1999

The Teacher Evaluation Process As A Means To Promote Professional Growth: An Examination Of Two Teacher Evaluation Instruments

Catherine M. Snyder

Seton Hall University

Follow this and additional works at: http://scholarship.shu.edu/dissertations

Part of the Educational Administration and Supervision Commons

Recommended Citation

http://scholarship.shu.edu/dissertations/1703
THE TEACHER EVALUATION PROCESS AS A MEANS TO PROMOTE PROFESSIONAL GROWTH: AN EXAMINATION OF TWO TEACHER EVALUATION INSTRUMENTS

BY

CATHERINE M. SNYDER

Dissertation Committee

Daniel Gutmore, Ph.D., Mentor
Frank J. Fehn, Ed.D.
Joanne K. Monroe, Ed.D.
Joseph Stetar, Ph.D.

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Education
1999
DEDICATION

This dissertation is dedicated to my daughter, Kristen, without whose encouragement and loving support I would not have completed this task.
# TABLE OF CONTENTS

**LIST OF TABLES** .................................................. v

**STATEMENT OF THE PROBLEM** ........................................ 1

Background .............................................................. 1
Purpose of this Study .................................................. 8
Summary of the Problem ............................................... 9
Research Questions ...................................................... 10
Assumptions ............................................................. 11
Definition of Terms ..................................................... 12
Limitations of the Study ............................................... 13

**THE REVIEW OF THE RELATED LITERATURE** ....................... 15

An Historical Perspective ............................................ 15
Impact of Research on Organizational Culture and Adult Learning Theory ........................................ 33
Teacher Attitudes Towards Evaluation ................................ 38
Current Models of Teacher Evaluation .............................. 41
Summary ................................................................. 47

**METHODOLOGY** ....................................................... 51

Purpose of the Study ................................................... 51
Methods of Research .................................................. 52
Description of the Sample ............................................. 53
The Instrument .......................................................... 58
Data Collection .......................................................... 65
Data Analysis ............................................................ 67

**PRESENTATION OF THE FINDINGS** .................................. 69

Introduction ............................................................. 69
Description of Respondents .......................................... 71
Presentation of Data Related to Research Question #1 ........ 72
Research Question #1 ................................................... 72
Summary of the Findings Related to Research Question #1 75
Presentation of the Data Related to Research Question #2 75
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Summary of Means and Standard Deviations Related to the Overall</td>
<td>73</td>
</tr>
<tr>
<td>Quality of the Evaluation Experience</td>
<td></td>
</tr>
<tr>
<td>2. One-Way Analysis of Variance Related to the Overall Quality</td>
<td>74</td>
</tr>
<tr>
<td>of the Evaluation Experience</td>
<td></td>
</tr>
<tr>
<td>3. Summary of Means and Standard Deviations Related to the Overall</td>
<td>76</td>
</tr>
<tr>
<td>Impact of the Evaluation Experience</td>
<td></td>
</tr>
<tr>
<td>4. One-Way Analysis of Variance Related to the Overall Impact of</td>
<td>77</td>
</tr>
<tr>
<td>the Evaluation Experience</td>
<td></td>
</tr>
<tr>
<td>5. Scheffé Test Related to the Overall Impact of the Evaluation</td>
<td>78</td>
</tr>
<tr>
<td>Experience</td>
<td></td>
</tr>
<tr>
<td>6. Correlations Between the Overall Quality of the Evaluation</td>
<td>80</td>
</tr>
<tr>
<td>Experience and the Specific Attributes of Teacher Evaluation</td>
<td></td>
</tr>
<tr>
<td>7. Correlations Between TEACHERS and Overall Quality</td>
<td>82</td>
</tr>
<tr>
<td>8. Correlations Between EVALS and Overall Quality</td>
<td>83</td>
</tr>
<tr>
<td>9. Correlations Between PROCS and Overall Quality</td>
<td>85</td>
</tr>
<tr>
<td>10. Correlations Between FDBACK and Overall Quality</td>
<td>87</td>
</tr>
<tr>
<td>11. Correlations Between CONTEXT and Overall Quality</td>
<td>89</td>
</tr>
<tr>
<td>12. Correlations Between the Overall Impact of the Evaluation</td>
<td>92</td>
</tr>
<tr>
<td>Experience and the Specific Attributes Of Teacher Evaluation</td>
<td></td>
</tr>
<tr>
<td>13. Correlations Between TEACHERS and Overall Impact</td>
<td>94</td>
</tr>
<tr>
<td>14. Correlations Between EVALS and Overall Impact</td>
<td>95</td>
</tr>
</tbody>
</table>
Table

15. Correlations Between PROCS and Overall Impact  
16. Correlations Between FDBACK and Overall Impact  
17. Correlations Between CONTEXT and Overall Impact  
18. Summary of Mean Scores and Standard Deviations Related to the 5 Subscale Areas  
19. One-Way Analysis of Variance Related to the Mean Scores for the Subscales  
20. Scheffé Test Related to the Mean Differences Between Groups on the 5 Subscales  

Page 97  99  101  104  105  106
CHAPTER I

Statement Of The Problem

Background

Current educational literature is replete with the theme that public schools are not teaching the majority of students the skills they will need to operate effectively in a society that is increasingly more complex, more information-based and more multicultural (Darling-Hammond, 1996; Schlechty, 1997). Calls for reform in education to deal with this issue have produced renewed interest in developing standards for teachers at the local, state and national levels. One vehicle used in public schools to address the issue of standards for teaching is the teacher evaluation process. All 50 states have some form of government-mandate for evaluating teachers.

Historically, research in the field reveals that the major purposes of teacher evaluation have been two-fold: to assure that teachers meet a minimum competency level and to promote the professional growth of teachers. The latter purpose occupies most of the current research into teacher
evaluation models. This research has sought to identify the critical attributes of a teacher evaluation system that fosters professional growth with the belief that the professional development of teachers has the potential to improve instruction and thus to address the need to prepare our students for a more complex society.

The research indicates that the teacher evaluation process has had little impact on the improvement of instruction. However, the research has identified elements found in effective evaluation systems that may guide future efforts in developing teacher evaluation systems that encourage teacher growth and the improvement of instruction.

Wise, Darling-Hammond, McLaughlin, and Bernstein (1984) conducted a study of the teacher evaluation systems in 32 districts and concluded that an effective teacher evaluation system needed to be aligned with the educational goals, management style, concept of teaching and community values of the school district. In addition, they found that the purpose of teacher evaluation needed to be clearly stated and then the implementation of a process for conducting teacher evaluation, which matched the purpose, was essential. Top level administrative commitment in terms of logistical and financial resources, a belief in
the utility of the evaluation process and teacher involvement in the development and implementation of the process were additional factors identified as important in an effective evaluation model. A more in-depth study of 4 of the 32 districts identified commonalities, which contributed to the success of their evaluation systems. Organizational commitment, evaluator competence and teacher-administrator collaboration were key factors in the implementation of the evaluation process. The evaluator’s ability to make sound judgments about the quality of teaching and his or her ability to make appropriate and specific recommendations for the improvement of teaching were primary indicators of competence. The belief that the competence of the evaluator is a key factor in the success of a teacher evaluation process is supported by further research (McGreal, 1988; McLaughlin and Pfeifer, 1986; Stiggins and Duke, 1988; Capasso, Monahan, and White 1996). Both teachers and evaluators should be jointly trained in the purposes and procedures of the evaluation process so as to understand each other’s roles. Evaluator feedback should be timely, specific, credible, and perceived as non-punitive.

The work of Wise and associates (1984), and others helped to clarify the contextual nature of the teacher
evaluation process. Teaching behaviors viewed as effective in one community could be viewed as less effective in a community with differing needs. A "one model fits all" definition of effective teaching with codified teaching behaviors that could be evaluated through the use of a checklist has been called into question. The belief that teaching is a complex activity that takes place in unique organizational cultures has led to further discussion of what factors in an organization impact on the teacher evaluation process. Since teachers are strongly influenced by the contextual clues and communications they experience in their school environments, their understanding of their work and their role in a school is largely defined by their interpretation of the contextual clues. If the goal of teacher evaluation is the improvement of instruction, the importance of the teachers' involvement in setting learning goals for themselves that are directed at specific instructional improvements, becomes critical (Rosenholtz, 1991). Glickman, Gordon, and Ross-Gordon (1995) encourage the greater involvement of teachers in setting their own professional growth goals.

Adult learning theory gives support to the importance of teachers playing a key role in the evaluation process. As learners, adults need to be self-directing. They bring
a rich background of experience to the learning situation. Their readiness to learn is effected by a need to solve real-life problems. Immediate application of their learning is important to adults. Teachers have historically expressed dissatisfaction with the evaluation process. They have not viewed the process as an effective motivator for improvement. Teachers point to the lack of quality feedback from supervisors and the lack of understanding of the complexity of the teaching process reflected in the evaluation process. Teachers feel that self evaluation and peer evaluation can play useful roles in fostering professional growth (Stark and Lowther, 1984).

