.2207	.1893	.4619*	.1168	.2404
11	-.1920	.3961	.0666	.2682	.3194	.6796**

\*  $p < .05$ , .423.

\*\*  $p < .01$ , .537.