

AN ABSTRACT OF THE THESIS OF

Sheridan Davis Jones for the degree of Doctor of Philosophy
in Education presented on October 27, 1980

Title: A MODEL FOR IDENTIFYING EVALUATION

PROCEDURES THAT HAVE A POSITIVE

INFLUENCE ON TEACHER ATTITUDE

Abstract approved: _____

Redacted for Privacy

Dr. Ed Strowbridge

Teachers and administrators spend many hours of work reviewing the State of Oregon Guidelines and drafting district guidelines in an attempt to make the process of evaluation positive and productive. Most districts agree the primary purpose of evaluation should be for the improvement of instruction.

The amended state law, ORS 342.850, does specify that the process include pre-conference, data collecting, and post conference. These are the primary elements of the clinical cycle of supervision.

The review of related research revealed that the research concentrates on the effectiveness of evaluation models in collecting evaluation data, not on their affect on attitude, or on improving instruction.

Therefore, research that measures the effect of one model, clinical cycle of supervision, on attitude is justified.

This research focused on a null hypothesis. "There is no

significant difference in the attitude of teachers toward the evaluation process in districts that have a clinical cycle for evaluation and those districts that do not." The null hypothesis was tested at the .05 level of significance.

The procedure for generating data and testing the null hypothesis was as follows:

1. Develop a Likert Style scale based on Rensis Likert's research as outlined in Oppenheim's (1966) text "Questionnaire Design and Attitude Measurement."
2. Establish the validity of the scale by using the Known Group Process (Criterion Related Validity).
3. Test the scale for reliability by using the split half method (Spearman/Brown).
4. Select the school districts based on how well their evaluation process meets the criteria in this study. (Clinical cycle, district package, trained evaluators, and supervision and evaluation separate.)
5. Sample by using the random sampling technique with a sample approximately twenty-five percent of total population.
6. Use the one-way ANOVA to determine the acceptance or rejection of the null hypothesis based on the .05 level of significance.

The scale for measuring teacher's attitude had a validity quotient of .3584 which suggests a small but definite correlation between the

scale and its ability to measure attitude. The reliability quotient was .6471 which indicates a substantial relationship between the scale and its ability to produce the same result with different populations.

The primary analysis based on the results of a one-way ANOVA indicated the null hypothesis must be rejected. The Least Significant Difference Test established the fact that the district labeled "D" had a significantly lower score than the districts labeled A, B, and C.

The item analysis indicated that of all the various elements in the evaluation process, those that are directly connected to the clinical cycle (Pre-conference, data collecting, and post conference) are viewed in a positive way by teachers. Unsolicited comments by the respondents indicate that there is a wide discrepancy between what the districts outlined in writing as the process of evaluating and the teacher's perception of that procedure.

The complete analysis of the data indicated a need for further research into other attitude objects (interpersonal relationships) and the importance of each component of the evaluation process on teachers' attitudes. (Clinical cycle, evaluation package, trained evaluators, separation of evaluation and supervision.)

The results of this study suggest that careful application of evaluation procedures and modifications, identified in this project, will provide valuable information to districts in their attempt to improve teachers' attitudes through the evaluation process.

© Copyright by Sheridan Davis Jones
All Rights Reserved
October 27, 1980

A Model for Identifying Evaluation Procedures
That Have a Positive Influence
on Teacher Attitude

by

Sheridan Davis Jones

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed October 1980

Commencement June 1981

APPROVED:

Redacted for Privacy

Professor of Education
in charge of major

Redacted for Privacy

Dean of School of Education

Redacted for Privacy

Dean of Graduate School

Date thesis is presented October 27, 1980

Typed by Opal Grossnicklaus for Sheridan Davis Jones

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Dr. Ed Strowbridge for his assistance throughout the project and in the completion of the program.

Many thanks to the members of my committee: Drs. John B. Corliss, Al Leeland, Frank Cross, and Knud Larsen for their astute direction and support for this study.

I am also grateful for the cooperation from public school superintendents, principals, and teachers who were kind enough to spare their valuable time for this study.

Special recognition goes to Ms. Brenda J. Geber for her dedicated effort in transcribing original manuscripts to typed copy.

Finally, the author wishes to express special appreciation and gratitude to his wife, Crichton Jones, for her consistent encouragement.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	Statement of Problem	1
	Justification	2
	Procedure	3
	Evaluation Process	4
	Criteria	5
	Attitude Scaling	6
	Limitations of the Study	7
	Assumptions	9
	Summary	9
	Definition of Terms	11
	Attitude	11
	Attitude Object	11
	Clinical Cycle	11
	Morris L. Cogan Style Evaluation	12
	Known Group Process	13
	Likert Scale	13
	Morale	13
	State Evaluation (ORS 342.850)	13
	State Guidelines for Evaluation	14
	Systematic and Objective Analysis of Instruction	14
	Verbatim Transcript	14
II.	REVIEW OF LITERATURE	15
	Introduction	15
	Current Findings	16
	Experimental Model	19
	Attitude Measurement	22
	Limited Research	28
	Summary	29
III.	MATERIALS AND METHODS	31
	Purpose of the Study	31
	Outline of Procedure	31
	Collection of Items for Scale	32
	The Preliminary Response Scale	33

CHAPTER		PAGE
III.	Administering the Preliminary Response Scale . .	35
	Identification of Items for Scale	36
	Validity Study	36
	High Score on ATEP Scale	37
	Low Score on ATEP Scale	37
	Reliability Study	38
	Selection of School Districts	39
	Criteria for Selection of School District A . .	39
	Criteria for Selection of School Districts B, C, D	40
	Administration and Scoring	42
	Contacting the School District	43
	Agreements Made	43
	Selection of Sample and Administration of Scale . .	45
	Treatment of Data	46
	Summary	47
IV.	RESULTS AND DISCUSSION	50
	Collection of Statements	51
	Administering the Preliminary Response Scale . .	51
	Validity Study	52
	Reliability Study	53
	Selection of School Districts	54
	Sample	56
	Treatment of Data	59
	Primary Analysis	59
	Respondents Compared by Grade Level	62
	Comparison of Districts by Grade Levels	65
	Summary	67
	Grade Level Within Districts	68
	Item Analysis	71
	Positive Scale Items	73
	Neutral Scale Items	74
	Unsolicited Written Comments	76
	Summary	77
V.	SUMMARY AND IMPLICATIONS	79
	The ANOVA	79
	Item Analysis	81

CHAPTER	PAGE
V.	
Unsolicited Comments	83
Implications of the Study	84
Summary	87
BIBLIOGRAPHY	90
APPENDICES	94
A. Likert Statement Request	94
B. Sample of Likert Scale Items Developed by Investigator	95
C. Results of Likert Statement Request	97
D. Preliminary Response Scale	100
E. ATEP Scale	101
F. Known Group Identification	103
G. Results of the ANOVA	104
H. Item Analysis for All Districts Combined . .	105
I. Tally of Unsolicited Comments per Item by District	106
J. Unsolicited Comments by District	107

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. COMPARISON DATA FOR SELECTING SCHOOL DISTRICTS	41
2. CRITERIA FOR STATEMENTS OF CORRELATION	54
3. COMPARISON DATA FOR DISTRICTS (DEMOGRAPHICS)	55
4. SAMPLING DATA FOR DISTRICTS	58
5. MEAN SCORES ON ATEP SCALE FOR DISTRICTS A, B, C AND D	61
6. MEAN SCORES BY GRADE LEVEL FOR ALL DISTRICTS COMBINED	63
7. MEAN SCORES BY GRADE LEVEL FOR DISTRICTS A. HIGH SCHOOL B. JUNIOR HIGH C. ELEMENTARY	66
8. MEAN SCORES BY GRADE LEVEL WITHIN DISTRICTS	69
9. ITEM ANALYSIS FOR HIGH AND NEUTRAL SCORES	72

A MODEL FOR IDENTIFYING EVALUATION PROCEDURES THAT HAVE A POSITIVE INFLUENCE ON TEACHER ATTITUDE

CHAPTER I

INTRODUCTION

One of the most important tasks for a public school administrator is that of evaluating teachers. The major purpose of an evaluation should be to improve district services to students, staff, and community. Performance evaluation is an essential part of staff development and is mandated by state law in Oregon Revised Statute (ORS) 342.850. The evaluation process should create a better understanding between teacher and administrator because the evaluator can assist the teacher by identifying strengths that should be encouraged and concerns that should become goals for improvement. Evaluation also assists the administration and teacher to recognize the teacher's potential for advancement.

Statement of Problem

The evaluation process should be fair and conducted in a positive way to provide a sense of satisfaction to the person being evaluated. Based on that concept, the purpose of this study is to identify evaluation procedures that will have a positive influence on teacher attitude.

The hypothesis is stated in null form: There are no significant differences in the attitudes of teachers toward the evaluation process in districts that have a clinical cycle model for evaluation and in those districts that do not. The clinical cycle model for evaluation, developed by Maurice Cogan (1950), consists primarily of preconference, data collection through classroom visitation, and post conference. All phases emphasize the need for positive interpersonal relations between the teacher and evaluator.

Justification

ORS 342.850 establishes the fact that there will be an annual evaluation of each classroom teacher. The Oregon Fair Dismissal Law (ORS 342.850 through 342.930) states the criteria by which teachers may be terminated. In 1977, the State Department of Education developed a document called "Evaluation Guidelines for School Personnel" that outlined suggested procedures for the evaluation process and provided samples of the documents to be used. This model parallels the Cogan clinical cycle.

Senate bill 354, approved by the governor of Oregon on July 23, 1979, amends ORS 342.850. Essentially, this amendment outlines elements of the process to be followed in evaluating a teacher which are similar to the clinical cycle of supervision. The pattern is the same as that suggested by the guidelines for evaluation made available in 1977.

Because the current law was enacted in the summer of 1979, and the previous law did not specify the evaluation procedure, school districts in the state of Oregon varied in the processes they used to evaluate teachers. Some districts had implemented a process similar to the state requirements as long as ten years ago. Others had no defined process, and the evaluator was free to complete the state evaluation form in any manner.

The fact that the law requires that teachers be evaluated and that it specifies some elements of the process strongly suggests that a study which would provide empirical data as to the merits of this process in influencing attitude could be valuable. If the process does create a positive attitude toward evaluation, the teachers are more likely to implement suggestions for improvement which should enhance classroom instruction.

Procedure

In order to test this hypothesis, the following procedure was used:

- A. Establish the criteria for the experimental group based on the clinical cycle of supervision model.
- B. Identify the public school districts and rank order them based on the established criteria (Table 1, p. 41).
- C. Develop an attitude scale following the Rensis Likert method.

- D. Use the analysis of variance to determine if there is a difference in the attitudes between the selected districts.
- E. Review item analysis to identify elements of the process that should be retained in order to have a positive effect on attitude.
- F. Outline suggestions for an evaluation model that is most likely to produce a positive attitude in the teachers being evaluated.

Evaluation Process

For the purpose of this investigation, the procedure used to evaluate teachers is based on the Systematic and Objective Analysis of Instruction model of supervision developed by Drs. James Hale and Allan Spanger, 1972. This model of evaluation is based on the Principles of Evaluation and Supervision developed by Dr. Morris L. Cogan. The essentials of the process are referred to as the clinical cycle of supervision and contain the following elements:

<u>Cycle</u>	<u>Principles</u>
Planning	Develop Shared Understanding
	Exchange Full Information
	Identify Learner Goals
Observation	Make Provisional Try
	Obtain Performance Data

<u>Cycle</u>	<u>Principles</u>
Analysis	Examine Performance Data Determine and/or Hypothesize Effect of Data
Strategy	Order Possible Behavioral Changes (Successive Approximation)
Conferences	Provide Feedback on Performance and Effect of Data Specify Desired Change

For the purposes of this study, though it is primarily a training program developed to teach evaluators the skills identified by Cogan, the Systematic and Objective Analysis of Instruction model and the clinical cycle of supervision will be used interchangeably.

Criteria

The criteria for selecting school districts for the study will be: an evaluation model based on the clinical cycle model of supervision; an information package outlining the procedure, including sample documents, dates, to be provided for all teachers; evaluators trained in the Systematic and Objective Analysis of Instruction procedures; and, finally that the process of evaluation and supervision are the responsibility of separate individuals.

All of the school districts selected for this research will be rank ordered based on the aforementioned criteria. The district

representing the control group for the experiment will be selected because it has not met any of the criteria. Other districts will be ranked according to the combination of criteria met (Table 1, p.41).

Attitude Scaling

This study recognizes the controversy facing investigators when it is necessary to define attitude. Oppenheim (1966) suggests that there are three components to an attitude.

Attitudes are reinforced by beliefs (the cognitive component) and often attract strong feelings (the emotional component) that will lead to particular forms of behavior (the action tendency component) (Oppenheim, 1966, p. 106).

He summarizes this by suggesting that,

Most definitions seem to agree that an attitude is a state of readiness, a tendency to act or react in a certain manner when confronted with certain stimuli (Oppenheim, 1966, p. 105).

This idea is reinforced by Allport (1935) who states that attitude is

a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related (Allport, 1935, p. 8).

Based on these definitions, the attitude object for this study is the attitude of the teacher toward the process used in evaluating. It is apparent that there are other attitude objects that will affect the results of the analysis. Blocker and Richardson (1973) suggest that interpersonal relationships between the evaluator and the teacher are

the most important factor affecting morale. Perhaps the most important factor would be the nature of the evaluation, whether the evaluation is for the improvement of instruction or the dismissal of the teacher. Both the Cogan theories and the principles of the Systematic and Objective Analysis of Instruction model deal with the affective component of evaluation, and, therefore, for the purpose of this study, the investigation will be limited to the analysis of the teacher attitude toward the evaluation process.

The Rensis Likert attitude scaling method is the most appropriate method for developing a scale. This process and the criteria for selecting items for the scale are outlined in detail in Chapter III.

Limitations of the Study

The Rensis Likert scaling procedure forces the investigator to make an assumption that the data collected on an attitude scale can be dealt with as Equidistant Interval data which can be processed by the analysis of variance (ANOVA) technique. Some statisticians would suggest that the data in this study is ordinal in nature and not appropriate for this type of analysis. Oppenheim (1966) supports the Likert scaling process but does suggest the difficulty by stating,

The most advanced scaling techniques and the most error-free procedures will not produce an attitude scale worth having unless there has first been a good deal of careful thought and exploration, and unless the ingredients of the

scale (the attitude statements) have been written and rewritten with the necessary care (Oppenheim, 1966, p. 105).

Because the evaluation process is regulated by state law and identified in contracts between school districts and the teachers' bargaining unit, the administrators and teachers will be reluctant to provide some of the information that will be requested. This suggests that there will be difficulty in identifying the actual procedures used in evaluation and that every effort must be made to assure the respondents of their anonymity.