As a response to the need to develop models of teacher evaluation that could foster professional growth, respond to the complex aspects of unique organizational cultures and incorporate best practices identified in the research, the 1980s and 1990s have seen the implementation of a wide variety of teacher evaluation models.

In the State of New Jersey the required teacher evaluation process for public school teachers is outlined in the New Jersey Administrative Code, Title 6 (N.J.A.C. 6:3-4.1 and 6:3-4.3). The Code specifies:
The purposes of this procedure for the observation and evaluation of nontenured teaching staff members shall be to identify deficiencies, extend assistance for the correction of such deficiencies, improve professional competence, provide a basis for recommendations regarding reemployment, and improve the quality of instruction received by the pupils served by the public schools (p. 21).

For tenured teachers the Code specifies that the purpose of teacher evaluation is to: (a) Promote professional excellence and improve the skills of teaching staff members, (b) improve pupil learning and growth, and (c) provide a basis for the review of performance of tenured teaching staff members (p. 23).

The Code requires that non-tenured teachers be observed three times a year. These observations are to be followed by conferences with the evaluator who has prepared a written evaluation of the observation. An additional written annual evaluation summarizing the non-tenured teacher's total performance is to be completed at the end of each school year. The annual evaluation is to include a professional improvement plan developed by the supervisor and the teaching staff member. Tenured teachers are
observed a minimum of once a year. The annual summary conference between the teacher and the evaluator should take place before the annual written evaluation is filed. This conference should include the development of an individual improvement plan.

The intent of the New Jersey Code is to foster improved practice and long term developmental growth in the teaching force. However, the processes and procedures that districts in New Jersey adopt to implement the Code vary from district to district (Capasso, et al., 1996). The degree to which districts have implemented those practices identified in the research on teacher evaluation as effective in promoting professional growth also vary from district to district.

In summary, a number of characteristics appear to be present in effective teacher evaluation models:

1. The district has clearly identified the purpose for teacher evaluation. All parties involved with the development and implementation of the teacher evaluation process understand its purpose.

2. Criteria for evaluation are clear and perceived as relevant and meaningful.

3. The processes and procedures for the implementation of teacher evaluation are clearly aligned to its purpose.
4. The district has allocated sufficient time, training and financial resources to teacher evaluation.

5. Teachers perceive the feedback they receive to be constructive.

6. Teachers are actively involved in setting their learning goals. There is an understanding that learning goals for non-tenured and tenured teachers will be different.

7. Multiple sources of data are used (including self and peer assessment).

8. The focus of evaluation is on teaching.

9. Mutual trust between the teacher and the evaluator is established.

10. Teacher evaluation and staff development are systematically linked.

Purpose of this Study

The purpose of this study was to increase the knowledge base regarding the key attributes of teacher evaluation systems that foster professional growth in terms of effective instructional practices. This study examined the potential of a newly designed teacher evaluation instrument to improve instruction. A k-8 suburban district in central New Jersey implemented a new teacher evaluation form during the 1997-98 school year. The district enrolls approximately 3500 students in 7 buildings (five k-5
elementary schools and two 6-8 middle schools). All non-tenured teachers (n= 43) and a volunteer group of tenured teachers (n= 62) used this new form. The remaining tenured teachers (n= 153) continued to use the old evaluation form for the 1997-98 school year. The new teacher evaluation form and the process for its implementation had been designed for the purpose of promoting professional growth. The purpose of this study was to determine if differences exist among the three groups of teachers: the non-tenured teachers evaluated using the new form, the tenured volunteers evaluated using the new form and the tenured teachers evaluated using the old form in terms of their perceptions of, if, and how the evaluation process fostered their professional growth in terms of effective instructional practices.

Summary of the Problem

The goal of teacher evaluation, both formative and summative, has historically served two purposes: establishing minimum competency for the purposes of personnel decisions and the promotion of professional growth. A review of the literature indicates that the process of teacher evaluation has had very little success with the goal of promoting professional growth (Duke & Stiggins, 1990; Guskey, 1988; Peterson, 1995). Most
recently, a series of alternative methods of teacher evaluation has been developed (e.g., professional growth portfolio, self-assessment, student assessment) and implemented in some public schools. However, the large majority of schools continue to rely on the traditional model of principal as evaluator. In this traditional model, teachers are observed by the principal and their performance is evaluated based on a set of predetermined criteria (usually a checklist format). A conference is usually held in which the principal reviews the completed checklist with the teacher. This model is widely used in New Jersey because such a model is spelled out in the New Jersey Administrative Code. Districts vary in their attempts to make this a meaningful process for professional growth.

Research Questions

This study addressed the following research questions:

1. What perceptions did the teachers in each of the three groups have regarding the quality of their most recent teacher evaluation experience (pre-observation, observation, post-observation, end-of-year evaluation)?

2. Following their most recent teacher evaluation experience, what perceptions did the teachers in each of the three groups have regarding the impact of the
evaluation process on the following: changes in teaching practices, attitudes about teaching, and/or understanding of the teaching process?

3. What relationships did teachers in each of the three groups perceive between specific attributes of teacher evaluation (attributes of the teacher, attributes of the evaluator, attributes of the evaluation procedures, attributes of the feedback on teaching performance and attributes of the evaluation context) and the overall quality of the teacher evaluation experience?

4. What relationships did teachers in each of the three groups perceive between specific attributes of teacher evaluation (see #3 above) and the overall impact (changes in teaching practices, attitudes about teaching, and/or understanding of the teaching profession)?

5. Were there differences among the three groups of teachers in the study regarding research questions 1-4?

Assumptions

The following assumptions were made by the researcher in carrying out the study:

1. Teachers have perceptions about teacher evaluation and supervisory practices.

2. Teachers will accurately report these perceptions.
3. Tenured teachers who have volunteered to be evaluated using the newly designed instrument have done so of their own volition.

4. Efforts will be made during the conduct of the study to assure a non-threatening environment and confidentiality.

Definition of Terms

The following definitions were assigned to terms and concepts in the study:

1. **Teacher evaluation process.** Throughout this study this term refers to a 4-step procedure. First, a preconference occurs prior to the observation of a lesson. The purpose of the preconference is to discuss the nature of the lesson to be observed and any details or aspects of the lesson that will need special attention. Second, the observation is the actual classroom visit, when the evaluator observes for an entire class period or for the duration of a subject lesson. Third, the postconference occurs after the lesson. During this conference the teacher and evaluator discuss the evaluator’s and the teacher’s analysis of the lesson and plan for instructional improvement. Fourth, a written evaluation is completed by the evaluator either before or after the postconference. A written evaluation is completed for
each observation and an annual written evaluation is submitted summarizing a teacher's total annual performance.

2. **Evaluator.** The administrator who has direct supervisory responsibility for a teacher's evaluation. At the K-5 level the evaluator is the building principal. At the 6-8 level the evaluator is the building principal.

3. **Professional growth.** This refers to the real or perceived improvement in a teacher's instructional skills. This improvement can be real or perceived by the teacher and/or the evaluator.

**Limitations of the Study**

The following have been identified as possible limitations of the study:

1. The data may be effected by the Hawthorne Effect of teachers perceiving that they are part of a new and exciting process.

2. The volunteer tenured teachers may have the characteristics of teachers who foster their own professional growth irrespective of the teacher evaluation process.

3. The attitudes that the new and tenured teachers bring to the study regarding teacher evaluation practices may have an effect on the outcome of the data.
4. The results of this study may not be generalizable to other populations.
CHAPTER II

The Review Of The Related Literature

This review of the related literature will include: an historical perspective of teacher evaluation, a brief discussion of the impact of the research on organizational culture and adult learning theory on the development of teacher evaluation models, a review of the research on teacher attitudes towards evaluation practices, and current models of teacher evaluation.

An Historical Perspective

Prior to 1920 the purpose of supervision was to eliminate any inefficiency or incompetence in the teaching force (Glanz, 1991). Rating forms were developed to assist in this objective (Boyce, 1915). Supervision before 1920 was rooted in a bureaucratic and inspectional-type of model. As a result, Glanz (1996) explained:

Supervision of this sort attracted vociferous criticism from teachers and others...The movement to alter supervisory theory and practice to more democratic and improvement foci, while at the same time minimizing the evaluative function, occurred in
the 1920s as a direct result of growing opposition to autocratic supervisory methods. More fundamentally, however, the rhetoric of democratic supervision now focused on making supervision more palatable and acceptable among teachers (p. 6).

Glanz (1996) viewed the past three quarters of this century in supervision as an effort to shift away from the bureaucratic, directive approach developed prior to the 1920s and more toward a model that reflected the potentially constructive aspects of teacher evaluation and supervision. Although this shift was underway, Darling-Hammond (1996) pointed out:

As recently as 10 years ago, the idea that teacher knowledge was critical for educational improvement had little currency. Continuing a tradition begun at the turn of the 20th century, policymakers searched for the right set of test prescriptions, textbook adoptions, and curriculum directives to be packaged and mandated to guide practice. Educational reform was 'teacher proofed' with hundreds of pieces of legislation and thousands of discrete regulations prescribing what educators should do (p.5)
Tuckman and Oliver (1968), Scriven (1973), Lortie (1975), Owens (1991), Bridges (1992) and Wang, Haertel and Walberg (1993) discussed the major concepts and perspectives that underlaid the design and use of teacher evaluation systems. The purposes of teacher evaluation included determining the value or merit of teaching, reassuring practitioners and their audiences, making staffing decisions, and the improvement of instruction.