There is a limited amount of research that focuses on the attitudes of teachers. Most of the research focuses on the interpersonal relationships between evaluators and teachers, the effectiveness of particular data collecting devices, and student achievement. In order to identify research related to student achievement, it was necessary to substitute the term morale for attitude while doing the computer search. Attitude and morale are not synonymous terms. Bentley and Rempel (1970) define morale in terms of beliefs and feelings which are considered by some as the affective component of attitudes. By making this substitution in terms, it was possible to identify some research indicating a correlation between high teacher morale and improved student achievement.

Assumptions

Public school administrators and teachers will be interested in the results of the study and will participate to the best of their abilities. The reason for their interest in the research is the fact that Oregon state Senate Bill 354 mandates that they must develop a process for evaluation which is similar to the clinical cycle and all teachers are evaluated each year.

The study is based on the assumption that the attitude object is the evaluation process, the Rensis Likert attitude scaling process, and the statistical tool called the analysis of variance. These components will provide the framework for successful research. The results of the research can be verified by utilizing the appropriate statistical model. The preliminary investigations suggest that the evaluation procedure does affect the teacher's attitude.

Summary

Public school teachers must be evaluated in order for school districts to remain in compliance with the Oregon Revised Statutes. ORS 342.850 (1979) has outlined the process to be used. This process is similar to the clinical cycle of supervision model based on the theories of Dr. Morris Cogan. In this study, when reference is made to the process of evaluation, the experimental group will be evaluated

by a process similar to the Cogan clinical cycle. The other districts will be selected according to the number of components of this process in use.

The thesis of this study is that an evaluation process based on the clinical cycle will have a positive effect on the teachers' attitudes. The attitude of the teachers will be measured by an attitudinal scale developed using the Rensis Likert method of constructing a scale which provides for reliability and validity studies.

Careful consideration must be given to the limitations of this study when analyzing the data. Attitudinal studies are difficult to conduct because of the various factors that can affect attitude. These factors have been identified in this chapter and should be considered when drawing conclusions based on the data.

Four districts have been selected for this study. One district meets all of the criteria for the experimental group, and the control district does not meet any of the criteria. Restated, this criteria is: a process of evaluation similar to the clinical cycle, an evaluation package provided for all teachers, training for evaluators in SOAI procedures, and evaluation and supervision are the responsibilities of separate individuals. The random sample will be approximately 25 percent of the districts' teachers, and all judgements will be based on the .05 level of significance with the Analysis of Variance being the primary statistical tool.

Based on results of the statistical analysis, the evaluation procedures that have a positive influence on teacher attitude will be identified.

Definition of Terms

Attitude

A mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related.

Gordon W. Allport, Attitude: Theory and Measurement.

Attitude Object

A particular object, value, concept, or goal with which an individual has a specific relationship.

Clinical Cycle

The essential elements of the clinical cycle are:

- a. Planning
 - (1) Develop Shared Understanding
 - (2) Exchange Full Information
 - (3) Identify Learner Goals

- b. Observation
 - (1) Make Provisional Try
 - (2) Obtain Performance Data
- c. Analysis
 - (1) Examine Performance Data
 - (2) Determine and/or Hypothesize Effect of Data
- d. Strategy
 - (1) Order Possible Behavioral Changes.
(Successive Approximation)
- e. Conference
 - (1) Provide Feedback on Performance and Effect Data
 - (2) Specify Desired Change

Morris L. Cogan Style Evaluation

The Cogan style of supervision contains the following elements:

- 1. Separation of Evaluation and Supervision
- 2. Communication Skills
- 3. Helper/Helpee Relationships
- 4. Data Collection and Analysis
- 5. Exchange of Information
- 6. Self Analysis

Known Group Process

A method for determining the validity of an attitude scale by correlating the value of the scale with groups that have an identifiable attitude. This is referred to as Criterion-Related Validity by Downie and Heath (1974).

Likert Scale

A scale for measuring attitudes based on the research of Rensis Likert. The scale items generally have a five point range.

1	2	3	4	5
Strongly	Disagree	Neutral	Agree	Strongly
Disagree				Agree

Morale

The mental and emotional condition of an individual or group with regard to the function or tasks at hand.

State Evaluation (ORS 342.850)

Teacher evaluation form personnel file content, (1) The district superintendent of every school district, including superintendents of intermediate education districts, shall cause to have made at least annually an evaluation of performance for each teacher employed by the

district in order to allow the teacher's development and growth in the teaching profession.

State Guidelines for Evaluation

The document outlines some procedures that parallel the SOAI cycle. These are only guidelines and are not state policy.

Systematic and Objective Analysis of Instruction (SOAI)

This is a system developed by James R. Hale and R. Allen Spanger based on the Cogan model. The emphasis is placed on combining evaluation and supervision in the least threatening procedure. The model contains the clinical cycle which consists of preconference, data collecting, analysis, and post conference.

Verbatim Transcript

A hand written transcript made while observing a teacher conducting a class. The transcript includes the time, the verbal interaction, and anecdotal comments.

CHAPTER II

REVIEW OF LITERATURE

Introduction

Teacher evaluation and supervision has been a subject of research for many years. Beecher (1901) discussed the reasons for evaluation and the process to be used in evaluation. Beecher indicates that evaluation of teachers should be for the improvement of instruction. He further states that evaluation and supervision are separate but essential aspects of an evaluation model. Evaluation should have its emphasis on making value judgements about the effectiveness of the teacher's work, and supervision should emphasize the diagnostic and prescriptive aspects that improve the teacher's performance in the classroom. This text used the term "appraisal of teaching" rather than evaluation of teachers and makes the point that the evaluation process is essential for the improvement of teaching effectiveness, and is measured by student success.

The interest in evaluation has been increasing because of the following factors:

1. The introduction of competency based education into Oregon schools.
2. Fair dismissal procedures identified by state law

(ORS 342.825)

3. Tendency for faculties to retain older and more experienced teachers because there are fewer other positions open.
4. Decrease in student population.

Since these factors have come into existence, teachers have been resisting dismissal attempts by utilizing the state law and any alleged breakdown in the evaluation process. For this reason, school district personnel have been currently studying and refining the process used in evaluating teachers.

Current Findings

Research, such as Biddle's (1964) Contemporary Research in Teacher Effectiveness, McNolte's (1968) How to Tell Which Teacher to Keep and Which Teacher to Lay Off or Barr's (1961) Wisconsin's Studies of the Measurement and Prediction of Teacher Effectiveness, all emphasize essentially the same concepts. These concepts are:

1. Separation of supervision and evaluation.
2. The use of objective measuring devices such as checklists.
3. The use of other sources of information on teacher performance such as student ratings.
4. The establishment of the primary purpose of evaluation as the improvement of instruction.

The major apparent variation in the literature is the sophistication of the devices used in the analysis of the teaching act. Devices such as the Purdue rating scale for instruction or Flanders Interactions Analysis system have validity and reliability studies completed.

The fact that teachers are more likely to initiate grievance proceedings after receiving a dismissal notice has forced school districts to develop procedures that meet the needs of the local district and, at the same time, meet state guidelines. (Example: District A packages include a philosophy statement, job description, procedural steps, timelines for completion, and sources of remediation available to teachers.)

The use of student evaluation, peer observation, and participatory management techniques are some new concepts appearing in the literature. For example, Page (1974), Student Evaluation of Teaching the American Experience, makes a strong case for student evaluation at the college level and suggests the same procedure is applicable to the public school sector.

McGee (1977), Clinical Supervision and Teacher Anxiety, has determined that peer observation and team planning will reduce the anxieties of teachers being observed.

Van Hoven (1974) suggests that peer involvement in job target setting and assessment can be an effective means of humanizing the assessment process, resulting in improved instructional performance

and higher staff morale.

Another theme in the current literature is participatory management. Participatory management, in public schools, is interpreted as the involvement of parents, students, teachers, and evaluators in any process that is used to determine the effectiveness of teachers. Jon Wiles (1979) in his Educational Innovations Probability Chart indicates that the most successful plan for change is one that involves the individuals affected. Why Process, by Littrell (1968), reinforces the theme by listing nine reasons for involving the people affected.

1. People have the right to participate in decisions which have an effect on their well-being.
2. Participating democracy is the superior method of conducting community affairs.
3. People have the right to strive to create that environment which they desire.
4. People have the right to reject an externally imposed environment.
5. Maximizing human interaction in a community will increase the potential for human development.
6. Implicit within a process of interaction is an ever widening concept of "community."
7. Every discipline and/or profession is potentially a contribution to a community's development process.

8. Motivation is created in man by association with his environment.
9. Community development is "interested" in developing the ability of human beings to meet and deal with their environment.

Current literature, therefore, supports the ideas that evaluation is for the improvement of instruction and that a systematic and objective approach will reduce anxiety and improve staff morale. The details of these methods are generally left to the districts.

Experimental Model

The evaluation model for the experimental group is based on the components of the Systematic and Objective Analysis of Instruction model by Hale/Spanger (1972). The components of the experimental group's evaluation procedures are as follows:

- a. Pre Conference
 - (1) Develop Shared Understanding
 - (2) Exchange Full Information
 - (3) Identify Learner Goals
 - (4) Discuss Evaluation Device to be Used
- b. Observation
 - (1) Class Visitation
 - (2) Obtain Performance Data

c. Analysis

(1) Examine Data

(2) Determine and/or Hypothesize Effect in Relation to
Specified Goals

d. Strategy

(1) Order Possible Behavioral Changes (Successive
Approximation)

e. Post Conference

(1) Provide Feedback on Performance and Effect Data

(2) Specify Desired Change

The Thomas (1972) Technical Report No. 9 on Systematic and Objective Analysis of Instruction is an analysis of his research of the effectiveness of the Systematic and Objective Analysis of Instruction training procedures for evaluators. The finding indicates that the evaluators trained in this system reach the goals as outlined in the manual. The goals of the system are that a person so trained will be able to:

1. Develop Shared Understanding

Exchange Full Information

Identify Learner Goals

2. Obtain Performance Data

3. Examine Performance Data

Determine and/or Hypothesize Effect of Data

4. Order Possible Behavioral Changes (Successive Approximation)
5. Provide Feedback on Performance and Effect Data
Specify Desired Change

An inference may be made that any evaluator using the Systematic and Objective Analysis of Instruction system is successful in improving instruction there by creating a more positive attitude toward evaluation. The Systematic and Objective Analysis of Instruction system is based on the tenets of the clinical cycle of supervision. Sullivan (1980) has investigated the effectiveness of the clinical cycle of supervision model. The research focuses on the use of the clinical cycle of supervision model. The research focuses on the use of the model for supervision of classroom teachers and the supervision of cadet teachers, rather than as a process for completing the annual teacher's evaluation. According to Sullivan,

While findings and indications can be summarized, no general conclusions can be drawn from the available research: the amount and quality of research is insufficient to support generalizations concerning the model (1980, p. 22-23).

The indications are that the clinical cycle tenets and practices are compatible with the desires of teachers and administrators.

This subjective analysis may be the reason why the model has been widely adopted as functional despite the fact that there is little objective data to support its value.

The state of Oregon has chosen to mandate the inclusion of the elements of the clinical cycle of supervision in the yearly evaluation procedures (ORS 342.835). This will assure widespread use of the clinical cycle.

Attitude Measurement

All of the aforementioned information relates primarily to the process to be used in evaluating teachers. The focus of this research is on measuring the attitude of teachers toward the process as a guideline for evaluating the value of the evaluation process.

Allport (1966) defines attitude as a mental and neural state of readiness, organized through experience, exerting a direct or dynamic influence upon an individual's response to all objects and situations with which it is related. Doob (1967) suggested another definition: "An implicit drive producing response considered socially significant in the individual's society" (p. 43).

Stearns (1964) and Wagoner (EDRS) have done their research in the area of teacher attitude toward the process of evaluation. One report is based on the analysis of Head Start teachers and the other on community college teachers.

Stearns (1964), in his report, Maximizing the Value of Evaluation for the Head Start Teacher, analyzed systematic feedback and its effect

on the development of a more favorable attitude toward evaluation. The teachers receiving feedback were more positive in their attitudes as shown by the TATE (Teacher Attitude Toward Evaluation) scores. Also, the report established that the feedback process produced increased rapport and understanding between teacher and evaluators.

The other citation related to attitude measurement is Teacher Attitude Toward Evaluation by Wagoner and Ottonion (EDRS). Ideas supported by this research are: better than average teachers had a more positive attitude toward evaluation, attitudes toward evaluation are predictable, and feelings of threat are not a necessary consequence of evaluation.

Pollack (1976), in a report for the North Carolina State Department, completed a study of full-time students, teachers and administrators drawn from the 57 North Carolina Technical Institutes and community colleges. It was hypothesized that the low level of attitude for teachers was due to the evaluation systems currently in use. The findings indicated a positive correlation between the process of evaluation and attitude.

These studies support the idea that positive attitudes toward evaluation procedures will be reinforced if the philosophical base for the evaluation process is improvement of instruction. The research also establishes the idea that the process followed does have an effect on teacher attitude.

Attitudinal research that focus on teachers and the evaluation process are limited. If the concept of morale is reviewed, there are numerous studies that identify factors that affect the morale of teachers. As stated in Chapter 1 (p. 13), the terms morale and attitude are not synonymous, but do share affective components. Since Bentley and Rempel (1970) have defined morale in a very broad sense not only considering the job aspect but the whole environment of the individual, a review of that type of research is warranted for this study.

Bentley and Rempel (1970) define morale as

the extent to which an individual's needs are satisfied, and the extent to which the individual perceives satisfaction as stemming from the total job situation. High morale is evident when there is interest in and enthusiasm for the job. What is important in morale is what the person believes and feels, rather than the conditions that may exist as perceived by others (p. 1).

Redefer (1959) found a close relationship between morale, the quality of education present in the school and the evaluation of the teachers by the administrators. This study was the only one that specifically included evaluation as a morale factor. Gragg (1955) found, by asking teachers to list factors affecting their morale, that the most frequently mentioned item that contributed to high morale was confidence in the leadership skills of the principals and other administrators. Other morale factors, such as teacher involvement in development of school policies that affect them, sufficient support from supervisory personnel and a fairly distributed work load were identified by Hand

(1948). Both Ross (1960) and Barber (1970) found that the single most important factor determining the climate or morale is the principal or other supervisory personnel.

A slightly different approach was taken by Hood (1965) who measured the congruence between a teacher's relationship with other teachers and the principal and the relationship's effect on their own morale. The results suggest (as indicated in other studies) the relationship between the principal and teacher was more important than the relationship between one teacher and another as far as the morale factor is concerned.

Napier (1966) identified 13 items associated with high morale factors in teachers. These are:

1. The administrator's understanding and appreciation of the teacher as an individual.
2. The confidence the teacher has in the administrator's professional competency.
3. The support the teacher receives from the administration regarding discipline problems.
4. Teacher participation in the formulation of policies that affect them.
5. Adequate facilities and equipment.
6. Adequate teaching supplies.
7. Teaching assignments which are commensurable with training.

8. Fair and equitable distribution of extracurricular assignments.
9. Professional training provided through the inservice program.
10. Job security.
11. An adequate policy for leaves of absence.
12. A fair and equitable distribution of the teaching load.
13. Salaries that are comparable with professions requiring equal training.

Gross and Heriott (1965) conducted a study that was to determine the organization and performance factors of the administrators that affect teacher morale. An instrument was used called EPL (Executive Professional Leadership) device that measured the effort of the principal to conform to the definition of his role that should stress his obligation to improve the quality of staff performance. There was a positive relationship between the EPL scores and the following factors:

1. Display a sense of pride in the school.
2. Enjoy working in the school.
3. Sense of loyalty to the school.
4. Work cooperatively with fellow workers.
5. Accept educational philosophy.
6. Respect judgement of administrator of the school.

Blocker and Richardson (1963) identified a similar relationship, that the administrator was the most important factor affecting teacher morale. Their study suggested another dimension, that of evaluating the administrator's performance by measuring teacher morale.

The research that investigated teacher morale and the factors responsible has a commonality that suggests the interpersonal relationship between principal and teacher is the most important factor. Only one, Redfer (1959), made direct reference to the fact that the evaluation process was also a factor affecting morale.

For the purposes of this research, the assumption was made that if teachers have a positive attitude toward the evaluation process, improvement of instruction will be the result. The improvement of instruction should result because a teacher with a positive attitude would view the results of the evaluation as a way to improve performance (thus attaining change for the better in the classroom). There was no research identified in the review of related literature that has addressed the issue of attitude and its effects on student achievement. However, there is some research that suggests high teacher morale does improve student achievement. Gross and Herriott (1965) found that there was a correlation between high morale and high productivity in students. Similarly, Anderson (1953), in a study of 20 Iowa secondary schools, found a similar correlation. In a study by Koura (1963), twelve secondary public schools in Dearborn, Michigan, found that student

achievement increased under teachers with high morale and decreased under teachers with low morale.

The research investigating aspects of teacher morale provides limited evidence that the evaluation process has any effect on morale but suggests that the administrator relationship with the teacher is probably the one most significant factor. The potential for improving student achievement by creating an environment that produces a high morale does seem to be verified by these studies.

Limited Research

Materials that relate to the purpose of this study, to identify an evaluation process that will produce a positive attitude in teachers, is limited. None of the studies with a focus on attitude measurement were based on the public school sector. The first Eric Search completed in the summer of 1978 utilized key words such as teacher evaluation, teacher attitude, teacher alienation, and teacher morale. Sixty-eight citations were identified, and only one related to this project.

The second Eric Search (summer of 1979) expanded the parameters by not restricting the search to the public school evaluation process. The word "anxiety" was included and the words "secondary and elementary schools" were removed. Forty-six citations were identified and three were applicable to the study. Even the inclusion of

citations identified by the keyword "morale" provided only one study that measured the effect of the evaluation process on morale.

Summary

The analysis of the research indicates that the concentration of effort has been in selecting the most effective way to analyze the teaching act, not on the effect the evaluation process may have on teacher attitude. The following are ideas supported by the research that are relevant to this study.

1. Evaluation should be for the improvement of instruction.
2. Supervision and evaluation are separate functions.
3. The evaluator should be as objective as possible.
4. Devices to collect data are recommended. Example:
Flander's Interaction Analysis coding sheet.
5. Teachers feel best about the process if they participate in the development of the process and have a complete understanding of the process.
6. Interpersonal relationships (between the evaluator and teachers) are of prime importance.
7. Studies that investigated the factors affecting the morale of teachers indicate that the principal is a prime factor and that student achievement is affected in a positive way.

There is limited evidence that measurement of attitude has been a criteria for the selection of the evaluation process. The research indicates that a positive attitude is more desirable than a negative one. Therefore, a research model that establishes the process of evaluation with the most positive effect on teacher attitude does merit the time and effort of a systematic study.

CHAPTER III

MATERIALS AND METHODS

Purpose of the Study

The purpose of the study is to determine if there is any significant difference in the attitudes of teachers in districts that use staff evaluation procedures based on the clinical cycle model of evaluation and in those districts that do not. The Rensis Likert scaling procedure was used to develop a twenty item scale. The attitude object for the scale was the teacher's attitude toward the evaluation procedure.

Outline of Procedure

The following is an outline of the procedure to be followed:

1. Collect sample questions about evaluation procedures from teachers in districts similar to the districts to be sampled.
2. Organize these questions into a Likert style based on criteria for editing the statements (p. 34).
3. Correlate the item score on the scale with the total scores to determine how well the test questions correlate with the total test score.
4. Rank order the items and choose the top twenty for the final scale.

5. Establish the validity of the scale by using the Known Group Process.
6. Test the scale for reliability by using the split half method (Spearman Brown).
7. Select the school districts for the study based on the criteria (Table 1, p. 41) for the experimental group and the control group. Rank order the other districts based on how close they match this same criteria.
8. Sample by using the Random Sampling technique with a Sample approximately twenty-five percent of total population.
9. Use the one way ANOVA to determine the acceptance or rejection of the null hypothesis based on the .05 level of significance.
10. Identify the relevant parts of the evaluation process by analyzing data from the item analysis.

Collection of Items for Scale

In order to collect the items for the scale, a form was developed. The form (Appendix A) requested that teachers write an idea or a sample question that focused on the evaluation process. An initial test of the device indicated that teachers thought the sample statement was asking for them to mark their response on the one to five scale rather

than write an additional statement. The problem was corrected by writing "SAMPLE" in block letters across the sample statement. This corrected model did elicit responses that were incorporated in the completed scale.

The investigator generated a list of items for use in the scale (Appendix B). This list was compared with the responses provided by teachers and used to make the initial selection of items for the scale.

The Preliminary Response Scale

This scale was called the Preliminary Response Scale and contained sixty-six items. The scale was designed to be self administered. All instructions and a disclaimer--the resulting data will not be identified by individual or building--was written at the top of the first page (Appendix C).

All of the statements were edited based on the criteria identified by Wang (1932), Thurstone and Chave (1929), Likert (1932), Bird (1940), and Kilpatrick (1948).

The criteria as stated is:

Criteria¹

1. The statement should express a desired behavior and not a statement of fact. (A person with the opposite attitude might agree with a statement of fact.)
2. Each statement should be clear, concise, straight-forward, and in the vocabulary of the target group.
3. Statements should not use double negatives or include more than one idea.
4. Select statements that cover the entire continuum one through five.
5. Select statements so that one half ranges toward the left hand, one through three of the continuum, and one half covers the continuum three through five.
6. Distribute these statements in a random manner through the instrument.
7. For tabulation purposes, select numbers one through five with three being the undecided position.
8. One would be assigned some designation such as heartily disagree with five representing the opposite attitude.

¹Paraphrase of criteria suggested established by Wang (1932), Thurstone and Chave (1929), Likert (1932), Bird (1940), and Edwards and Kilpatrick (1948).

9. Check the statements to correct for any reversal of responses to the one through five designations.
10. Avoid statements that refer to the past rather than to the present.
11. Avoid statements that may be interpreted in more than one way.

The items were edited and improved after three limited samples were taken. At this point, these samples indicated that there was no discernible problem with the instructions or with understanding the purpose of the instrument.

Administering the Preliminary Response Scale

The preliminary response scale was administered to a random sample of teachers (N=105). The Pearson correlation coefficient was used to establish the correlation between the score on each question and the average of the total score for each respondent. The average total score was selected rather than the actual score on the scale in order to compensate for non-response to an item. (This alternative was recommended by Ms. Susan Maresh, OSU Computer Center Consultant.)

Identification of Items for Scale

Each item on the scale was identified with the appropriate correlation value and manually ranked from highest to lowest. Starting with the item with the highest correlation value, the first twenty items were identified as suitable for use.

Validity Study

The Known Group Identification checklist was developed, based on the criteria for the four populations to be sampled in the study (Appendix D, p. 100). The respondents were instructed to check one of three items labeled A, B, or C. Item A represented the district evaluation procedures that met all the criteria for the experimental group (Table 1, p. 41). Item B represented the control district criteria which essentially means the district has no prescribed procedure (Table 1, p. 41). Item C would indicate a district used some combination of A and B in the evaluation process.

An evaluation statement was added to this criteria list which required the respondent to check a continuous scale ranging in value from 1 to 5 points.

My feelings about the procedure used in evaluating me are:

1	2	3	4	5
Negative		Neutral		Very Positive

The Known Groups were identified in the following way:

High Score on ATEP Scale

1. Respondents that checked A or C and marked the scale four or five.
2. Respondents that checked B and marked the scale one, two, or three.

Low Score on ATEP Scale

1. Respondents that checked A or C and marked the scale one, two, or three.
2. Respondents that checked B and marked the scale four or five.

A random sample of high school, junior high, and elementary teachers was taken. This sample was manually sorted by using the aforementioned criteria (known group identification checklist) into two groups. The hypothesis, based on the known group identification checklist criteria, is that one group will have high scores on the ATEP scale, and one group will have low scores.

The statistical consultant (Susan Maresh, OSU Computer Center) recommended that the Pearson Correlation product moment be used to determine the relationship between the scores on the ATEP scale and the ratings on the known group identification continuum. In order to

determine the acceptance or rejection of this hypothesis at the .05 level of significance, the "P" value was computed. The value of "P" is the probability that the value established is equal to the level of significance.

This process completed, the initial analysis indicated that the groups were successfully identified, and the scale's validity is adequate for this study.

Reliability Study

The split half method of determining reliability was selected because one test is needed for the computation of the reliability coefficient. The correlation factor for the ATEP scale was computed by correlating the odd numbered items with the even numbered items. According to Downie and Heath (1974), a Pearson Product-Moment Correlating coefficient is computed between the two sets of scores. A reliability coefficient of this type is often called a coefficient of internal consistency.

The reliability of a test is directly related to the length of the test (Downie and Heath, 1974). The Spearman-Brown formula is used to correct for this problem.

$$r_{sB} = \frac{2r_{xy}}{1+r_{xy}}$$

Where r_{sB} = The reliability of the original test

r_{xy} = The reliability coefficient obtained by correlating the scores on the odd items with the scores of the even items.

For the purposes of this research, the averages of the odd/even responses were tabulated to compensate for the non-response to an item. This process was completed successfully and the initial analysis indicated an acceptable reliability quotient for the study.

Selection of School Districts

The following criteria were developed by the researcher to identify the district that will be referred to as District A. This district represents the experimental component of the study since all of the items listed in the criteria have been put into effect. These items are:

Criteria for Selection of School District A

A. Evaluation Package which contains:

1. Rationale with emphasis on improvement of instruction.
2. Process specified which is similar to the clinical cycle pattern based on S.O.A.I. model.
3. Job descriptions for those covered by the evaluation process.

4. Goals or standards for the teacher's performance in the classroom.
 5. Guidelines for completing state teacher's performance evaluation form.
 6. Timeline for completion of steps in process.
- B. Evaluation and supervision responsibilities invested in separate individuals, either on a building level or district level.
- C. Evaluators and supervisors are trained in the Systematic and Objective Analysis Method.
- D. Optional miscellaneous information (specific data collecting devices, example: checklists, etc.).

Criteria for Selection of School Districts B, C, and D

The district that will represent the control component for the research will not contain any of the elements present in the criteria above and will be labeled District D. Two other districts will be selected that have some variation or combination of the criteria identified for the experimental group and control group. These districts will be referred to as Districts B and C.

State law governing evaluation of teachers (ORS 342.825) applies the same criteria for evaluation to all districts regardless of student

Table 1. COMPARISON DATA FOR SELECTING SCHOOL DISTRICTS (Based on Evaluation Process)

Type of Evaluation Process	Clinical cycle prescribed for process of Evaluation	No process prescribed (Clinical Supervision Implied)	Clinical cycle prescribed fall, 1979. Previously no process prescribed	No process prescribed
Materials provided Teachers	Evaluation packet Philosophy Job description Process Explained Dates for completion Copies of documents	Evaluation packet Philosophy Job description Level of achievement No process specified Dates for completion Copies of documents	Evaluation Package introduced 1979 No district package before this date	No packet
Training provided Evaluators	District provided training in clinical cycle thru S. O. A. I. training (Systematic and Objective Analysis of Instruction)	Training left to individual evaluators	Some evaluators trained in ITIPS* *Instructional Theory into Practice	No district prescribed training. Some evaluators trained in Cogan cycle
Person responsible for Assisting Teachers Improve Instruction	Teacher development specialist provided for teacher's assistance (no responsibility for dismissal procedures)	Responsibility assigned to personnel director only active in dismissal procedures	None with specific responsibility	No one with specific responsibility
	DISTRICT A	DISTRICT B	DISTRICT C	DISTRICT D

population, geographic location, or organizational pattern. Therefore, for the purposes of this study, the districts were chosen primarily for their relationship to the criteria identified in Table 1 (p. 41). In addition to the fact that the districts were selected based on their relationship to the criteria outlining the evaluation process, the demographic data was identified and outlined (Table 3, p. 55). When the data is analyzed, the demographic material will be reviewed as to its potential effect on the results.

Administration and Scoring

The ATEP scale consists of twenty statements to which the teacher responds with one of five choices: Strongly Disagree (1), Neutral (3), and Strongly Agree (5), with two and four being used to complete the continuum. The score is determined by summing up the numerical values of the twenty items. The maximum score would be one hundred if the respondent circled five for each item and a minimum score would be twenty if the respondent circled the value of one for each of the twenty items. The higher the total score, the more positive the attitude toward the evaluation process, and the lower the score, the more negative the attitude.

The directions for completing the ATEP scale are written on the top of the scale and are self-explanatory. There is no time limit, and it takes approximately five minutes to complete the scale.

Contacting the School District

Phone contact was made with the superintendents of six districts that generally fit the criteria established for the study. In each case, an interview was granted and the superintendents reviewed the purpose of the project, the materials to be used, and the effect the project would have on the district.

Two superintendents, after reviewing the scale with the administrative staff, felt that the questionnaire might interfere with the collective bargaining (Teachers' Union vs. School Board) that was in process. One district was involved in the same bargaining unit but agreed to the use of the questionnaire on the basis that there was the possibility of obtaining useful information that would be of value to the district.

Four superintendents, after the exchange of information with the other administrative levels in the districts, accepted the study as valuable and agreed to cooperate.

Agreements Made

All four superintendents that agreed to the use of the scale in their districts required that the district not be identified in the research and that the scale not be coded in any way except by grade level (high school, junior high, elementary). Information such as sex, tenure, non-tenure, identified as having a need for improvement, or any identification of

actual practices followed by the evaluators was viewed as an imposition on the teachers and a risk of developing data detrimental to the school district.

For the purposes of the study, since all teachers by law (ORS 342.825) are technically treated the same, the conditions were accepted by the investigator. (All staff members are covered by the same evaluation procedure, and the evaluation document provided by the state is the same for all. The only variation in the process for new teachers and tenured teachers is the number of formal evaluations. The districts have a general procedure of evaluating new teachers two or more times and the tenured teachers once per year.)