Wise, and associates (1984) conducted a study of 32 districts in a broad range of rural, suburban and urban communities. These districts all had reputably well-developed teacher evaluation instruments. Wise and associates (1984) described a dual purpose for teacher evaluation: to help facilitate decisions about teacher status and to help teachers improve their performance. They concluded:

1. To succeed, a teacher evaluation system must suit the educational goals, management style, conception of teaching, and community of values of the school district.

2. Top-level commitment to and resources for evaluation outweigh checklists and procedures.
3. The school district should decide the main purpose of its teacher evaluation system and then match the process to the purpose.

4. To sustain resource commitment and political support, teacher evaluation must be seen to have utility. Utility depends on the efficient use of resources to achieve reliability, validity, and cost-effectiveness.

5. Teacher involvement and responsibility improve the quality of teacher evaluation (p.xi-xiii).

Wise and associates (1984) expanded the view of what factors needed to be included in the development of an effective evaluation/supervision model. They concluded that the usefulness of a teacher evaluation system would largely depend on the degree to which the system was congruent with what it sought to measure and on whether or not a district could tolerate the logistical and financial costs. They recognized that the usefulness of any given approach to teacher evaluation was dependent on changes in the political and organizational climates as well as on changes in the purposes of teacher evaluation. These researchers were among the first to discuss the importance of teacher involvement in the improvement of the teacher evaluation process. The idea that teachers could
contribute to the evaluation process was a shift from the model of supervision and evaluation that was in place in the 1920s.

As part of their study, Wise and associates (1984) selected 4 districts from among the 32 involved in the study to be case studies. These 4 districts represented diverse teacher evaluation processes and different organizational environments. Although different in a number of ways, Wise and associates (1984) identified commonalities in each of these four districts, which contributed to the success of their evaluation systems. These commonalities were found in the implementation of the evaluation systems. Organizational commitment, evaluator competence, teacher-administrator collaboration and strategic compatibility were key factors in the implementation of the evaluation process. These districts provided time for evaluation. They focused on developing the evaluator's ability to make sound judgments about the quality of teaching and on the ability to make appropriate and specific recommendations for the improvement of teaching. All 4 of these districts also had mechanisms for verifying the accuracy of the evaluators' reports about teachers. Mechanisms were also in place to provide for collaboration between teachers and administrators about the
evaluation process. Lastly, teacher evaluation was viewed as part of a larger strategy for school improvement.

Further support for viewing teacher evaluation contextually was provided by the work of Brophy and Evertson (1974), Cronbach and Snow (1977) and Peterson (1995) who found that effective teaching behaviors could vary for students of different socioeconomic backgrounds or for students with certain mental or psychological characteristics. Gage (1978) and McDonald and Elias (1976) found the same true for students in different grade levels or in different subject areas. Peterson and Kauchak (1982) and Soar (1972) posited that teaching practices that were proven effective when used in moderation in one situation could actually produce significant negative results when used in another. A “one model fits” all concept of effective teaching was called into question.

The consensus of the research shifted from a belief that teaching behaviors could be codified or that there was any one set of effective teaching behaviors to a belief that teaching was a context-dependent, complex phenomenon. The focus moved from evaluation for the purpose of accountability to evaluation for the purposes of improvement of instruction and professional growth. The shift from a focus on a bureaucratic, directive model of
teacher supervision whose goal was to ferret out incompetence to a model which included a view of supervision which examined the professional growth of teachers was definitely reflected in the literature about teacher evaluation and supervision by the 1980s.

The definition of good teaching and efforts to codify good teaching behaviors has continued to be problematic for researchers. After reviewing a dozen studies from 1921 to 1959 Medley and Coker (1987) concluded that the correlation between a principal’s rating of a teacher and direct measures of teacher effectiveness were near zero. Scriven (1981) discussed the difficulty of trying to evaluate good teaching from the limited perspective of just a few visits to the classroom. The definition of good teaching needed to include indicators that moved beyond the classroom into the context of the school as a whole.

At one point, popular belief was that one could measure good teaching by correlating it to student achievement. However, Berk (1988), who recounted 70 years of research in teacher evaluation showed the link between teacher performance and pupil learning to be indirect. He discussed 50 other factors, including pupil effort and parent support, as factors beyond the control of teachers. Lieberman and Miller (1992) commented:
The sad fact is that, as a profession, we have not been able to codify teaching under a variety of contingencies in a way that is satisfying to practitioners. The knowledge base in teaching is weak: there is simply no consensus (as there is in medicine and law) about what is basic to the practice of the profession (p. 3).

Darling-Hammond (1990) viewed the efforts in teacher evaluation in the 1970s and 1980s to be focused on improving curriculum and programs. She viewed the development by many states of teacher competency tests as designed to check on whether teachers were effectively delivering these curricula and programs. She observed a new shift gain momentum in the 1980s and projected a direction for the 1990s:

Teacher professionalism and school restructuring are the major watchwords used to describe efforts to reform teaching and schooling so that they will focus more directly on learners' needs... These initiatives rely on teacher evaluation for a wide variety of purposes, including selection, training, improvement, and advancement. They often envision broader roles for teachers in evaluation and for evaluation (p. 17).
Throughout the 1990s much of the research in teacher evaluation has focused on the role that the teacher evaluation process could play in the professional improvement of teachers. Duke and Stiggins (1990) reviewed the literature on teacher evaluation for the period 1981-1988 to discover 95 citations that discussed teacher evaluation and professional growth. However, only three of these citations dealt with professional growth beyond minimal standards and none of them included any empirical investigation. Duke and Stiggins (1990) conducted three studies to find out what conditions in the teacher evaluation process would most likely foster the professional development of competent teachers. Their first study involved case studies of 4 teacher evaluation systems. They concluded that none of these highly structured systems promoted the improvement of teachers' skills. Their second study involved the collection of narrative data from 33 teachers who reported experiencing professional growth as a result of teacher evaluation. From this narrative data, Duke and Stiggins (1990) concluded that teacher attributes, evaluator attributes, attributes of the information gathered on teaching performance, the attributes of the feedback on teaching performance, and the attributes of the evaluation context
could play a role in professional growth. Based on these findings, Duke and Stiggins (1990) developed the Teacher Evaluation Profile (TEP) Questionnaire. Their third study was used to validate the TEP. A pilot test of the questionnaire with 450 teachers in 5 districts determined that the questionnaire provided valid data regarding teacher perceptions "...and that a strong relationship existed between specific attributes of an evaluation system and the growth-oriented outcomes of that system as perceived by teachers" (p.121). Their research indicated that some teachers would be unwilling to take advantage of the opportunities for professional growth. Others, however, who had the following characteristics would benefit from evaluation that included opportunities for professional growth: strong professional expectations, a positive orientation to risk taking, openness to change, willingness to experiment in class, openness to criticism, strong knowledge of technical aspects of teaching, strong knowledge of subject matter, some positive prior experience with teacher evaluation (p. 121).

The evaluator attributes identified by this third study as contributing to the professional growth of teachers were: credibility as a source of information,
interpersonal communication skills, persuasiveness, patience, trustworthiness, a track record for being helpful, and an ability to model.

McGreal (1988) confirmed Duke and Stiggins (1990) findings: "The bottom line of effective evaluation is the quality of what happens when the administrator and teacher get together" (p. 3). McGreal believed that if an evaluation system was designed around effective supervisory and teaching behaviors and if appropriate training was provided to administrators and to teachers, then a single system could serve multiple purposes.

McLaughlin and Pfeifer's (1986) 4 case studies of highly regarded district teacher evaluation systems further confirmed the value of the evaluator who can provide sound, growth-producing feedback.


More recently the work of Bulach, Boothe and Pickett (1998) further confirmed Duke and Stiggins (1990) research. They surveyed 375 practicing teachers in 23 graduate classes. These teachers were asked to identify and then
rank the most harmful mistakes their principals made as
supervisors. They identified the top 5 mistakes:
ineffective human relations, poor interpersonal
communications, lack of educational priorities, avoiding
conflict, and lack of knowledge about instruction and
curriculum.