Another stipulation agreed to was the fact that there would be no follow up on teachers that did not respond to the ATEP scale in order to guarantee their anonymity. This was the most serious limitation imposed on the study. It created the problem of sample bias due to the fact that the return would be based on a voluntary sample. In order to continue with the project, this limitation was accepted.

The next issue settled was the process for delivering the ATEP scale to the teachers and the process to be used for the return of the scale.

Three of the districts agreed to use the school mail to get the scale to the teachers. A self addressed stamped envelope was used to return the scale. One district allowed the use of the school mail both

for delivering the scale to the teacher and for returning the completed scale.

Negotiations with the superintendents concluded with the agreement that they would write a letter stating that permission was granted for the scale to be used in the district. In addition, it was agreed that each package of scales would be addressed to the building administrator and accompanied by a letter explaining the purpose of the scale and listing both the name and phone of the superintendent and the investigator if additional information was needed.

Selection of Sample and Administration of Scale

A sample (all academic areas, grade levels, counselors, special education, and media specialists) was selected by using the Random Digit Tables and the district rosters provided by the superintendents. Due to the limitation resulting from not allowing follow up on returns, a random sample of approximately forty percent of each district was selected. The teachers were identified by name, grade level, and building. This information was placed on a memorandum to the teacher which explained the purpose of the study, stated that it would be anonymous, and suggested a date for return. Each memorandum was modified slightly to fit the district.

A second memorandum was addressed to the building administrator. This document contained the following information:

1. The fact that the scale was approved for use.
2. What to do with the scale if the teacher named is no longer employed.
3. How to obtain further information.

The scale with the memorandum (cover letter) to the teacher was placed in the self addressed envelope and all of the scales intended for delivery to separate buildings were packaged in one large manila envelope. The memorandum to the administrator was attached to the outside. All of the packages were delivered to the secretaries of the superintendents in the appropriate districts.

Treatment of Data

An ANOVA or analysis of variance was selected as the means of determining the acceptance or rejection of the null hypothesis. The Newman-Kells procedure and the Least Significant Difference test will be used to identify where the difference lies, if the ANOVA rejects the null hypothesis. This will be based on the .05 level of significance.

The mean scores on the ATEP (Attitude Toward Evaluation Process) scale were identified in the following pattern:

1. Mean score by district (primary data to be compared to determine if hypothesis is to be accepted or rejected).

2. Mean score by grade level including all districts.
3. Mean score for each district by grade level.

These combinations of mean scores were analyzed for patterns that support the results found when comparing the mean score by districts.

In order to identify elements of the process of evaluation that have the most positive effect on attitude, an item analysis was completed. The item analysis was done in the following manner:

1. An item analysis by district including all grade levels.
2. An item analysis combining all districts including all grade levels.
3. An item analysis of all districts combined by grade level.
4. An item analysis for each district by grade level.

(The grade level designations were high school, junior high, and elementary.)

Based on consultations with the O.S.U. Computer Consultant, the mean and the mode are selected as most informative data for comparison of districts and the identification of elements in the evaluation process.

Summary

The ATEP scale (Attitude Toward Evaluation Process), consisting of twenty items designed to measure attitude toward evaluation, was

completed by following a process: Identification of items for scale, a statistical process for elimination of items, a validity study, and reliability study. This process was outlined by Oppenheim (1966) and Rensis Likert (1932) and was selected as appropriate for the study.

Six districts were investigated as meeting the needs of this study. A criteria was established to identify the experimental group and the control group. The criteria included items such as the type of information provided teachers, the training provided the evaluators, and the identification of a process for evaluation. Consideration was given to the demographic data such as size and geographic location of districts. Two districts did not wish to participate because they were involved in negotiations. As a result, four districts were identified that met the needs of the study.

During the interviews with the superintendents, limitations were placed on the study. These limitations included restrictions on the identification of actual practice in the district and guarantees of complete anonymity with no identification of individuals, or buildings. The most severe limitation was that of not allowing a follow up contact for the reluctant respondent who chose not to complete the scale.

A random sample in four districts was identified, and the districts cooperated in the distribution and return of the ATEP scale.

The data is best analyzed by an ANOVA that compares the mean scores on the ATEP scale between districts. As a further refinement

of the data for analysis, an item analysis was generated that provided mean scores on each item as well as the mode for each item.

The discussion of materials and methods in this chapter indicates the complex nature of this study. It also makes apparent that it is possible to identify a process by which the attitudes of teachers toward the evaluation process can be measured and the resulting data can provide a basis for acceptance or rejection of the null hypothesis.

CHAPTER IV

RESULTS AND DISCUSSION

The purpose of this study was to test the null hypothesis "There is no significant difference in the attitude of teachers toward the evaluation process in districts that use a clinical cycle style of evaluation and those districts that do not." This hypothesis was tested at the .05 level of significance.

The following procedure was used to generate the data for this study:

- I. Collection of statements to be used in the teacher attitude scale.
- II. Completion of a preliminary response scale in order to select the most relevant statements.
- III. Validity and Reliability tests for the completed scale.
- IV. Identification of districts to be sampled.
- V. Selection of sample and administration of scale.
- VI. Treatment of Data.

The data generated by this process indicates that the null hypothesis should be rejected. The L. S. D. (Least Significant Difference) test and the Newman-Kells procedure verified that there was a significant difference in the mean scores of the four districts. District D was identified as having a significantly lower mean score than Districts A,

B, and C. The following paragraphs outline the data generated at each step in the process which support the above conclusion.

Collection of Statements

The Likert Statement Request form was administered to 99 teachers. A total of 56 forms were returned for a response of 57 percent. This form generated 86 statements that relate to the evaluation process. After reviewing the 114 statements generated by the researcher and the statements generated by the sampling device, 66 were selected as appropriate for the preliminary response scale.

Administering the Preliminary Response Scale

The scale was administered to a random sample of teachers, grade levels K through 12, in two school districts similar to the districts that were used in the final sample. The sample consisted of 105 total responses with 49 from high school, 22 junior high, and 34 elementary.

The Pearson R correlation factor was used to correlate the rating (1-5) of each item against the average total score on the scale. The average total score on the scale was selected in order to compensate for non-response to an item. The correlations ranged from .5679 to -.2531. Twenty items were chosen with correlations that ranged from a definite but small correlation to a moderate relationship

(Table 2, p. 54).

These twenty items were accepted as appropriate for the ATEP scale (attitude toward the evaluation process).

Validity Study

The results of the Known Group Validity Study produced a Pearson R correlation coefficient for the high group of .3584 and a value of "P" equal to .041. The size of the random sample for this part of the study was 84.

Based on the criteria for statements of correlation (Table 2, p. 54), these values indicate there is a definite, but small relationship between the teacher's attitude and its measurement by this device. Oppenheim (1966) suggests the following.

Inadequate questionnaire construction is not the only cause of poor validity. The problem of validity remains one of the most difficult in social research, and one to which an adequate solution is not yet in sight (p. 78).

Downie and Heath (1974) suggests similar limitations by stating

These validity coefficients tend to be much lower than reliability coefficients. An examination of the research over the years will show that they tend to fall within the band of .4-.6, with a median value of about .5 (p. 244).

The correlation coefficients for this study are .3584 and .3379. Based on the pattern identified by Downie and Heath (1974), the correlation coefficients were accepted as adequate to establish the validity of the ATEP scale for the completion of this study.

Reliability Study

The Spearman/Brown split half correlation was completed, utilizing the odd and even items and the average of the total to compensate for non-response on an item.

The results of this are as follows:

Correlations between forms = .64171

Unequal - length Spearman/Brown = .78176 (10 items, even numbers on scale)

Alpha for part 1 = .51758

Equal - length Spearman/Brown = .78176

Guttman Split-half = .78173

Alpha for part 2 = .52843 (10 items, odd numbers on scale)

Based on the criteria for statements of the correlation table 2 (p. 54), there is a substantial relationship indicated by the correlation figures which indicates that the scale will be reliable.

A summary of the results of the validity study indicates a definite, but small relationship between the teacher's responses and the scale's ability to measure attitude. The reliability study indicates that there is a substantial relationship between the scale and its ability to repeat the results with different populations. This completes a development of the ATEP scale and suggests that it is adequate for the completion of this research.

Table 2. CRITERIA FOR STATEMENTS OF CORRELATION

Correlation Value	Approximate Descriptive Meaning
less than .20	Slight, almost negligible relationship
.20-.40	Low correlation; definite but small relationship
.40-.70	Moderate correlation; substantial relationship
.70-.90	High correlation; marked relationship
.90-1.00	Very high correlation; very dependable relationship
r r ₂	Guilford (1965) P 145

Selection of School Districts

Four public school districts agreed to participate in this study. All of the districts are considered suburban school districts, and all are within fifteen miles of Portland, Oregon. The student population ranged from a high of 20,154 to a low of 7,046. The teacher population ranged from a high of 1,244 to a low of 389. All of the districts had elementary, junior high, and high schools. Two of the districts had kindergarten programs, and two did not.

The four districts are generally the same in geographic location (suburban), and organization (high school, junior high, and elementary). The major difference is in student population. One district is approximately 37 percent larger than the other three (Table 3, p. 55).

Table 3. COMPARISON DATA FOR DISTRICTS (DEMOGRAPHICS)

Geographic Location	Suburban	Suburban	Suburban	Suburban
Total Student Population	7,562	20,154	7,046	7,546
Total Teacher Population	719	1,244	389	402
High Schools				
Number	3	3	1	2
Student Population	3,858	5,052	1,350	2,866
Staff	188	295	75	166
Grades	9-12	10-12	10-12	9-12
Junior High/Middle School				
Number	4	6	3	2
Student Population	1,988	5,161	1,571	1,709
Staff	96	271	101	84
Grades	7-8	7-9	7-9	6-8
Elementary				
Number	19	29	12	5
Student Population	1,716	9,946	4,025	2,971
Staff	435	678	211	150
Grades	K-6	1-6	K-6	1-5

These districts are regulated by the state in terms of the evaluation process, with no regard to size, geographic location, or organizational pattern. The common element in the public school districts is the fact that all teachers are involved in an evaluation process. Since the focus of this study is the relationship of teacher's attitude toward the evaluation process, the demographic dissimilarities should not have a major impact on the results of the study.

For the purpose of this study, the four districts were compared (see Table 1, p. 41) against the following criteria: Type of evaluation process, materials provided the teacher, training provided the evaluator, and whether or not they have a person responsible for improvement of instruction. One of the districts was designated as District A because it met all of the criteria. Another district was identified as District D because it did not meet any of the criteria. Districts B and C were so designated based on their partial compliance with the criteria.

Sample

A random selection of teachers was accomplished by using the district rosters and the random digit tables. The teacher population of the four districts totaled 2754. The random sample identified was 1100, which is approximately 39.9 percent of the total population. A return of 707 scales was 64 percent of the random sample or 25.6

percent of the total teacher population. Table 4 (p. 58) provides the same information by district. (Total number of teachers, size of sample, percentage of total population, number of returns, percentage of sample, percentage of total population.)

The random sample obtained meets the requirements of this study since the combined total returns equal 25.67 percent of the population. The four school districts range from a low of 22.9 percent of the population to a high of 35.9 percent.

According to Oppenheim (Questionnaire Design and Attitude Measurement) a return of 40 percent to 60 percent is typical for respondents with no special interest in the subject matter. The combined returns for all four districts was 64.27 percent. The districts individually ranged from a low of 67.10 percent to a high of 69.50 percent. Oppenheim (1966) speaks to the question of bias.

The important points about these poor response rates is not the reduced size of sample, which could easily be overcome by sending out more questionnaires, but the possibility of bias. This is because almost invariably the returns are not representative of the original sample drawn; non-response is not a random process; it has its own determinants, which vary from survey to survey (p. 34).

Chapter III, in the discussion of the limitations placed on the by the school districts, makes note of the fact that no follow up would be allowed.

The investigator would have preferred face to face interviews, use of additional scales, or follow up reminder notes to the respondents.

Table 4. SAMPLING DATA FOR DISTRICTS

	District A	District B	District C	District D	Combined Totals
Total number of teachers	719	1, 244	389	402	2, 754
Size of sample	274	410	215	201	1, 100
Percentage of population	38%	32. 9%	55%	50%	39. 94%
Returns	168	285	140	114	707
Percentage of sample returned	61. 31%	69. 50%	65. 12%	56. 71%	64. 27%
Percentage of total population	23. 36%	22. 90%	35. 9%	28. 35 %	25. 67%

Since none of these techniques was used, any generalization from the sample back to the population must be done with the awareness that experimental bias does exist.

Treatment of the Data

The one way ANOVA (analysis of variance) was used as the primary statistical tool to analyze the data generated by this project. All decisions were based on the .05 level of significance.

Mean scores for all respondents were compared by districts in order to determine the acceptance or rejection of the null hypothesis. Other combinations were analyzed in order to determine if there was any effect. These combinations were:

1. Comparison of all respondents by grade level.
 - a. Effect of district
 - b. Effect of grade level
 - c. Effect of grade and district
2. Comparison of district by grade level.
3. Comparison of grade levels within districts.

Primary Analysis

The respondents' mean scores were used to compare the four districts, and the statistical tool used was the ANOVA. The results of the ANOVA were accepted or rejected based on the value of "p".

This value, "p", is based on a probability factor that suggests the resulting numerical value represents the level of significance.

The ANOVA indicates that the mean score for District D is significantly lower than the mean scores for Districts A, B, and C (Table 5, p. 61).

District A: Mean score 77.3214

District B: Mean score 78.0546

District C: Mean score 76.8929

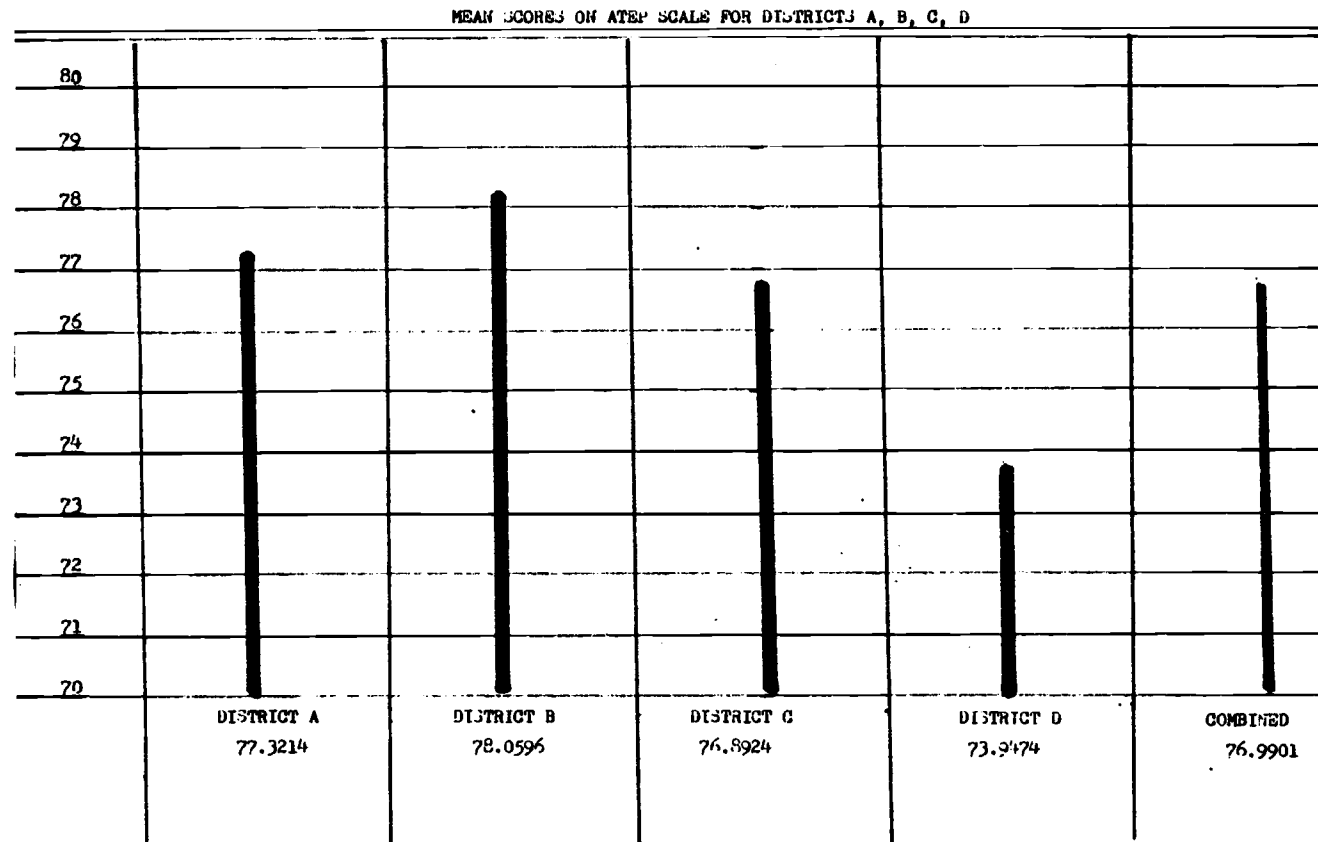
District D: Mean score 73.9474

All decisions were measured at the .05 level of significance and the results were $p = .0178$. The L.S.D. test and the Newman-Kells procedure verified that the difference was indeed between District D and Districts A, B, and C.

For this reason, the null hypothesis has been rejected. This implies that District A, which meets the criteria for this study (Table 1, p. 41), would have the highest score, indicating the most positive attitude. The other districts in order should rank as follows: District B, C, and D.

However, when the districts are rank ordered by mean scores, the order is District B first, District A second, District C third, and District D fourth. Since the ANOVA shows no significant difference between Districts A, B, and C, there is no statistical significance to this change in rank order. Table 1 (p. 41) shows two variations in the

Table 5



criteria between Districts A and B. District B has no specific training program for evaluators, and the evaluation and supervision responsibilities are not the responsibilities of separate individuals.

The only information that may account for the change in rank order between Districts A and B is the fact that District B's procedures for evaluation served as a model for the development of the procedure used in District A. As a result, the procedures in District B have been in use longer than in District A.

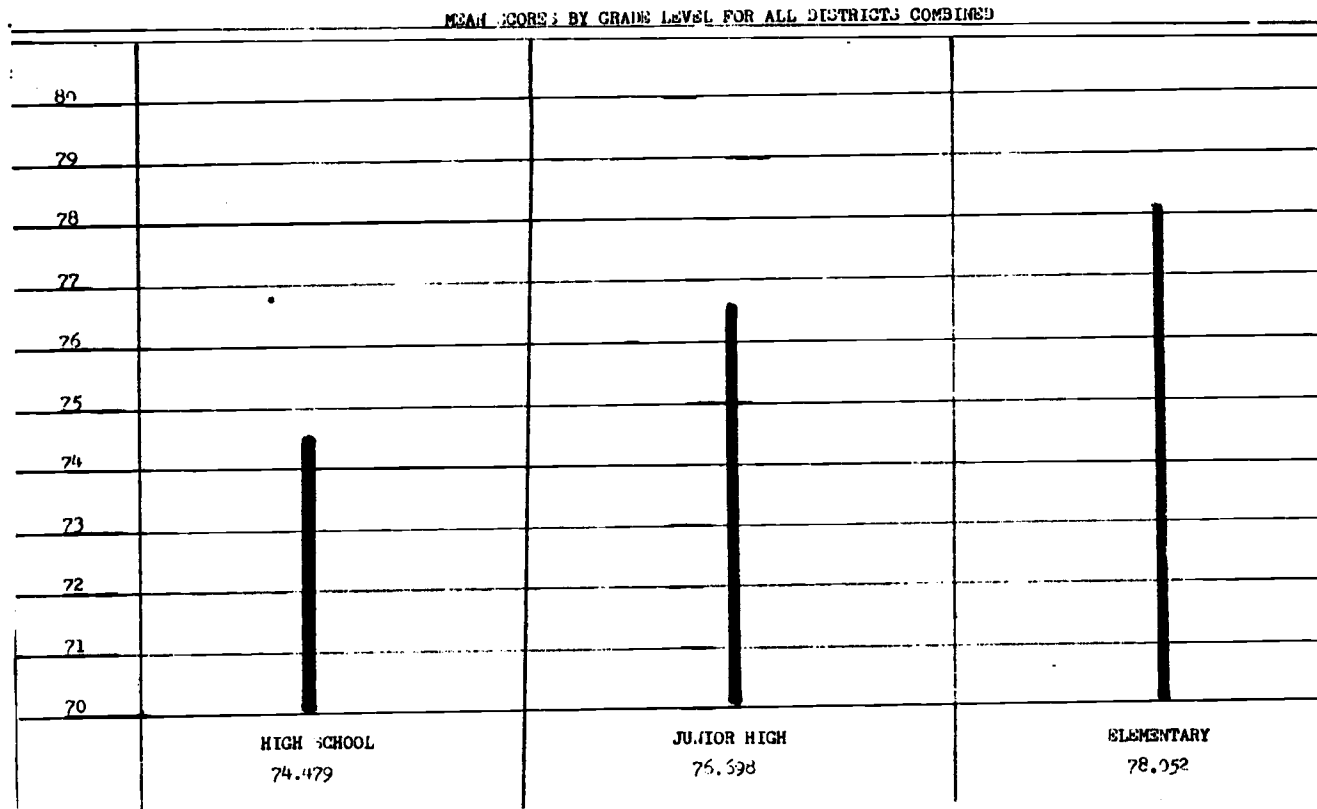
When the criteria used for designating the experimental group and the control group is compared to the results of the ANOVA, the only conclusion to be drawn is that a district wide, specified procedure for evaluating teachers is likely to produce a positive attitude in teachers. Conversely, it may be stated that teachers are more likely to hold a negative opinion in districts that have no district wide procedure.

Respondents Compared by Grade Level

The statistical consultant (Susan Maresh, O.S.U. Computer Center) suggested the mean scores be compared by a two way ANOVA. The reason for suggesting the two way ANOVA was the fact that a more complete analysis can be made of the two factors which were grade and district.

For all districts, the mean scores, when separated by grade level, are elementary teachers with a mean score of 78.052, the junior

Table 6



high with a mean of 76.698, and the high school with the lowest mean of 74.479. There are no main effects of districts since $p = .209 > .05$, which means that the average for all four districts indicates no significant difference among the grades. Also, there is no Grade X District interaction $p = .177 > .05$. This means that the relationships between the grade levels are the same for all districts.

The results of this analysis suggest that either there is no difference in the attitude of the teachers in each grade level category, or the instrument is not discrete enough to detect a significant difference. One could assume the rank order of results would occur that would indicate the elementary teachers have the most positive attitude, junior high second, and high school teachers have the least positive attitude.

The research suggests (Chapter II) that the most important factor is the principal's interpersonal relationships with the teachers. Obviously in the elementary schools used in the study (Table 3) the administrator to teacher ratio is lowest in elementary schools and highest in the high schools. Therefore, further research is needed to establish the fact as supported by this data, that no difference occurred in grade levels by district or, if, there is a variation in attitudes as suggested by the ranking of the grade levels by the mean score, it is undetected by the ATEP scale.

Comparison of Districts by Grade Levels

The mean scores for the high school respondents, when rank ordered by criteria (Table 1, p. 41) follow the pattern: District A, first mean score 78.250; District B, second mean score 77.880; District C, third mean score 75.106; and District D, fourth mean score 71.104 (Table 7, p. 66).

The ANOVA comparing the high school mean scores indicates that there is a significant difference in scores between districts due to the fact that $p = .0231 < .05$. Neither the L.S.D. nor the Newman-Kells procedure was able to identify where the difference exists between the districts. All that can be said is that there are significant differences among the districts for the high school mean, but where the difference lies cannot be determined.

The junior high mean scores rank the districts in the following order:

District D: 79.271

District B: 77.139

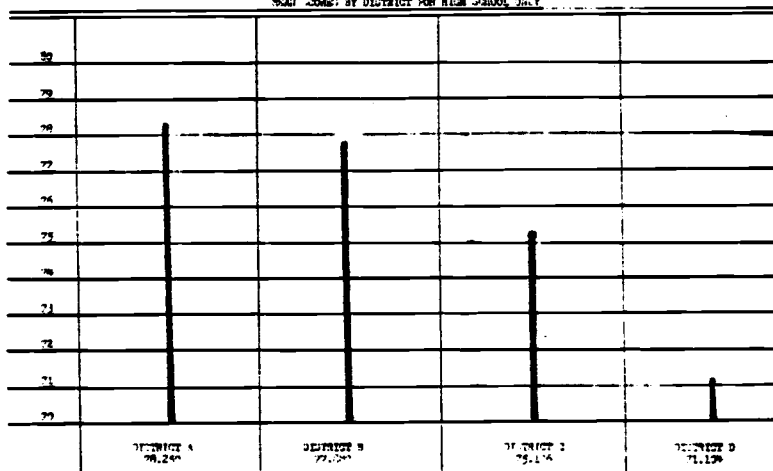
District C: 74.234

District A: 73.800

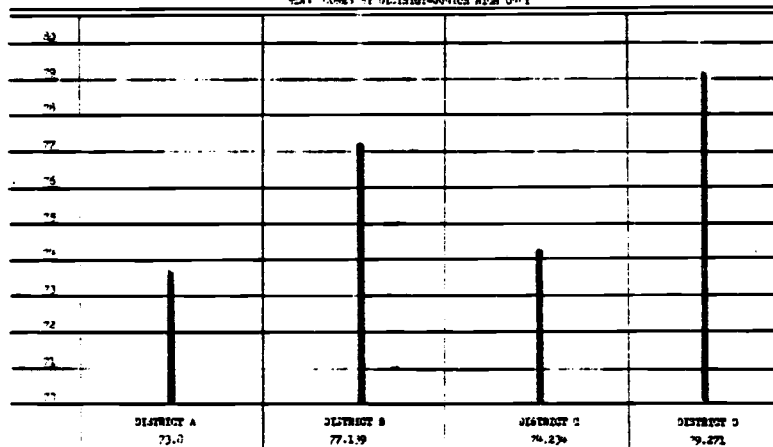
This rank order does not follow any obvious pattern. It is of interest to note that District D, which was ranked last at the high school and elementary grade levels was ranked first at the junior high level (Table 7,

Table 7

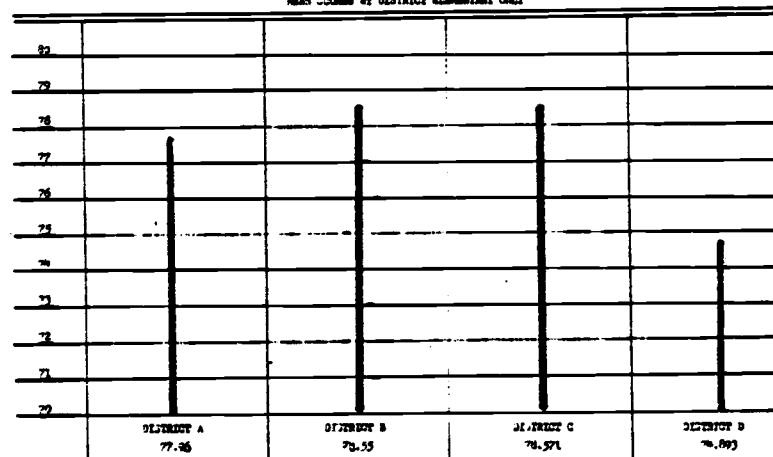
WEAT SCORES BY DISTRICT FOR HIGH SCHOOL ONLY



WEAT SCORES BY DISTRICT-JUNIOR HIGH ONLY



WEAT SCORES BY DISTRICT-ELEMENTARY ONLY



p. 66).

However, the one-way ANOVA indicates $p = .3953 > .05$ which means there is no significant difference in the mean scores for the junior high level.

The elementary mean scores rank the districts in the following order:

District C: 78.571

District B: 78.55

District A: 77.96

District D: 74.893

The ranking indicates no pattern that can be attributed to the data available. The ANOVA indicates $p = .3896 > .05$. There is no significant difference in these scores (Table 7, p. 66).

Summary

For the purposes of this study, the ANOVA has not indicated a significant difference in mean scores by district when comparing grade levels. This suggests the districts are generally administering the evaluation process as outlined to all grade levels in the same manner. The results also suggest that teachers view the process in a similar manner regardless of grade level.