Duke and Stiggins (1990) studies also raised the
question of whether or not any single teacher evaluation
model could meet the dual purpose of teacher evaluation:
professional development and professional accountability.
Although McLaughlin and Pfeifer (1988) had collected some
case study data on exemplary systems that presumably served
both purposes, Duke and Stiggins (1990) based their
explanation of this difference of opinion on two different
definitions of growth:

Teachers may grow toward greater competence, as
represented by some set of performance standards. But
few expand their repertoire of skills, develop new
perspectives on old problems, or become virtuoso
performers in specific areas of teaching. This latter
type of professional development may not be possible
within the confines of evaluation systems serving the
accountability purpose (p. 127).
Peterson (1995) provided a review of the research in teacher evaluation. He indicated that 70 years of empirical research on teacher evaluation had resulted in evidence that current practices neither improved teacher performance or accurately reflected what was happening in the classroom. Peterson (1995) cited Medley and Coker (1987): "To this day, almost all educational personnel decisions are based on judgments which, according to the research, are only slightly more accurate than they would be if they were based on pure chance" (p.12). Peterson (1995) proposed field-tested, innovative practices for improving the state of teacher evaluation. He proposed that these practices included gathering of information regarding what aspects of good teaching already exist rather than focusing solely on a deficiency model. He suggested further that the purposes for teacher evaluation be clarified and that teachers play a more central role in the evaluation process. He also encouraged the use of more than one person to judge teacher performance and the collection of multiple sources of evaluation data. Peterson (1995) recommended that the teacher evaluation process should:
Attend to the sociology of teacher evaluation. Current practice is to ignore the powerful effects of expectations, roles, rewards, sanctions, and relationships in the workplace. Evaluation is done as simple actions of feedback and retention decisions, rather than as a complex, human, organizational interaction and transaction (p. 9).

Peterson's (1995) work expanded the field of teacher evaluation by proposing a model, which he referred to as Quality Teaching as an Emergent Phenomenon. He defined emergent evaluation as:

The difference between discrepancy and emergent evaluation is fundamental in how teachers are judged. In discrepancy evaluation, the design is to have one means by which all teaching is evaluated, or, if multiple measures are used to specify a weighting among measures to prespecify quality. In the emergent view, the evaluation effort is to document the practice or performance after it has happened, and then to recognize it and describe its value through human judgments (p.45).

Peterson’s (1995) proposed emergent evaluation model downplayed the principal-based checklist routine in favor
of the use of multiple and varied data sources (parents, peers, student performance, self-evaluation, artifact collection), increased teacher control and participation and the use of panel decision-making bodies.

This expanded view of supervision and evaluation was reflected in the work of Glickman et al. (1995). They reviewed the wave of effective schools research done in the 1970s and the wave that followed in the 1980s. They concluded that a developmental model of supervision was appropriate. This model encouraged greater involvement, autonomous thinking, and collective action by teachers. Glickman and associates reported research which showed that schools that developed links between instruction, classroom management, and discipline and staff development, direct assistance to teachers, curriculum development, group development, and action research for the attainment of a common goal or purpose were schools that achieved their objectives. Glickman and associates (1995) posed the following:

Our own supervisory platform is based on the premise that human development is the aim of education. Therefore, supervision should be eclectic in practice, directed toward the goal of nondirective, existentialist supervision. Our goal as supervisors
is eventually to return control to the teaching faculty to decide on collective, instructional improvements...Regardless of the entry point, the supervisor should always strive to shift control to teachers... (p. 102-103).

Glanz (1996) supported Glickman's and associates model of supervision that emphasized a variety of supervisory approaches (nondirective, collaborative, and directive methods). The direction of the 1990s for teacher supervision and evaluation was a model that incorporates both directive and nondirective measures selected in the context of the complex environment of schools.

Harrington-Lueker (1996) reported on the current trend to view teacher evaluation as a long-term process in which different kinds of assessments were appropriate at different points in a teacher's career.

Districts in California began to develop models of teacher evaluation that offered remediation components including peer support, allocation of resources and district wide as well as county wide committee review (McLaughlin and Pfeifer, 1988).

State and national efforts called for the reform of teacher evaluation. All 50 states had some form of government-mandated procedure for evaluating teachers.
This legislation was intended to foster improved practice and professional growth (Capasso, Monahan and White, 1996). The National Commission on Teaching and America’s Future in its report *What Matters Most: Teaching for America’s Future* (1996) called for standards for both teachers and students, the reinvention of both teacher preparation and professional development, improved teacher recruitment practices, the encouragement of teacher knowledge and skill, and schools that were organized for both student and teacher success.

The National Council for Accreditation of Teacher Education (NCATE), the Council of Chief State School Officers (CCSSO), and the National Board for Professional Standards (NBPTS) were all in the process of developing and implementing standards and assessments that reflect the skills and knowledge that were needed both for state licensing and for the continuous process of professional development (Wise, 1996).

The Center for Research on Educational Accountability and Teacher Evaluation (CREATE), a federally funded research center at Western Michigan University, assessed the strengths and weaknesses of several new methods of teacher evaluation, including self-evaluation instruments (Harrington-Lueker, 1996).
Manatt and Benway (1998) reported on the work of the School Improvement Model research team (SIM) at Iowa State University. SIM has been working with selected districts in the development of a new approach to teacher evaluation called team evaluation or 360-degree feedback. In this model multiple sources of feedback are used: supervisor feedback, self-evaluation, student achievement, peer feedback, student feedback, and parent feedback.

Another key influence on the research in teacher evaluation has been the impact of the constructivist model of learning. This model changed the way teachers view learning and instruction. Brooks and Brooks (1993), Smith (1993) and Walen and DeRose (1993) reported on the inadequacy of the traditional checklist model of teacher evaluation in relationship to the constructivist theory. Searfoss and Billie (1996) concluded:

In our recent work in 20 schools, we observed that teachers engaged in holistic practice received far less feedback in their evaluation, as the philosophical mismatch between holistic/constructivistic and direct instruction approaches hit head-on (p. 38).
Impact of Research on Organizational Culture and Adult Learning Theory

The research on teacher evaluation evolved to include an understanding that teaching is a contextually dependent activity. The context in which teachers worked was the individual school site. Each school site had its own unique organizational culture. The 1983 Rand study completed by Wise and associates (1984) not only examined the instruments and procedures used for teacher evaluation, but also the implementation processes and the organizational contexts in which these instruments and procedures were used. Wise and associates (1984) emphasized that an effective teacher evaluation system needed to be responsive to both individual teachers and the organizational culture. An understanding that school cultures differed supported the premise that a "one model fits all" approach to teacher evaluation was inappropriate.

Not only were school cultures reflective of individual teachers and students, but they were also reflective of the educational goals established by these cultures. Wise and associates (1984) found:

...as one ascribes different degrees of generalizability to effective teaching behaviors and different weights to context-specific variables, one implicitly embodies
different conceptions of teaching. The more complex and variable one considers the educational environment, the more one relies on teacher judgment to guide the activities of classroom life and the less one relies on generalized rules for teacher behavior (p.10).

A teacher evaluation system must match the educational goals, management style, concept of teaching, and the values of the school community (Wise et al., 1984). McLaughlin and Pfeifer (1988) confirmed that:

Teacher behaviors and activities interact with numerous factors to shape student performance. Student socioeconomic status, school climate, pupil abilities, previous instructional treatment, family life and home conditions are but a few of the many factors that influence a student’s performance. ‘Teacher effectiveness,’ however defined, is highly contextual and conditional (p.2).

The belief that teaching was a complex activity that took place in unique organizational cultures led to an expansion of the concept of teacher evaluation. A single model, a single evaluator, a single set of data, little involvement of the teacher no longer seemed viable components of an effective teacher evaluation process. In
her 1991 work *Teachers' Workplace: The Social Organization of Schools*, Rosenholtz studied 78 elementary schools in 8 Tennessee districts. Her study was based on the assumption that teachers' own definitions of their work was guided by how they sought to make sense of the world around them. She believed that teachers were strongly influenced by the contextual clues and communications they experienced in their school environments. Her research was valuable because it not only confirmed earlier research regarding the contextually dependent nature of teaching but it raised questions about what could be changed in organizational cultures to bring about desired outcomes, in this case, the improvement of teaching. Rosenholtz's study indicated the importance of teachers' involvement in setting learning goals for themselves that were directed at specific instructional objectives, the need for frequent evaluation by principals who had identified specific improvement areas and monitored teachers' progress in achieving them, and frequent opportunities for teachers to share learning goals and to collaborate on how to achieve them. Rosenholtz concluded:

> In sum, teachers' regard for their work—their sense of optimism, hope and commitment—tends to reside in workplace conditions that enable them to feel
professionally empowered and self-fulfilled, that keep them reaching for new teaching challenges, fresh opportunities, and ever-expanding technical knowledge (p.165).

Glickman and associates (1995) discussed the role of supervision in the work culture of the school. Their model of developmental supervision called for the encouragement of greater involvement, autonomous thinking and collective action by teachers. "If supervision is to improve instruction, it must reshape norms and beliefs about the work culture of schools..." (p.28).

Teachers as learners was a more recent view of the teacher evaluation process. Adult learning theory shed some light on what conditions might foster the learning environment for teachers. Malcolm Knowles (1980) popularized the theory of andragogy. There are 4 basic tenents of this theory: (a) Adults have a need to be self-directing, (b) adults bring a rich background of experience to the learning situation, which should be tapped, (c) an adult's readiness to learn is effected by a need to solve real-life problems, and (d) adults are concerned about making immediate application of their learning.