Grade Level Within Districts

The last analysis, that of comparing grade levels within districts, was completed with the following results (Table 8, p. 69):

District A - mean scores by grade level

High School: 78.250

Elementary: 77.960

Junior High: 73.800

Analysis of Variance: $.6280 > .05$ (no difference)

District B - mean scores by grade level

Elementary: 78.55

High School: 77.88

Junior High: 77.139

Analysis of Variance: $.7268 > .05$ (no difference)

District C - mean scores by grade level

Elementary: 78.571

High School: 75.106

Junior High: 74.234

Analysis of Variance: $.4302 > .05$ (no difference)

District D - mean scores by grade level

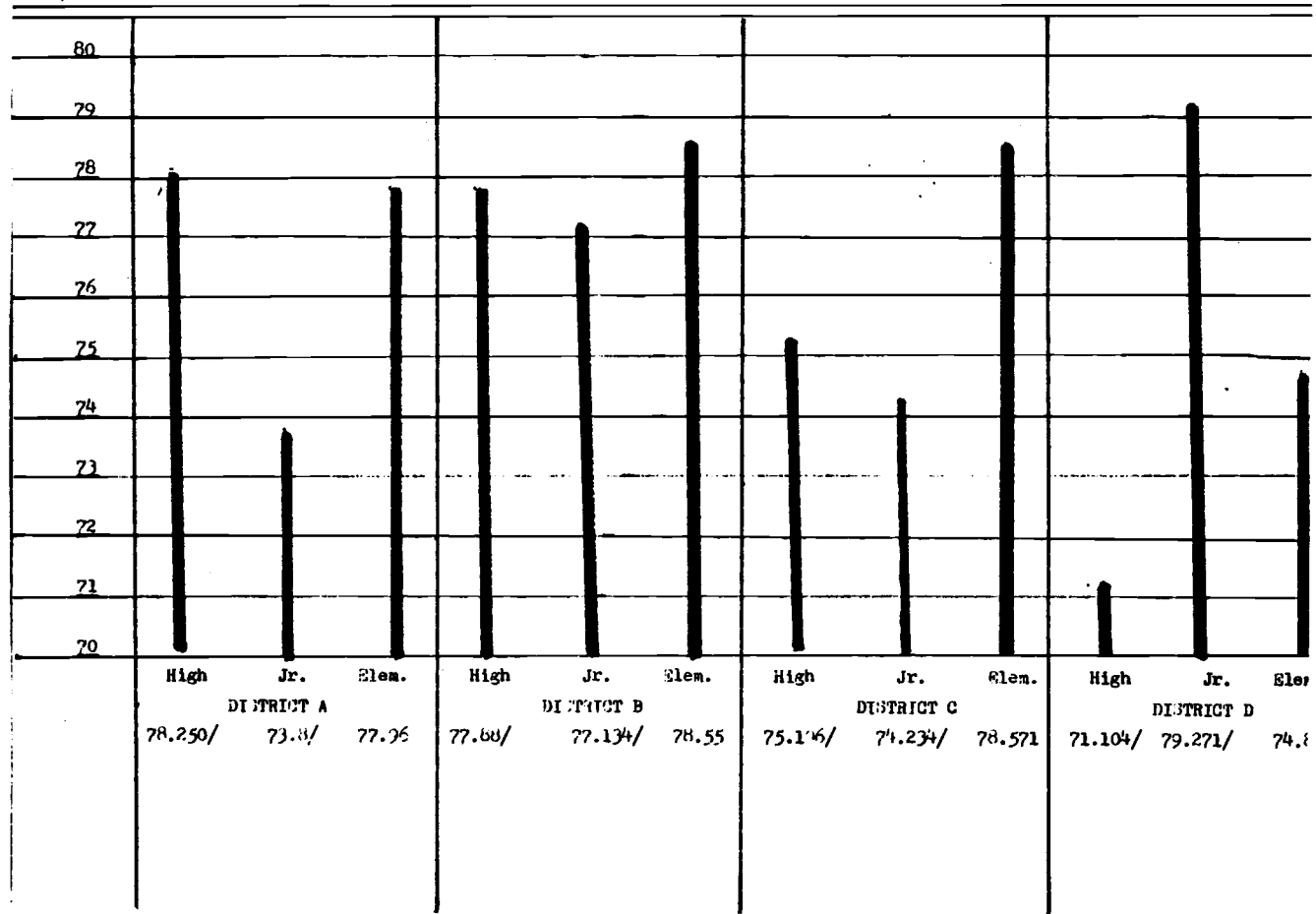
Junior High: 79.274

Elementary: 74.893

High School: 71.104

Table 8

MEAN SCORES FOR ALL DISTRICTS BY GRADE LEVEL



Analysis of Variance: $.0226 < .05$ (difference exists)

The rank order of grade levels by district suggested only two patterns. The junior high ratings are lowest in Districts A, B, and C, and highest in District D. The elementary teachers were ranked first in Districts B and C, and second in Districts A and D. The researcher can attribute no significance to the patterns.

Four one way ANOVAS, one for each district comparing the three grade levels (District A $p = .6280 > .05$, District B $p = .7268 > .05$, District C $p = .1302 > .05$, indicated no significant difference in the mean scores for the grade levels. The L.S.D. (Least Significant Difference) test and the Newman-Kells procedure at the $.05$ level verify these results. District D with $p = .0226 < .05$ indicates that there is a significant difference among the grade levels. The Newman-Kells and L.S.D. (Least Significant Difference) procedures both agree in concluding that $HS = EL$ and $EL = JR$, but HS is significantly lower than JR ($HS < JR$).

Here again, the results suggest that the districts tend to implement the evaluation process in a similar manner regardless of grade level, and that teachers react to this process in a similar manner. The significant difference found in District D (the high school teachers' attitudes more negative than the junior high teachers) introduces an anomaly that cannot be analyzed with the data available to the project.

Item Analysis

An item analysis was completed which provided the mean and the mode for the respondents in the following combinations: All respondents compared by grade level, high school, junior high, and elementary; respondents compared by grade level within each district. For the purposes of this research, the mode was selected as the statistical base for analysis. The purpose for comparing the modes in these combinations is to establish a statistical base for the selection of items that are relevant to the evaluation process.

The mode for the combined scores for all districts was selected as the primary source of data for selecting the items. These were separated by scores of five and three. When rating the items by grade level, the fives and threes were identified when they were consistent in all grade levels. This initial screening indicates that items 1, 2, 3, 6, 8, 10, 11, 17, and 18 are rated five on the ATEP (Attitude Toward Evaluation Process) scale. Items 4, 9, 12, 13, 15, and 16 are rated three on the ATEP (Attitude Toward Evaluation Process) scale. The same process was used to identify the items for all respondents by grade level and to compare the items in the four districts.

The only apparent pattern, as indicated by Table 9 (p.72), was that items 8, 11, 17, and 18 were consistently rated five. There was no apparent pattern to the items rated three.

Table 9. ITEM ANALYSIS FOR HIGH AND NEUTRAL SCORES

No.	All Districts Combined		All Respondents By Grade Level		District A By Grade Level		District B By Grade Level		District C By Grade Level		District D By Grade Level	
	5	3	5	3	5	3	5	3	5	3	5	3
1	X											
2	X											
3	X								X			
4		X										
5												
6	X		X		X		X					
7												
8	X		X				X		X		X	
9		X										
10	X		X				X					
11	X		X		X		X		X		X	
12		X		X						X		X
13		X										
14												
15		X		X								X
16		X		X						X		
17	X		X		X		X		X		X	
18	X		X		X		X		X		X	
19						X						
20										X		

Positive Scale Items

1. It really makes me feel more relaxed to talk over the class visit directly after it has occurred.
2. I feel better about the evaluation process when I know my colleagues are undergoing the same process.
3. It makes me feel more at ease if the evaluator explains how he/she will collect data during the classroom visit.
6. It is important to me to be notified by the evaluator if he/she is going to change the pattern of evaluation.
8. It is important to me for the evaluator to visit my class during the evaluation process.
10. The process of evaluation should be the same on a district wide basis.
11. It is important for me to be notified in writing when the evaluator wants a change in my teaching habits.
17. It is important for the evaluator to notify me in writing if there is a change from improvement to dismissal.
18. The process by which the evaluator arrives at his/her evaluation statement is of importance to me.

The positive items rated five are divided into two groups. Items 1, 3, 8, and 18 are elements of the clinical cycle of supervision.

The analysis supports the importance of the ideas found in the clinical cycle of supervision which are pre-conference (item 3), classroom visitation (item 8), and post conference (item 1). The principle of clinical supervision which suggests that a clearly defined process is important is supported by item 18.

The criteria for the district package (Table 1, p. 41) identifies elements that are reflected in the items 2, 6, 11, 17, and 10. A generalized statement about these elements would be that teachers have a more positive attitude toward the process if they know that the process is district wide and that any concerns or changes in the process will be put in writing.

Neutral Scale Items

4. There really is a need for a formal meeting before a class visit by the evaluator.
9. Writing out goals and objectives, in preparation for a classroom visit, is important.
12. I am more secure when the performance level is the score throughout the district. (Indicates a neutral response really don't care.)
13. The evaluator is more effective if he/she uses some formal method of collecting data (checklist, interaction analysis, etc.).

15. The evaluators who have had training in the process of evaluation used by the district make me feel more at ease.
16. The verbatim transcript taken by the evaluator provides me with valuable information.
19. It is important for me to explain my objectives in the lesson to the evaluator before he/she visits the class.
20. Goal setting early in the year makes me feel more comfortable.

The items that were rated three are 4, 9, 12, 13, 15, 16, 19, and 20, and refer to the elements of the clinical cycle and the criteria for the district package (Table 1, p. 41).

Items 4, 9, 19, and 20 deal with goal setting as part of the evaluation process. According to the ratings, this is not a particular concern for the teachers, yet it is part of the clinical cycle and included as part of the district package as outlined by state law (ORS 342.850).

One of the criterion (Table 1, p. 41) concerns training processes for the evaluator. Item 15, which reflects this idea, is rated as neutral, suggesting that formal training in the process is of no great importance to teachers. Item 16 is of special interest because it suggests that the verbatim transcript (primary data collecting device in clinical supervision) is not important in the process. Of the 18 unsolicited comments, only one said that this was a valuable tool. The majority either did not know what a verbatim transcript was or said that it was never used.

Item number 12 had 24 unsolicited responses which indicated that teachers did not understand what the performance level comment referred to. The only generalization that can be drawn from the analysis of these items is that teachers respond in a neutral way to the value of goal setting, formal methods of collecting data, and the verbatim transcript.

Unsolicited Written Comments

The ATEP scale was designed to make it difficult for a respondent to write in comments. However, the teachers chose to put short comments, such as "what does this mean" or write rather extensive comments expressing an idea or concern (Appendix E). Table 9 (p. 72) shows that items 3, 6, and 18 received no comments. Item 12 received the largest number, 24. These comments ranged from statements such as "how am I supposed to know" to some indicating a problem with the item such as "I don't understand the question." Some statements support the validity of the ATEP scale through comments such as "this is not the way I'm evaluated" and "this is how I feel about the process." This type of statement indicates that the item measures attitude rather than what the evaluator may or may not do.

The notes written by the teachers (Appendix E) reflect no pattern related to the rankings of the districts by the ANOVA. In District B, rated as having the most positive attitude, some comments were

positive and some were negative. The comments generally indicate that there is a discrepancy between the actual evaluation practice and the procedures outlined in the district packages. The other group of comments centered on either suggestions for the improvement of the scale or questions of clarification.

Summary

This study has as its purpose the identification of an evaluation process that will produce a positive attitude within public school teachers. Essentially, there were two parts to the study. One, the development of a scale to measure attitude, and the other, to test the hypothesis with the ATEP scale.

Reliability and validity studies were completed which suggest that the scale will be quite reliable and that there is a definite, but small relationship when measuring the teacher's attitude.

A sample bias has been introduced due to the fact that no follow-up was allowed by the superintendents in the four districts. The ANOVA indicated that there was a significantly lower score in District D as compared to Districts A, B, and C. The mean score changed the ranking of Districts from A, B, C, and D to B, A, C, and D. One possibility for this change in the ranking was the factor that the process developed in District A was based on the process used

in District B. District B, therefore, had this procedure functioning over a longer period of time.

When all of the respondents were divided into grade levels (high school, junior high, and elementary), the ANOVA indicated that there were no significant differences in the mean scores for these three groups. When the grade levels were compared by district, the ANOVA indicated there was a significant difference between the districts for the high schools, but the L.S.D. and the Newman-Kells procedure were unable to determine where the difference lies. The procedure also indicated there was no significant difference between the mean scores of the junior high and elementary teachers. When the grade levels were compared within each district, the ANOVA indicated that in Districts A, B, and C there was no significant difference in the mean scores. The ANOVA indicated that in District D the high school score was significantly lower than the junior high score. The item analysis identified items that indicated a positive attitude toward the primary elements of the clinical cycle and the items identified in the district package. The item analysis indicated that the primary data collecting method (verbatim transcript) was rated as neutral. The unsolicited comments tend to support these ideas and suggest that there is a difference between what is written in the district procedure and what actually happens. The implications of these findings and how they relate to the attitudes of teachers toward the evaluation process will be discussed in Chapter V.

CHAPTER V

SUMMARY AND IMPLICATIONS

The basic problem in this study was to investigate the relationship between the teachers' attitudes and the evaluation procedure used by the district. The hypothesis stated in null form was that there would be no significant differences in the attitudes of teachers toward the evaluation process in districts that have a clinical cycle model for evaluation and those districts that do not. The hypothesis in this study was tested at .05 level of significance. The ANOVA indicated that there was a significant difference between one district (District D) and the other three (Districts A, B, and C). Therefore, the null hypothesis was rejected.

The ANOVA

The fact that District D had a significantly lower mean score is in keeping with the rank order identified by the criteria in Table 1 (p. 41) ranking District D fourth. Further comparisons do not identify any patterns that match the rank order produced by the ANOVA. For example, there are no significant differences in the mean scores for Districts A, B, and C, yet the mean scores rank order the Districts B, A, and C. District B only met two of the four criteria, that of having a package available, and the implied use of the clinical cycle.

District C met one of the criteria by having evaluators trained in another system for evaluation called ITIPS (Instructional Theory into Practice System). Another reason for ranking District C third was the fact that the Evaluation Process (Clinical Cycle Model) and the Evaluation Package were introduced to the teachers after the sample was taken.

Under the given conditions, there is no apparent way to separate one of the four criteria as being more important than another. District B was ranked ahead of District A by the mean score. There again is no statistical evidence to account for this change in original ranking suggested by the criteria (A, B, C, D).

The two way ANOVA indicated there was no significant difference in the mean scores when all respondents were compared by grade level, nor was there any effect by district.

When the grade level mean scores were compared by district, the high school mean scores by district were significantly different, but the LSD test and the Newman-Kells procedure were unable to locate the difference. The fourth analysis, comparison of grade level within districts, indicated that only District D had a significant difference in scores between grade levels. The high school mean scores were significantly lower than the junior high school scores.

These procedures indicate that the districts must be evaluating all teachers in approximately the same way and that the teachers'

attitudes toward the process are the same regardless of grade level.

These conclusions are supported by the fact that the ANOVA did rank the districts in approximately the same relation as forecast by the criteria and that the grade level comparisons indicate no consistent pattern.

The fact that the districts were rank ordered by the ATEP (Attitude Toward Evaluation Process) scores in approximately the same order as determined by the criteria (Table 1, p. 41) indicates the elements of the process of evaluation are significant in affecting teacher's attitudes. The ANOVA supports this idea by identifying a significantly lower rating in District D when compared to Districts A, B, and C. However, the ANOVA indicates no significant difference in the scores between Districts A, B, and C.

An alternative analysis would be that the ATEP scale is not sensitive enough to identify the variations between grade levels or differentiate between districts that meet some but not all of the criteria in Table 1 (p. 41).