Ellis and Bernhardt (1988) studied a random sample of 425 classroom teachers from public elementary and secondary
schools within 24 school districts in Fairfield County, CT. and found that "Teachers in the high andragogical supervision group were significantly more satisfied with their jobs and had higher levels of internal work motivation than were teachers whose supervisors are not perceived to use andragogical behaviors" (p. 15).

Fullan (1991) wrote about educational change being a learning experience for teachers. If so, principles of adult learning had particular relevance for teacher evaluation models that included the improvement of instruction and professional growth as goals.

Mezirow's (1990) theory of perspective transformation shed additional light on adult learning theory. Transformative learning occurred for adults following a disorienting dilemma or a series of events that changed the routines of life. Adults were called to examine the assumptions that guide their decisions and their actions. Reflection and critical thinking were important parts of adult learning in Mezirow's theory.

Glickman and associates (1995) contrasted the research on adult learning theory to conclude:

Effective supervision responds to the principles of adult learning. Teachers' learning should be related to their experiences, needs, and learning strengths;
should include opportunities for collaborative action, reflection, and critical thinking; and should be directed toward teacher empowerment (p. 78).

Teacher Attitudes Towards Evaluation

Historically, teachers have mistrusted evaluation (Wolf, 1991 and Johnson, 1990). Teachers viewed their ratings to be too dependent on the idiosyncrasies of the raters. For the most part teachers saw nothing to be gained from evaluation. They did not view it as an effective motivator for improvement. Teachers found supervision practices to be largely a waste of time and described their relationships with supervisors as being like a cold war (Blumberg, 1980). Wise and associates (1984) in their survey of 32 districts found two major problems: "...principals lacked sufficient resolve and competence to evaluate accurately...Teacher resistance or apathy was the second most frequently cited problem" (p. 22).

One of the major themes in the research on teacher attitudes towards evaluation was that the models in practice had not kept pace with the new understandings about the complexities of the evaluation process. More recently evaluation was viewed as serving multiple purposes
ranging from accountability to instructional improvement to professional growth. Evaluation took place in unique school cultures which reflected the values of the culture as well as the beliefs and values of the teachers who work in the school. Recent evaluation models included a more expanded role for teacher involvement. However, McLaughlin and Pfeifer (1988) pointed out that "Claims for teacher evaluation as an aid to improvement fall on cynical ears. Teacher evaluation, teachers remind policymakers, has no tradition as a strategy to foster improvement. Instead, inspection and control have characterized teacher evaluation since colonial times" (p. 1).

In the spring of 1994 the National Center for Education Statistics reported on a nationally representative survey of approximately 1000 elementary school teachers:

A majority of teachers reported that formative goals, that is, goals associated with professional development...should be an objective to a great extent in teacher performance evaluations. However, approximately 20% fewer teachers reported that each of these 4 goals had been an objective to a great extent at their school when they were last evaluated (p. 11).
Another aspect of the evaluation process that teachers expressed concern about was the quality of feedback they received. Teachers believed that the quality of feedback was closely tied to the professional growth and instructional improvement aspects of evaluation. Capasso, et al. (1996) reported on preliminary research data regarding the New Jersey Administrative Code for the evaluation of professionally certified school employees. Their sample included 225 teachers from 8 counties in southern New Jersey. Since one of the purposes of the Code was to provide meaningful feedback to teachers, these researchers sought to determine teachers' impressions of the feedback they received. They found:

The survey data suggests that respondents were not very enthusiastic about the quality of the feedback they received during their post-observation or annual summary conference. As can be seen, 29% indicated that the quality of feedback during post-observation conferences was meaningless. Twenty-six percent indicated that the quality of feedback during their annual summary conference was meaningless (p. 21).

Teachers were also critical of evaluation practices because of their inability to reflect the complex nature of teaching. Searfoss and Billie (1996) concluded that: "All
of the teachers expressed anger and disappointment that the ‘quality of discovery learning and student interactions was discounted’...Teachers also felt that they missed opportunities for meaningful feedback and collegial discussions about the complex pedagogy of their practice” (p. 39).

Duke and Stiggins (1990) reported on teachers' ratings of the quality and impact of teacher evaluation. They found that a teacher's rating of evaluation was highly dependent on the perceived skill, integrity and caring of the evaluator.

Stark and Lowther (1984) described the results of a survey of nearly 1000 teachers concerning the evaluation process. Eighty-nine percent of the teachers agreed that self-assessment was the most appropriate method of evaluation, 85% viewed administrator judgments as also appropriate, 75% selected teacher/peer assessment as appropriate, and 79% of the teachers found student and parent judgment as inappropriate.

Current Models of Teacher Evaluation

The literature remained consistent in identifying the improvement of instruction and the enhancement of teachers' professional growth as the goals of supervision and evaluation. Historically, accountability concerns have
overpowered both these goals. During the 1980s and 1990s teacher evaluation and professional growth appeared to be more closely linked to discussions about school reform and restructuring. Changes in organizational climate have brought about changes in rules, roles and responsibilities of the various stakeholders in the evaluation process.

Nationwide support endorsing teacher evaluation for formative purposes prompted many school districts to add a professional development component to their standard evaluation systems (Stiggins and Duke, 1988).

The 1980s and 1990s saw the advent of the performance objectives approach offered by Redfern (1980), the goal-setting models proposed by Iwanicki (1981), Glatthorn’s (1984) differentiated supervision model, cognitive coaching (Costa and Garmston, 1994), developmental supervision (Glickman et al., 1995), and the professional portfolio.

In the performance objectives approach, objectives were established based on teacher needs; action plans were developed and implemented; and, results were collected, assessed and discussed. Evaluative judgments were grounded by evidences, which when monitored were superior to assumptions and unsupported opinions often associated with the traditional supervisory model.
The goal-setting model first involved the teacher in a process of self-evaluation. Next, the teacher and supervisor would pre-conference to develop a goal-setting contract. Near the end of the evaluation cycle, the teacher and supervisor would meet again to assess goal accomplishments and to plan future directions for improvement.

The differentiated model of supervision posited that no one way of evaluation appeared appropriate for all members of a school staff. One of the growth-oriented options suggested by Glatthorn (1984) was clinical supervision. The prescribed 5 stages of clinical supervision (pre-observation conference, observation, analysis and strategy, supervision conference, and post-conference analysis) encouraged reflection and analysis, both essential to professional growth.

Another option offered by Glatthorn (1984) was a model described as self-directed development. It provided opportunities for teachers to work independently to carry out a systematic plan for their own professional growth. The principal served as a key resource person to help the teacher develop the plan, find resources, and assess progress.
The cognitive coaching model created a supervision strategy that focused on enhancing the teachers' capacity for and practice of critically self-reflective teaching.

The recognition of the uniqueness of teachers' developmental needs led the field of teacher evaluation to continually seek new models that more effectively placed an emphasis on self-directed learning. The professional portfolio was one of these models.

Bird (1990) offered a context for portfolios in the teaching profession. Portfolios were kept by artists and architects who purpose was to exhibit their work. These portfolios contained a selected sample of one's finest work. Portfolios were kept by lawyers and social workers as works in progress. Bird suggested that the material in the teacher's portfolio would include material about a specific problem that a teacher had chosen to investigate at a given time. If entries in the portfolio were produced mainly by the teacher, they might included elective entries, guided entries, or required entries. If the portfolio entries were jointly produced, they included collegial projects, negotiated entries or proctored entries. If the portfolio entries are mainly produced by someone other than the teacher, they included commentary (parent letter), attestations (ratings or observations), or
official records (license, diploma). Bird posited: "The main attraction of portfolios appears to be that they might provide opportunities to see a schoolteacher's work relatively whole and in context..." (p.248).

Shulman (1988) and Wolf (1991) and their colleagues participated in the Teacher Assessment Project at Stanford University (TAP). The project was a 4-year effort to develop new approaches to teacher evaluation. The resulting prototypes and methods were intended to serve as resources for the National Board for Professional Teaching Standards, which wanted to create a voluntary program for national teaching certification. TAP focused on two approaches: simulation exercises done at an assessment center and portfolios. The portfolios were designed to capture the complexities of teaching by documenting the process of teaching and learning over time in the context of the classroom.

Teachers involved in the project followed a sequence of events during the investigation of portfolios. First, teacher teams identified several critical teaching tasks in their content areas. Next, the teams designed portfolio entries related to these tasks which involved teacher collaboration and peer evaluation. The portfolios contained examples of student work, teacher-developed lesson plans,
videotapes of teaching, and teacher reflections on his or her own teaching (Wolf, 1991). Prior to the TAP study Shulman and his colleagues attributed the ineffectiveness of portfolio designs to the unwieldy collection of nonselective content.

The Professional Development Portfolio as designed by Dietz (1994) focused on facilitating teachers’ learning and increasing the impact on learning in the classroom. Dietz posited that professional development activities were more effective if professionals were involved in setting their own goals, choosing a method for learning and deciding how best to integrate new learning. She viewed change as a process of resocialization that took place over time and that requires interaction.

Dietz (1994) described the 4 features of the portfolio: a purpose which examined why the portfolio was being done; a focus which described the theme for learning; a process which told how collaboration, learning, planning and reflecting would take place; and, the outcomes which described conclusion, reflections and some exhibits of the learning.