Item Analysis

The items which the ATEP scale indicated had the most frequent positive response dealt with the preconference, data collecting, post conference, and district wide evaluation procedures. These items were selected by comparing the mode across groups (example: All groups chose 5 (strongly agree) most often. See Table 9, p. 72). These three

elements (preconference, data collecting, and post conference) are the key elements of the clinical cycle of supervision. The fact that these elements were identified in the item analysis suggests their value in creating a positive attitude in teachers toward the evaluation process.

The items that received a neutral response (rating of 3) indicate that some of the elements of the clinical cycle and the district evaluation packages are not of great concern to the teachers. As an example, scale item 16 (verbatim transcript) consistently received neutral ratings and 18 unsolicited comments that, in general, indicated the verbatim was not used or the teachers did not know what it was. Scale item 12, that dealt with performance level for all teachers throughout the district, was rated as neutral (3) and received a total of 24 unsolicited comments. The majority indicated that it was either of no concern or that the teachers did not understand the concept of "performance level." These comments suggest that further study is needed to determine if the teachers are uninformed about the verbatim and performance levels or if they see no value in them. Identifying a district wide performance level will become a concern of the districts because ORS 342.850 specifies that standards be identified for meeting established criteria (performance level) and that evaluators be trained in various methods of collecting data, one excellent method being the verbatim transcript.

Unsolicited Comments

The unsolicited comments, in general, reflect a negative feeling on the part of the teachers toward the evaluation process. District D, the district rated with the least positive attitude, had comments such as, "the principal needs to care, which is not the case," "administrators are not trained as competent evaluators," and "the statements in the scale do not relate to the manner in which the evaluation is carried out." In contrast to these statements, one respondent stated, "I feel very comfortable with my principal so I never feel threatened by her evaluation. If I did not feel comfortable with a principal, I would want a mere formal evaluation."

The comments, in general, can be categorized into three types. One group of comments indicates there is a need to correlate the actual practice used in evaluating with the written process. Another group of comments are esoteric to the particular district and, therefore, would be meaningful only to a knowledgeable person in that district. The third type of comment referred directly to the ATEP scale. Some questioned the ability of the scale to measure attitude and others qualified the way they they marked the scale.

The discrepancy between what is supposed to be and what is in relation to the evaluating process has been verified by the analysis of these comments. Further investigation of the discrepancy between the written procedure and actual practice was not possible due to the fact

that the superintendents requested that no follow up be made on the actual performance of the administrators (evaluators). The fact that teachers felt the need to make comments suggests that if an administrator chose to use the ATEP scale in a district, the scale should be re-evaluated in terms of incorporating space for written comments.

Implications of this Study

The process used by districts in the evaluation of teachers does seem to affect the attitude of the teachers as indicated by the ATEP scale. Attitude as defined by Allport is, "a mental or neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related" (1935, p. 8). Attitudes are comprised of three dimensions, cognitive, behavioral and affective. Attitudes are formed by direct experience, explicit or implicit learning and personality pre-disposition.

It follows that the procedure used in evaluation would affect the attitude of teachers. One could postulate that a process that deals with the three components of attitude and takes into account the manner in which attitudes are formed would create a positive attitude in teachers.

The clinical cycle model of evaluation contains the three dimensions considered responsible for the formation of attitude. One of its primary goals is the establishment of a trust relationship between the

evaluator and the teachers which represents the affective dimension. Another is the exchange of detailed information in the preconference about the manner in which the evaluation is to occur which attempts to meet the cognitive component. Classroom visitation and the post conference contain the behavioral components.

Sullivan (1980), in support of the clinical supervision model, states,

Taken together, these studies yield some findings in support of the clinical supervision model. There is evidence which points to validation of the model and indications that the model tenets and processes are compatible with the desires of teachers and administrators (p. 22).

For the purposes of this study, however, the clinical cycle model of evaluation is being investigated in the context of other elements that affect the process and have been outlined in Table 1 (p.41). These additional elements are an evaluation package provided to each teacher, administrators trained in the process, and the roles of supervision and evaluation are the responsibilities of separate individuals.

Any implications from this study must be developed with the awareness that limitations are inherent in the study. From a statistical basis, the most severe is the fact that the school districts would not allow further contact with the respondents or allow a parallel investigation into the actual practice of evaluation. The attitude object for this study was the evaluation process and the data generally supports the idea that teachers view the process as an entity to be recognized as

important. However, other investigations show that other attitude objects may be of equal importance. For example, Blocker and Richardson (1963) support the idea that the administrator's interpersonal relationship is the most important factor affecting the morale of teachers.

Despite the above mentioned limitations, it is possible to suggest an evaluation model that will have the most positive effect on the teachers' attitudes toward evaluation.

The empirical data has indicated that the null hypothesis should be rejected, and the analysis of the additional data suggests that the process is important to teachers and that elements of the process should be retained.

The process should include the elements of the clinical cycle model, pre-conference, class visitations, post conference. The district should have a package available to the teachers that explains the process, identifies philosophy, and contains a job description, criteria to be used in evaluating the teachers, sample forms, and timelines for completing the process. There is limited evidence that the administrators should be trained in the clinical cycle through the S. O. A. I. procedure. However, there are enough unsolicited comments to suggest that at least the evaluator should be able to demonstrate competence in collecting data, regardless of a particular model used. The evaluation and supervision responsibilities should be separated according to the

principles developed by Morris Cogan (1980). However, in most districts this is not feasible for economic reasons, and, therefore, it is important to use the clinical model because built into the cycle are elements that assist the person responsible to function both as a supervisor and an evaluator.

Summary

The implications implicit in this study are that the evaluation process is important in creating a positive attitude in teachers. The elements of the process most likely to influence the attitude are the pre-conference and data collecting procedures, the post conference, the expertise of the evaluator, and the awareness of the need to separate elements of evaluation and supervision. Also implied in the data is the need for a district to have a written policy and procedure for evaluation. These implications appear valid in the context of the limitations of the study and the ratings produced by the validity and reliability studies.

The ATEP scale has established a basis for rejecting the null hypothesis, and at the same time, the data generated has made it possible to identify three areas for further research. These areas are:

- A. The need to investigate the ability of the ATEP scale to discriminate between the attitudes of teachers in districts with similar evaluation processes.

- B. The need to investigate the criteria in Table 1 (p. 41) to identify the most important of the four criteria and its effect on teachers' attitudes.
- C. The need to investigate the relationship of the teachers' attitudes toward the evaluation process and the other factors that affect attitude (example: Evaluation statements that are used for dismissal).

The first area for further research is suggested by the fact that, even though the scale did identify a difference between the attitudes in the Experimental population and the control, there was no significant difference in the three districts that had some similarities in their processes of evaluation. The analysis of the data by grade level, both within districts and between districts, indicated no significant differences.

The validity and reliability studies, as well as the analysis of the data, indicate the results are valid but further research in this area would add additional information to support the findings of this project.

There are several reasons for suggesting the second area of further research. The criteria identified in Table 1 (p. 41) was found to be in limited use. School districts either see no need to incorporate all four elements of the criteria or find it too expensive. The separation of evaluation and supervision, and the need for training all evaluators are two of the criteria that are directly related to the school

district's ability to provide funding. Some of the unsolicited comments indicated that a competent evaluator was very important.

Further research into the effect of other attitude objects on attitude is a need indicated by the information gathered in the review of related research. There is strong evidence that interpersonal relationships between the evaluator and teacher is the most important factor affecting attitudes. Therefore, a study that would compare the attitudes of teachers toward the process identified in this study and the other attitude objects would be justified.

When all aspects of the study are reviewed and analyzed, it is important that this study has identified an evaluation process that has a positive effect on the teachers' attitudes as well as areas for further research. The subject of teacher evaluation will continue to be of interest to public school employees since ORS 342.835 has mandated a model of evaluation similar to the one identified in this study.

BIBLIOGRAPHY

- Allport, G. W. Attitudes. In C. Murchism (Ed), A handbook of social psychology. Worcester, Mass. Clark University Press, 1935. pp. 798-844.
- Anderson, L. W. 1953. Teacher morale and student achievement. *Journal of Educational Research* 46:693-698.
- Barber, W. R. 1970. The principal and the morale. *School and Community* 56:23.
- Barr, A. S. Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness, Dembar Publications, Inc., 1961.
- Beecher, Dwight Everett. The Evaluation of Teaching Backgrounds and Concepts, Syracuse University Press, 1901.
- Bentley, R. R. and A. M. Remple. 1970. Manual for the Purdue Teacher Opinionaire. West Lafayette, Indiana, The University Book Store.
- Bhella, Surjit Kaur. 1975. An investigation of the relationship between administrator leadership style and teacher morale. Unpublished Doctoral Dissertation, Oregon State University. Corvallis, Oregon.
- Biddle, Bruce J. Contemporary Research on Teacher Effectiveness, Holt, Rinehart and Winston, 1964.
- Blocker, C. E. and R. C. Richardson. 1963. Twenty-five years of morale research--a critical review. *Journal of Educational Sociology* 36:200-210.
- Cogan, M. L. *Clinical Supervision*. Boston: Houghton Mifflin Co., 1973.
- Doob, L. W. The Behavior of Attitudes. *Psychol. Rev.* 1947. 54 135-156.
- Downie, N. M. and R. W. Heath. Basic Statistical Methods. Harper and Row, NY, 1974.

- Ellena, William J. Who's A Good Teacher, American Association of School Administrators, 1961.
- Fishbein, Martin. Readings in Attitude, Theory and Measurements, John Wiley and Sons, Inc., 1967.
- Gragg, W. L. 1955. Teacher Morale. *Clearing House* 29:493-494.
- Griffith, Francis. A Handbook for the Observation of Teaching and Learning, Pendell Publishing Co., 1973.
- Gross, N. and R. E. Herriott. 1965. Staff leadership in public schools: A sociological inquiry. New York: John Wiley and Sons, Inc., 247 p.
- Hale, James R. and Allen R. Spanger. Systematic and Objective Analysis of Instruction, Northwest Regional Lab, 1972.
- Hand, H. C. 1948. What makes for high teacher morale. *Educational Leadership* 5:279-280.
- Hood, E. C. 1965. A study of congruence of perceptions concerning factors which affect teacher morale. Unpublished Doctoral Dissertation, East Texas State University.
- Koura, H. S. 1963. An experimental study of students' achievement in relation to the morale of selected secondary school teachers. Unpublished Doctoral Dissertation. The University of Michigan, Ann Arbor, Michigan.
- Littrell, Donald W. The Theory and Practice of Community Development--A Guide for Practitioners, Extension Division, University of Missouri, Columbia, MP 184, 1968.
- Marks, Merle B. Effective Teacher Evaluation, NASSP Bulletin, September 1976.
- Martin, Dan T., George E. Overholt, and Wayne J. Urgan. Accountability in American Education: A Critique, Urban Education, April 1977.
- McGee, Jerry C. and Robert Eaker. Clinical Supervision and Teacher Anxiety, *Contemporary Education*, F 77.

- McNolte, J. How to Tell Which Teacher to Keep and Which Teacher to Lay Off, American School Board Journal, June, 1976.
- Napier, T. G. 1966. Teacher morale. Unpublished Doctoral Dissertation, The University of Nebraska Teachers College. Lincoln, Nebraska.
- Oppenheim, A. N. Questionnaire Design and Attitude Measurement. Basic Books, Inc., Publishers NY. 1966.
- Page, Colin Flood. Student Evaluation of Teaching the American Experience, Society for Research into Higher Education, 1974.
- Pollack, J. David. The Development and Testing of a Criterion Referenced Evaluation System for Faculty and Administrators in Technical Institutes/Community Colleges, Office of Education (DHEW), Washington, D.C., June 1976.
- Redefer, F. L. 1959. Factors that affect teacher morale. Nations Schools 63:59-63.
- Ross, W. E. 1960. A study of personal factors affecting the morale status of teachers of two rural school systems. Unpublished Doctoral Dissertation. New York University, New York.
- Simpson, Ray H. Teacher Self-Evaluation, MacMillan Co., 1966.
- Stern, Carolyn. Maximizing the Value of Evaluation for the Head Start, Office of Economic Opportunity, Washington, D.C., August 1969.
- Sullivan, Cheryl Granade. 1980. Clinical Supervision: A state of the art review. Association for Supervision and Curriculum Development publication.
- Thomas, Tommy Eldon. A Summarative Evaluation of S. O. A. I. Technical Report No. 9, 1972.
- Van Hoven, James B. Peer Involvement: A Working Model of Staff Evaluation, Catalyst for Change, F 74.
- Wagoner, Roderic L. and James Ottonlon. Teacher Attitude Toward Evaluation, EDRS.

Wiles, Jon and Joseph Bondi, Jr. Curriculum Development, Charles E. Merrill Publishing Company, 1979.

"Better Than Rating New Approaches to Appraisal of Teaching Services," National Education Association, Commission on Teacher Evaluation, 1950.

"Evaluation Guidelines for School Personnel," Oregon Department of Education, Kipp, Don, May 1977.

APPENDICES

APPENDIX A

HELP

I need your ideas to make a questionnaire to
measure attitudes toward the yearly evaluation process.

```

*****
*
* Sample statement that may measure attitude:
*
* I feel most comfortable when my evaluator pops in
* unannounced.
*
*      1          2          3          4          5
* Disagree          Neutral          Agree
*
*****

```

Please write your ideas or sample statements that
will help measure attitudes toward the evaluation process.
Do not limit your comments to the concept suggested in
the sample.

Thanks for your help

Return to Sheridan Jones

APPENDIX B

SAMPLE OF LIKERT SCALE ITEMS DEVELOPED BY INVESTIGATOR

1. Helpful suggestions from the evaluator make me feel as though he or she is displeased with my performance.
2. The process by which the evaluator arrives at his evaluation statements is of no importance to me.
3. Accuracy of the evaluation statements is my only concern.
4. The verbatim transcript taken by the evaluator provides me with valuable information.
5. I feel just as comfortable talking over the visit without any formal data.
6. The timelines for goal setting, visits, conferences, and notifications are a waste of time.
7. If I were a tenured teacher, there would be no benefit to me if evaluated after March 15.
8. The easiest and most meaningful way to be evaluated is to let the evaluator choose the process.
9. It doesn't bother me to have my evaluator include data in the evaluation that I was unaware was collected.
10. If there is a problem with my evaluation I would rather call in another person to give their opinion after a class visit.

11. The process for evaluation should be the same on a district wide basis.
12. The level of performance should be spelled out in writing for the entire district.
13. The basic points of concern or objectives for all teachers should be the same in the district.
14. How well each teacher meets the goals should be between the teacher and the evaluator, not a district standard.
15. The process used by the evaluator is more important to my feelings than the ratings I receive.

APPENDIX C

RESULTS OF LIKERT STATEMENT REQUESTELEMENTARY

At what point in evaluation is the evaluator observing, gathering evidence primarily for the purpose of non-renewal?

My evaluator hires and fires. Therefore, I cannot share my inadequacies without risk and fear that they will show up in the evaluation.

A prepared bit makes me feel like an actor and all that 'fill-out-the-purpose' makes it seem like other times my act is unnecessary.