The implementation of the portfolio occurred in 4 phases. In phase one, the teachers wrote a credo reflecting their beliefs about their profession and about
how individuals learned. In phase two, participants met with their portfolio partners to develop banner questions, which reflected a point of inquiry for the teachers. The third phase involved collaborative support for the development of the growth plan. The final phase of the Professional Development Portfolio provided a context for summarizing, concluding, and rethinking one's progress with the portfolio process. The portfolio and the learning process was shared with others and a final evaluation of the process was recorded.

Brogan (1996) advocated for the use of portfolio-type assessment because it provided clear evidence of what was important to teachers and motivated teachers to be selective in identifying what they valued as teachers and as members of the school community.

Summary

The evaluation of teachers has been in practice since before the turn of the century. Despite attempts to make the evaluation process a meaningful one for teachers, the research indicated that there is little evidence that teachers benefited from the process. Early models of evaluation focused on eliminating incompetence. Later models attempted to measure how well teachers followed prescribed curricula and programs or how well teachers
performed against a set of codified teaching behaviors. Teacher evaluation was a low stakes activity with the focus on weeding out the ineffective teacher rather than on promoting the growth of the competent teacher.

Later models like differentiated supervision, developmental supervision and cognitive coaching attempted to deal with evaluation for the purposes of improving instruction and promoting professional growth. These models added the components of inquiry, reflection and collaboration.

As the literature on adult learning theory and organizational culture began to inform the teacher evaluation process, the need for teachers to become active participants in the evaluation process was highlighted. Teachers as learners had individual needs that would require evaluation models to become more complex and collaborative. Teacher evaluation would become a high stakes activity if instruction and professional growth could be impacted.

An increased understanding of the nature of learning impacted the field of teacher evaluation. A constructivist model of learning gained popularity in many classrooms. Matching teacher performance to prescribed set of teaching behaviors was no longer an appropriate approach.
Teachers have continually expressed a desire for meaningful feedback about their instructional practices. However, the traditional accountability procedures of checklists and tallies have provided them with little useful feedback.

As school reform efforts began to focus on improvement at the individual school site, teacher evaluation systems needed to be aligned to the educational goals of the school site and to the conversations teachers and supervisors were having about the nature of teaching and learning.

Several current efforts hold promise for the reform of teacher evaluation: rethinking professional development, involving teachers in research, collaborative inquiry the use of multiple sources of evaluation data, and standard-setting in the profession.

The model for teacher evaluation most commonly used continues to be the principal as evaluator. The vehicle used for evaluation continues to be a checklist of codified behaviors, which is completed after the principal conducts a classroom observation. However, the research revealed that the critical factor in the success of teacher evaluation is not the instrument that is used for evaluation purposes. The critical factors lie in complex organizational factors that include the attributes of the
teachers, the attributes of the evaluators, the evaluation procedures, the attributes of the feedback teachers receive and the characteristics of the context in which evaluation occurs. The research revealed the need to continue exploration of those factors in the teacher evaluation process that can contribute to professional growth.
CHAPTER III

Methodology

Purpose of the Study

The purpose of this study was to increase the knowledge base regarding the key attributes of teacher evaluation systems that foster professional growth in terms of effective instructional practices. This study examined the potential of a newly designed teacher evaluation process to improve instruction. A k-8 suburban district in central New Jersey implemented a new teacher evaluation form during the 1997-98 school year. The district enrolled approximately 3500 students in 7 buildings (five k-5 elementary schools and two 6-8 middle schools). All non-tenured teachers (n= 43) and a volunteer group of tenured teachers (n= 62) were evaluated using this new form. The remaining tenured teachers (n= 153) were evaluated with the old evaluation form for the 1997-98 school year. The new teacher evaluation form and the process for its implementation were designed for the purpose of promoting professional growth. The purpose of this study was to determine if differences existed among the three groups of
teachers: the non-tenured teachers evaluated using the new form, the tenured volunteers evaluated using the new form and the tenured teachers evaluated using the old form in terms of their perceptions of, if and how the evaluation process fostered their professional growth in terms of effective instructional practices.

Methods of Research

The research methodology in this study was quantitative, as reflected in the reporting of the data gathered from the survey instrument, The Teacher Evaluation Profile (TEP). The survey data was summarized in three forms for each of the three groups: a frequency distribution summary, a profile of means and standard deviations and a correlational analysis. The correlational analysis will examine: (a) the relationship between the ratings of the teachers in each of the three groups for each of the items on the survey and their ratings of the overall quality and overall impact of their most recent evaluation experience, and (b) the relationship between the ratings of the teachers in each of the three groups in each of the survey subscale areas (attributes of the teacher, attributes of the evaluator, attributes of the evaluation procedures, attributes of the feedback on teaching performance and attributes of the
evaluation context) on the survey and their ratings of the overall quality and impact of their most recent evaluation experiences. This aspect of the study may be classified as descriptive research. Leedy (1997) states, "The branch of statistics that describes what data looks like—where their center is, how broadly they are spread, and how they are related in terms of one aspect to another aspect of the same data—is called descriptive statistics (p. 252)." A one-way analysis of variance was used to investigate differences among the mean scores of the groups in terms of their ratings of overall quality and impact and in terms of their ratings on the subscales.

A survey research method was used to answer the research questions posed in the study. This method allowed the researcher to gather data from a relatively large sample.

Description of the Sample

The school district involved in the study was a suburban, public school district located in central New Jersey. The district served a student population of 3500 students in kindergarten through grade 8 in 7 school buildings (five k-5 elementary schools and two 6-8 middle schools). Each school building was administered by a building principal who had direct responsibility for the
observation and evaluation of the building’s professional staff.

The professional staff in the district were evaluated according to the procedures required in The New Jersey Administrative Code, Title 6 (N.J.A.C. 6:3-4.1 and 6:3-3.3). The Code required non-tenured teachers to be observed three times a year. Tenured teachers were to be observed one time a year. Written annual evaluations, which included individual improvement plans, were completed for all staff at the end of each school year. In the district under study, all observations were preceded by a conference between the teacher and the building principal to discuss the goals and objectives of the lesson to be observed. Following the observation, postconferences were held to discuss the observation, to set goals for improvement and to review the written observation report.

The school district involved in this study began a revision of its teacher evaluation system in 1996-1997. A 15-member committee was charged with developing a revised teacher evaluation system that would reflect both a growth model and the district’s commitment to excellence. The committee included teachers, administrators, guidance personnel, nurses, parents and Board of Education members. The committee developed a mission statement for the newly
revised teacher evaluation process, which held the promotion of professional growth and excellence as the primary goals of the evaluation process (Appendix B). The ultimate aim of the new process was to be the realization of effective teaching and learning. There were 7 objectives (Appendix B) identified for the new process:

1. To provide a professional profile which establishes criteria for effective teaching, reflective of acknowledged principles of learning and methods of instruction.

2. To develop and implement procedures that support collaboration, collegiality, and communication within the educational team.

3. To develop a model which encourages observation, analysis, questioning, and conferencing skills as a means of promoting reflection and self-analysis.

4. To provide training for the supervision/evaluation process.

5. To develop a comprehensive supervision plan which includes the opportunity to utilize multiple measures and divergent/alternative means of assessment.

6. To develop a comprehensive evaluation plan which specifies the procedures for the performance review of staff.
7. To design procedures and monitor and evaluate the supervision and evaluation process.

The committee developed several new instruments to be used for the first time during the 1997-98 school year. These evaluation instruments included a Teacher Observation Report and an Annual Performance Summary (Appendix B). As a part of the Annual Summary process, an optional Teacher’s Supplemental Comments sheet and a Summary of the Professional Improvement Plan for the year were included. The Observation Report and the Annual Summary included sets of descriptors and indicators, which addressed a teacher’s competence in the areas of curriculum, instruction and assessment, classroom environment and management, and professionalism. The descriptors identified desirable qualities, behaviors or activities on the part of the teacher. Indicators specified the behaviors that gave evidence of the descriptors. For each indicator teachers could be ranked at one of five levels: (a) U for Unsatisfactory indicated absence of requisite competence, (b) B for Basic indicated fundamental competence, (c) P for Proficient indicated thorough competence, (d) E for Expert indicated extraordinary proficiency, (e) NO/A for Not observed/Not applicable.
During the summer of 1997 district administrators received training in the use of the newly developed evaluation tools. Twelve additional training sessions for administrators were scheduled for the 1997-98 school year. These sessions focused on training administrators in the essential elements of instruction as identified by Madeleine Hunter’s Instructional Theory Into Practice Model.

All non-tenured teachers and all volunteer tenured teachers who would be evaluated using the new tools received training on the new instruments in the fall of 1997. The proposed plan for implementing the new evaluation instruments provided for them to be used with all non-tenured teachers and volunteer tenured teachers during the 1997-98 school year. During the 1997-98 school year, all remaining staff would be evaluated using the old evaluation tool (Appendix C). During the 1998-99 school year, all staff, both non-tenured and tenured would be evaluated using the new tools. The committee planned to gather feedback about the new instruments during the pilot year, 1997-98, and incorporate any needed revisions for the 1998-99 school year.

For the purposes of this study all teachers in the school district were surveyed. Such a sample provided data that
was useful in comparing results within and among the three groups included in the study.

The Instrument

This researcher used the Teacher Evaluation Profile (TEP) to collect data for the study (see Appendix A). The TEP is a 44-item questionnaire developed by Stiggins and Duke (1988). This questionnaire was selected because it addressed a number of the elements identified in the research as related to growth-producing teacher evaluation. The use of the TEP allowed the researcher to address the research questions posed in this study and to analyze the elements of the teacher evaluation environment of this district under study. Stiggins and Duke's (1988) research leading to the development of the TEP was conducted with the purpose of finding out why evaluations have failed to improve teaching and how evaluation procedures might be changed to promote teacher growth and development. They conducted a series of three studies. In their first study, they examined the teacher evaluation systems of 4 Pacific Northwest school districts. Their goal was to gain detailed knowledge of the nature of the teacher evaluation environment and the management of teacher improvement issues within that environment.
The districts were recruited on a volunteer basis. Two districts were located in Washington and two were located in Oregon. One of the Washington districts was suburban and served 13,220 students in 2 high schools, 4 middle schools and 13 elementary schools. The other Washington district was located about 30 miles outside a metropolitan area and served 7100 students in 2 high schools, 4 middle schools and 5 elementary schools. Both Oregon districts were suburban. One district served 6,000 students in 1 high school, 2 middle schools and 9 elementary schools. The other Oregon district served 11,500 students in 3 high schools, 4 junior high schools and 18 elementary schools.

Stiggins and Duke (1988) interviewed district administrators and principals. Each principal was asked to identify 4 teachers to be either interviewed or surveyed. Interviewees were asked to describe teacher evaluation practices from their perspective and if and how they used the results of evaluation to plan for teacher development. Over half the teachers interviewed reported that they wanted more opportunity for collegial observation and self-evaluation through goal setting and videotaping. Teachers tended to feel that for evaluation to be effective it must be a schoolwide priority, occur with regular frequency and
be implemented by evaluators trained to provide relevant, specific and complete feedback. Administrators identified lack of time and lack of teacher trust in the evaluation system as key problems.

For their second study Stiggins and Duke (1988) sent an inquiry to 5000 teachers in the Pacific Northwest asking any teacher who had experienced important and demonstrable professional growth as a result of the teacher evaluation process to contact them. Thirty-three teachers volunteered for the study. The goal of the study was to profile the individual evaluations of these teachers in terms of the attributes that appeared to make them effective.

In addition to the 33 teachers involved in the study, these researchers identified two principals who had established outstanding records of conducting high-quality, growth-producing teacher evaluations. This second study resulted in the identification of 5 factors that contribute to the quality and impact of a particular teacher evaluation experience: the teacher, the evaluator, the procedures, the feedback and the context. Six teacher attributes emerged as important: instructional competence, personal expectations, openness to suggestion, orientation to change, subject knowledge and previous experience with useful evaluations. Six evaluator attributes were
identified: credibility as a source of information for teachers, careful interpersonal communications, persuasiveness, patience, trust, a track record for being helpful and modeling. Two attributes of the procedure were identified as helpful: performance criteria that are appropriate for the individual context and the capabilities of the teacher and data collection procedures that foster an ongoing sequence of regular visits and discussions.

Two attributes of the feedback were identified: the source of the feedback is credible and the feedback describes specific aspects of teaching along with ideas and suggestions for improvement that make sense within the context. Three aspects of the evaluation context were identified: a positive history of labor relations, sufficient time spent on evaluation and availability of resources for growth (released time for visiting other classrooms, attending workshops, staff development activities).

The first study, which focused on district evaluation systems, revealed some of the barriers that prevent teacher evaluation from promoting professional growth. The second study provided evidence that those barriers can be removed. However, the key attributes of evaluation were identified by a content analysis of the interviews of 33 volunteers.
The generalizability of the above mentioned attributes to other teachers and their experiences remained in question. Therefore, the goal of Stiggins and Duke's (1988) third study was to validate the list of key attributes of teacher evaluation by determining if these attributes were related to perceived growth outcomes of evaluation. Stiggins and Duke developed a questionnaire that asked teachers to describe their most recent evaluation experience, rate its overall quality and describe its impact on them. Responses to the questionnaire allowed the researchers to explore the relationships among components of the evaluation process and any relationship between key attributes and various perceived outcomes of the evaluation process.

In this third study the questionnaire asked 470 teachers from the Pacific Northwest to describe their most recent evaluation in terms of 5 categories which included a total of 55 items. In addition to responding to the 55 items, teachers were also asked to rate the outcome of their latest evaluation taking into account the entire process. They were to rank the experience from 0 to 9. A rating of 0 would indicate poor quality. A rating of 9 would indicate high quality. Next they were asked to rate the evaluation experience from 0 to 9 in terms of the
impact the experience had on their teaching attitudes, practices about teaching and/or their understanding of the teaching profession (Stiggins & Duke, 1988).

The analysis of the responses revealed that 44 of the original items combined to create an internally consistent picture of teacher evaluation practices that provided a fairly accurate prediction of the overall quality and impact of those practices (Stiggins & Duke, 1988). This questionnaire has been revised to become the Teacher Evaluation Profile (TEP) which will be used in this study.

The TEP has been found to be an instrument with high validity. Its intent was to identify some of the key attributes of a teacher evaluation environment in a school district. Its validity was established during its development by conducting a content analysis of growth producing teacher evaluation environments and designing the TEP to include key dimensions of those environments. Subsequent collection and analysis of questionnaire responses verified the predictive validity of those dimensions. Regression analyses using the TEP items to predict overall quality and impact consistently produced multiple correlations in excess of .80 (Northwest Regional Educational Laboratory, 1989).
The TEP has been found to be an instrument of high reliability. The internal consistency reliability of the instrument as a whole is .93 (Northwest Regional Educational Laboratory, 1989).

The TEP was originally administered to 470 teachers in 5 different pilot districts. The results were analyzed to determine if the instrument could detect differences in the profiles of those districts. Results of a multivariate analysis of variance of the 5 subscale scores across the 5 districts revealed a sufficiently sensitive instrument (Northwest Regional Educational Laboratory, 1989).

The TEP asks teachers to reflect on their most recent evaluation experience. As they think about this experience, teachers are asked to rank the overall quality of the experience and the impact of the experience in terms of leading to profound changes in their teaching practices, attitudes about teaching and/or understanding of the teaching profession on a scale of 0 to 9. A rating of 0 is a poor rating. A rating of 9 is a high rating.

Teachers are then asked to rate their most recent evaluation experience on a scale of 1 to 5 in terms of 44 key attributes related to the teacher, the evaluator, the evaluation procedures, the evaluation feedback and the
context of the evaluation. The TEP takes approximately 15 minutes to complete.

This researcher contacted the Northwest Regional Educational Laboratory in November 1997. The researcher inquired about the use of the TEP for this study. The researcher was given verbal permission to use the TEP. Written permission was received on March 11, 1998.

Data Collection

This researcher received written permission to conduct this study in the selected district from the district superintendent in January 1998. On May 13, 1998 this researcher made a presentation to the entire district administrative team to explain the purpose of the study and to outline the procedures for data collection and for the maintenance of confidentiality. Principals were assured that individual building results would not be reported.

On May 18, 1998 potential subjects in the study received a letter from the researcher which described the study and invited the subjects to participate. Enclosed with this letter of introduction and invitation was an Informed Consent Form. This letter of introduction and the Informed Consent Form were distributed to the subjects' mailboxes in 7 buildings by this researcher. Subjects were asked to return by May 26, 1998 the signed Informed Consent
Forms in self-addressed stamped envelopes provided by the researcher. The researcher checked off the returned Informed Consent Forms on a district staff list.

On June 1, 1998, the researcher distributed the questionnaire materials to the mailboxes of all subjects in each of the 7 buildings who had consented to participate in this study. Each subject received a cover letter, a copy of the Teacher Evaluation Profile, the accompanying NCS forms to record their responses and a self-addressed stamped envelope to be used to return the completed NCS form to the researcher. Since this study examined the responses of three groups of teachers: (a) non-tenured teachers who were being evaluated during the 1997-98 school year with a pilot evaluation instrument, (b) volunteer tenured teachers being evaluated during the 1997-98 school year with the pilot instrument, and (c) tenured teachers being evaluated during the 1997-98 school year using the traditional district instrument) subjects were asked to indicate in the Identification Box on the NCS form whether they belonged to group number 1, 2 or 3. The subjects were asked to complete the questionnaire and to return the NCS forms to the researcher by June 15, 1998.

The researcher sent a reminder letter to all subjects who agreed to participate in the study on June 8, 1998.
The researcher accepted any completed surveys until August 31, 1998. Forty-two non-tenured teachers, 57 volunteer tenured teachers and 128 tenured teachers returned completed NCS forms.