Although I feel I must be "on my toes" so to speak, I don't feel nervous or threatened by my present evaluator. Perhaps this would not be true with a different evaluator.

I approve of an evaluator who makes frequent appearances in the lunchroom, on the playground, and throughout the building to keep aware of what is going on.

Many times when an evaluator plans to observe, the classroom is excited and a true teaching picture isn't observed. Sometimes teachers tend to do something special rather than everyday lessons.

JR. HIGH SCHOOL

I feel most comfortable when my evaluator has been evaluated as an evaluator.

Young teachers with fewer than seven years experience need an annual evaluation. Teachers with over seven years don't need to be evaluated every year. Maybe every two or three years.

Evaluators must find out ahead of time what they are seeing, goals, etc. How well qualified are our evaluators? Do teachers know what the evaluator's standards are?

After being in the building for five years or more, a formal evaluation of a classroom situation is not necessary.

A classroom evaluation situation is representative of an everyday classroom situation.

I would eliminate the word most in the above sample and probably add or announced.

I feel that the evaluation helps me improve my teaching skills.

If I were a first year teacher, I would be more comfortable with the "staged" evaluation, but I feel that a true evaluation can only be conducted over a long period of time and from many different samplings of informal situations.

I feel uncomfortable when my evaluator comes in unannounced.

I feel that my evaluator sees my best teaching techniques when he observes me.

I feel that the yearly evaluation is necessary.

HIGH SCHOOLS

I feel most comfortable when my evaluator has current knowledge of my subject area.

A more thorough pre-evaluation conference is a must.

I'm not sure that our evaluators have adequate academic background to judge a teacher's performance.

Evaluation of goals at the beginning of the year in relation to your teaching performance.

I would like evaluators from outside our building who would work with me on new and creative teaching ideas.

If a teacher is prepared to meet with the students, he should be ready to meet with the evaluators.

I feel most comfortable when my evaluator understands what I am teaching/doing.

I feel most comfortable when my evaluator follows the contract to the letter. (Especially the part where the evaluator makes it perfectly clear what is being evaluated - or - what the evaluation standards are.

I feel most comfortable when my evaluator visits more than one class.

If one of the purposes of evaluation is renewal of contracts, then the teacher wants to feel he has really prepared to show his best effort.

Evaluations should be conducted in a manner to help the teacher grow and improve teaching techniques.

APPENDIX D
[Sample: Page 1 of 4]

PRELIMINARY RESPONSE SCALE

Your response to these statements will help determine the twenty most indicative statements for a Likert style attitude scale.

The resulting data will not be identified by individual or building. Thank you for your cooperation.

INSTRUCTIONS: Circle the number that corresponds to your feelings about statements.

	1	2	3		
	Strongly	Neutral	Strongly		
	Disagree		Agree		
1. It is important for me to have the evaluator follow a procedure approved by the district.	1	2	3	4	5
2. A good evaluator is able to arrive at an accurate evaluation without engaging in any prescribed procedure.	1	2	3	4	5
3. Evaluation is for the purpose of hiring and firing.	1	2	3	4	5
4. The process by which the evaluator arrives at his/her evaluation statements is of importance to me.	1	2	3	4	5
5. Accuracy of the evaluation statements is my only concern.	1	2	3	4	5
6. The verbatim transcript taken by the evaluator provides me with valuable information.	1	2	3	4	5
7. The presence of any formal data makes me feel more comfortable in the post-conference.	1	2	3	4	5
8. To really help me, my evaluator should spend as much time as possible in my classroom.	1	2	3	4	5
9. Helpful suggestions from the evaluator make me feel as though he/she is displeased with my performance.	1	2	3	4	5
10. Most evaluators already know which teachers are to be rated high and low without going through any process.	1	2	3	4	5
11. It is important for me to explain my objectives in the lesson to the evaluator before he/she visits the class.	1	2	3	4	5
12. The performance level should be left to the discretion of the teacher and evaluator.	1	2	3	4	5
13. The evaluator should be able to help me with my problems without affecting my evaluation.	1	2	3	4	5
14. The most ideal situation is to have the responsibilities for supervision and evaluation carried out by different people.	1	2	3	4	5

APPENDIX E
*ATEP SCALE

The resulting data will not be identified by individual or building. Thank you for your cooperation.

INSTRUCTIONS: Circle the number that corresponds to your feelings about the following statements.

	1	2	3	4	5
	Strongly Disagree		Neutral		Strongly Agree
1. It really makes me feel more relaxed to talk over the class visit directly after it has occurred.	1	2	3	4	5
2. I feel better about the evaluation process when I know my colleagues are undergoing the same process.	1	2	3	4	5
3. It makes me feel more at ease if the evaluator explains how he/she will collect data during the classroom visit.	1	2	3	4	5
4. There really is a need for a formal meeting before a class visit by the evaluator.	1	2	3	4	5
5. The cycle of pre and post conferencing is a good way to evaluate a teacher.	1	2	3	4	5
6. It is important to me to be notified by the evaluator if he/she is going to change the pattern of evaluation.	1	2	3	4	5
7. It makes more sense to keep the focus of evaluation on the classroom teaching act.	1	2	3	4	5
8. It is important to me for the evaluator to visit my class during the evaluation process.	1	2	3	4	5
9. Writing out goals and objectives, in preparation for a classroom visit, is important.	1	2	3	4	5
10. The process of evaluation should be the same on a district wide basis.	1	2	3	4	5
11. It is important for me to be notified in writing when the evaluator wants a change in my teaching habits.	1	2	3	4	5
12. I am more secure when the performance level is the same throughout the district.	1	2	3	4	5
13. The evaluator is more effective, if he/she uses some formal method of collecting data (checklist, interaction analysis, etc.)	1	2	3	4	5
14. It is important for me to have the evaluator follow a procedure approved by the district.	1	2	3	4	5

- | | | | | | |
|---|---|---|---|---|---|
| 15. The evaluators who have had training in the process of evaluation used by the district, make me feel more at ease . | 1 | 2 | 3 | 4 | 5 |
| 16. The verbatim transcript taken by the evaluator provides me with valuable information. | 1 | 2 | 3 | 4 | 5 |
| 17. It is important for the evaluator to notify me in writing if there is a change from improvement to dismissal. | 1 | 2 | 3 | 4 | 5 |
| 18. The process by which the evaluator arrives at his/her evaluation statements is of importance to me. | 1 | 2 | 3 | 4 | 5 |
| 19. It is important for me to explain my objectives in the lesson to the evaluator before he/she visits the class. | 1 | 2 | 3 | 4 | 5 |
| 20. Goal setting early in the year makes me feel more comfortable. | 1 | 2 | 3 | 4 | 5 |

*Attitude toward Evaluation Process.

Paid for by Sheridan Jones

APPENDIX F

PLEASE RESPOND TO THE ITEMS ON THE THREE PAGES AND RETURN TO SHERIDAN JONES.
THANK YOU FOR YOUR HELP.

KNOWN GROUP IDENTIFICATION

Instructions: Please check the evaluation procedure that most closely describes the procedure used by your evaluator.

___ A. My evaluator follows a written, district approved procedure that includes guidelines for:

1. Goal setting
2. Pre Conference
3. Classroom visitation
4. Formal Data Gathering
5. Post Conference
6. Time lines for completion

___ B. My evaluator is free to use his/her own procedure that may or may not include items listed above. The procedure is known to vary from building to building and within buildings because there is no written district approved procedure of which I am aware.

___ C. My evaluator uses a combination of processes stated in A and B. Example: There is a district procedure but the evaluator doesn't follow it.

Explain Combination: _____

Instructions: Please circle the appropriate number 1 through 5.

My feelings about the procedure used in evaluating me are:

1	2	3	4	5
Negative		Neutral		Very Positive

APPENDIX G
RESULTS OF THE ANOVA
DISTRICTS COMPARED BY TOTAL SCORES

	<u>Count</u>	<u>Mean</u>	<u>Standard Deviation</u>
District A	168	77.32	11.06
District B	285	78.05	12.14
District C	140	76.89	11.49
District D	114	73.94	12.01
	.0178 at .05 level		
Total	707	76.9901	11.8049

SCORES BY GRADE LEVEL

$p = .209 > .05$	Effect of district	no significant difference
$p = .393 > .05$	Effect of grade	no significant difference
$p = .177 > .05$	Effect of grade X district	no significant difference

COMPARISON OF DISTRICTS BY GRADE LEVELS

High School	$p = .0231 < .05$	significant difference
Junior High	$p = .3953 > .05$	no significant difference
Elementary	$p = .3896 > .05$	no significant difference

GRADE LEVELS WITHIN DISTRICTS

District A	one-way anova 3 grade levels	$p = .6280 > .05$	no significant difference
District B	one-way anova 3 grade levels	$p = .7268 > .05$	no significant difference
District C	one-way anova 3 grade levels	$p = .1302 > .05$	no significant difference
District D	one-way anova 3 grade levels	$p = .0226 < .05$	no significant difference

APPENDIX H
ITEM ANALYSIS FOR ALL DISTRICTS COMBINED

<u>Number</u>	<u>Mean</u>	<u>Mode</u>
1	4.038	5
2	4.029	5
3	4.135	5
4	3.528	3
5	3.914	4
6	4.074	5
7	3.817	4
8	4.174	5
9	3.342	3
10	4.106	5
11	4.142	5
12	3.340	3
13	3.560	3
14	3.798	5
15	3.484	3
16	3.238	3
17	4.622	5
18	4.483	5
19	3.770	4
20	3.714	5

APPENDIX I

TALLY OF UNSOLICITED COMMENTS PER ITEM BY DISTRICT

<u>No.</u>	<u>District A</u>	<u>District B</u>	<u>District C</u>	<u>District D</u>	<u>Total</u>
1	2	5	2	1	10
2	0	2	0	0	2
3	0	0	0	0	0
4	2	4	2	0	8
5	2	2	1	1	6
6	0	0	0	0	0
7	3	6	2	2	13
8	2	4	1	4	11
9	2	7	2	4	15
10	1	6	0	0	7
11	1	3	2	1	7
12	4	11	5	4	24
13	3	3	0	1	7
14	2	1	0	1	4
15	2	4	3	2	11
16	2	8	6	2	18
17	6	5	1	2	14
18	0	0	0	0	0
19	2	3	1	1	7
20	1	1	1	1	4

APPENDIX J

UNSOLICITED COMMENTS BY DISTRICT

DISTRICT A

1. All of those questions refer to a formal evaluation of one classroom setting. Many of my fellow staff members feel that an evaluation should also be based on several informal unannounced visits by the principal, giving a more accurate day-by-day picture of a teacher's performance.
2. If this document is intended to measure attitudes toward evaluations requested by teachers, my responses might be quite different.
3. I feel that this whole thing depends on the principal that you happen to be teaching under. How you happen to feel about that person. How well you work together.
4. It makes me uneasy when an evaluator fails to make contact and finish the process - leaves it hang fire for months.

DISTRICT B

1. What can one really find out from this when people will mostly mark 5 - It asks how you want it, not how it is.
2. I have taught for 32 years and I am very rarely observed for a set

time period. However, the principal is in and out almost every day to bring messages.

3. I hope this helps in getting your Ph.D. It sure won't help me or my aids.
4. There needs to be more positive feedback after evaluations - not just criticism.
5. Thus far, I've been quite satisfied with the evaluation process. I'm doing a good job, my evaluator knows I'm doing a good job and tries not to bother me with any more than the absolute minimum of required hassles.
6. Many of these questions seem to have a bias toward a positive response.
7. This scale does not relate to my attitude toward the District.
8. I always manipulated my own goals - all busy work. Somewhere I'd like to state that the whole process seems like an expensive waste of time - those who shouldn't teach are still teaching because administrators can't evaluate properly.
9. As a school counselor, I find these questions do not apply in my case. I feel that the evaluation process is fair, but I do not find it to be particularly helpful as a tool to assist in Prof. growth. It is a procedure I tolerate since it is required. It's not threatening--a bit time consuming.

10. I do not know how I feel about some of this. I know you can't do anything about it, but I just thought you might find it interesting. The language in the evaluation handbook is so full of educational jargon and so vague I have difficulty understanding it. My whole concept of teacher evaluation in this district is definitely negative (It seems as if we are only supporting vague bureaucratic dogma). I have never had what I consider a realistic evaluation with a planned visitation by an evaluator in the past 7 years I have taught in the district! So what is the meaning of all this? I think many share my feelings but are afraid to say anything.
11. You are not asking the right questions. Over and above everything - do the techniques used by the evaluator do what they're supposed to do. In other words, can the evaluator, on the basis of the existing procedure, really know if I am or am not an effective teacher? There is the implication that classroom management and teacher effectiveness mean the same thing. Do they?
12. For my part and most of the people that I have talked to felt that evaluation by a formal evaluator is of not much help and must be quite costly.

I believe that a teacher who has fulfilled the required work to teach and has been selected to teach is much better qualified to evaluate himself or herself than a specialist who may not have much background in subject matter and/or teaching experience.

From my experience I know that most teachers are always improving themselves in order to provide the best and latest subject matter in the best way. When teachers need help they can best get it from other teachers and more opportunity should be made available for them to do so.

Since formal evaluation has been in vogue I have not received any suggestion for improvement that I was not already aware of and would put into practice with time available. On the other hand, my experience has been that the evaluation offered nothing and sometimes was negative. I cannot see how any well educated teacher would be very positive about an evaluation system that offers so little toward improving the educational product. Evaluation is not the area where the problem lies.

I believe that I would prefer to encircle a response to the questions on the reverse side only on the basis of qualifications.

DISTRICT C

1. I feel like the professional pressure is nearly there on teachers. Principals, seemingly can get away with quite a lot unchecked. This unbelievable attitude baffles me.
2. If you have a "good" (caring and intelligent) principal -evaluations can be well done just by his dropping by every so often. If you have a "bad" (uncaring, out of step, etc.) principal, no form of

evaluation is going to be of any importance! Therefore, I am basically neutral on the subject.

3. The positive attitude of the evaluator himself can put me at ease.
4. I think less formal evaluations - lots of drop in visits would be a more reliable way of evaluating.
5. These are pretty vague questions. What is your main direction. Most evaluations are nitpicking items of little importance.

DISTRICT D

1. This really is somewhat difficult to answer. Some of the statements are not appropriate to how the evaluation process is carried out. It depends on the attitude of the administrator.
2. The principal needs to care, which is not the case, based on his behavior or interaction. In 6 months, the principal has not once asked about the class education, or my needs as a teacher. The school is sterile, boring and very average. You need dynamic humanistic principals that put education first, not how clean the rug is in my room.
3. Not all building administrators are trained or competent in evaluating their teachers. Can't (referring to 15) respond accordingly, never allowed to see verbatim.
4. I feel very comfortable with my principal, so I never feel threatened by her evaluations. If I did not feel comfortable with a

principal, I would want a more formal evaluation.

5. Ambiguous statement - who writes them? (referring to 9)