**Data Analysis**

The data collected from the TEP questionnaire was summarized and analyzed in three forms for each of the three groups: a frequency summary, a profile of means and standard deviations and a correlational analysis. The correlational analysis examined: (a) The relationship between ratings of the teachers in each of the three groups for each of the items on the survey and their ratings of the overall quality and overall impact of their most recent evaluation experience and (b) The relationship between the ratings of the teachers in each of the three groups in each of the subscale areas (attributes of the teacher, attributes of the evaluator, attributes of the evaluation procedures, attributes of the feedback on teaching performance and attributes of the evaluation context) on the survey and their ratings of the overall quality and impact of their most recent evaluation experiences.

A one-way analysis of variance was used to determine any statistically significant differences among the three groups as related to their mean scores of overall quality,
overall impact and their mean scores on each of the 5 subscales.
CHAPTER IV
Presentation Of The Findings

Introduction

The purpose of this study was to increase the knowledge base regarding the key attributes of teacher evaluation systems that foster professional growth in terms of effective instructional practices. This study examined the potential of a newly designed teacher evaluation process to improve instruction. A k-8 suburban district in central New Jersey implemented a new teacher evaluation form during the 1997-98 school year. The district enrolls approximately 3500 students in 7 buildings (five k-5 elementary schools and two 6-8 middle schools). All non-tenured teachers (n= 43) and a volunteer group of tenured teachers (n= 62) were evaluated using this new form during the 1997-98 school year. The remaining tenured teachers (n= 153) were evaluated with the old evaluation form for the 1997-98 school year. The new teacher evaluation form and the process for its implementation had been designed for the purpose of promoting professional growth.

This study examined differences among the three groups of teachers: the non-tenured teachers evaluated using the
new form, the tenured volunteers evaluated using the new form and the tenured teachers evaluated using the old form in terms of their perceptions of, if, and how the evaluation process fostered their professional growth in terms of effective instructional practices.

This study addressed the following research questions:

1. What perceptions did the teachers in each of the three groups have regarding the quality of their most recent teacher evaluation experience (pre-observation, observation, post-observation, end-of-year evaluation)?

2. Following their most recent teacher evaluation experience, what perceptions did the teachers in each of the three groups have regarding the impact of the evaluation process on the following: changes in teaching practices, attitudes about teaching, and/or understanding of the teaching process?

3. What relationships did teachers in each of the three groups perceive between specific attributes of teacher evaluation (attributes of the teacher, attributes of the evaluator, attributes of the evaluation procedures, attributes of the feedback on teaching performance and attributes of the evaluation context) and the overall quality of the teacher evaluation experience?
4. What relationships did teachers in each of the three groups perceive between specific attributes of teacher evaluation (see #3 above) and the promotion of teacher growth (overall impact in terms of changes in teaching practices, attitudes about teaching, and/or understanding of the teaching profession)?

5. Were there differences among the three groups of teachers in the study regarding research questions 1-4?

This chapter reports the results of the three groups of teachers' responses to the Teacher Evaluation Profile (TEP). The data presented are reported in order to answer the five research questions investigated in this dissertation. The data is presented as follows: (a) Each research question is stated, (b) a table/s of the data for each question is presented, (c) a discussion of the findings in the table/s follows, and a summary of the findings for each question is given.

**Description of Respondents**

The researcher surveyed 43 non-tenured teachers, 62 volunteer tenured teachers and 153 tenured teachers. Completed surveys were returned from 42 (98%) non-tenured teachers, 57 (92%) volunteer tenured teachers and 128 (84%) tenured teachers.
Presentation of Data Related to Research Question #1

Research Question #1.

What perceptions did the teachers in each of the three groups have regarding the quality of their most recent teacher evaluation experience (pre-observation, observation, post observation, end-of-the-year evaluation)?

Teachers rated the overall quality of their most recent evaluation experience on a scale of 0 to 9 with 0 representing very poor quality and 9 representing very high quality. Mean scores greater than the midpoint of the scale (5.00) were assumed to be favorable. Mean scores and standard deviations were calculated for each of the 3 groups (non-tenured, volunteer tenured and regular tenured).
Table 1

Summary of Means and Standard Deviations Related to the Overall Quality of the Evaluation Experience

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Valid</th>
<th>Missing</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-tenured</td>
<td></td>
<td></td>
<td></td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>7.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>1.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vol-tenured</td>
<td></td>
<td></td>
<td></td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>6.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>2.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reg-tenured</td>
<td></td>
<td></td>
<td></td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>6.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>2.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. In all tables the following abbreviations will be used: vol-tenured for the volunteer tenured group and reg-tenured for the regular tenured group.

All three groups had a favorable rating of the overall quality of their most recent evaluation experience with mean scores ranging from 6.43 to 7.00. Non-tenured teachers who were evaluated using the new evaluation process reported the highest mean score relative to the overall quality of the evaluation experience with a mean score of 7.00 and a standard deviation of 1.53. The volunteer tenured teacher group who was evaluated using the
new evaluation process reported the second highest mean score related to the overall quality of their most recent evaluation experience with a mean score of 6.47 and a standard deviation of 2.09. The regular tenured teacher group who was evaluated using the traditional evaluation process reported a mean score of 6.43 and a standard deviation of 2.02.

A one-way ANOVA was used to investigate whether or not there were significant differences among the three groups' mean rating scores related to the overall quality of their evaluation experience.

Table 2

One-Way Analysis of Variance Related to the Overall Quality of the Evaluation Experience

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUALITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between groups</td>
<td>10.687</td>
<td>2</td>
<td>5.343</td>
<td>1.369</td>
<td>.250</td>
</tr>
<tr>
<td>Within groups</td>
<td>857.578</td>
<td>224</td>
<td>3.828</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>868.264</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The above table shows that there were no significant differences (p<.05) among the three groups in terms of how
they rated the overall quality of their most recent evaluation experience.

Summary of the findings related to research question #1.

Overall, there are no statistically significant (p<.05) differences among the mean scores of the three groups in their rating of the overall quality of their most recent evaluation experience. All three groups had mean scores, which were above the midpoint level (5.00) of the 0 to 9-point scale, and therefore assumed to be favorable.

Presentation of the Data Related to Research Question #2

Research question #2.

Following their most recent teacher evaluation experience, what perceptions did the teachers in each of the three groups have regarding the impact of the evaluation process on the following: changes in teaching practices, attitudes about teaching, and/or understanding of the teaching process?

Teachers rated the overall impact of their most recent evaluation experience on a scale of 0 to 9 with 0 reflecting no impact at all (no changes in practices, attitudes and/or understanding of the teaching profession) and 9 representing a strong impact (leading to profound changes in teacher practices, attitudes about teaching, and/or understanding of the teaching profession). Mean
scores greater than the midpoint on the scale (5.00) were assumed to be favorable. Mean scores and standard deviations were calculated for each of the three groups.

Table 3

Summary of Means and Standard Deviations Related to the Overall Impact of the Evaluation Experience

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Valid</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-tenured</td>
<td>42</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>7.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Vol-tenured</td>
<td>57</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>5.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>2.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Reg-tenured</td>
<td>128</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>5.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>2.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

All three groups had a favorable rating of the overall impact of their most recent evaluation experience with mean scores ranging from 5.12 to 7.40.

Non-tenured teachers who were evaluated using the new evaluation process had the highest mean score related to the overall impact of their most recent evaluation experience of 7.40 with a standard deviation of 1.50. The
volunteer tenured teachers who were evaluated using the new evaluation process had a mean score related to the overall impact of their most recent evaluation experience of 5.12 with a standard deviation of 2.16. Regular tenured teachers who were evaluated using the traditional evaluation process had a mean score related to the overall impact of their most recent evaluation experience of 5.46 with a standard deviation of 2.49.

A one-way ANOVA was used to investigate whether or not there were differences among the means of the three groups' scores for impact.

Table 4

One-Way Analysis of Variance Related to the Overall Impact of the Evaluation Experience

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPACT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between groups</td>
<td>148.077</td>
<td>2</td>
<td>74.038</td>
<td>14.522</td>
<td>.000</td>
</tr>
<tr>
<td>Within groups</td>
<td>1142.064</td>
<td>224</td>
<td>5.099</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1290.141</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The one-way analysis of variance shows that there was a statistically significant difference at the p<.05 level.
among the three groups related to how each group rated the impact of the evaluation experience.

The Scheffe Multiple Comparisons Test was used to test each of all possible comparisons.

Table 5

Scheffe Test Related to the Overall Impact of the Evaluation Experience

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(i) Group</th>
<th>(j) Group</th>
<th>Mean Diff. (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact</td>
<td>non-tenured</td>
<td>vol-tenured</td>
<td>2.28</td>
<td>.459</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reg-tenured</td>
<td>1.94</td>
<td>.402</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>vol-tenured</td>
<td>non-tenured</td>
<td>-2.28</td>
<td>.459</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reg-tenured</td>
<td>-.34</td>
<td>.360</td>
<td>.643</td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>non-tenured</td>
<td>-1.94</td>
<td>.402</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>vol-tenured</td>
<td>.34</td>
<td>.360</td>
<td>.643</td>
</tr>
</tbody>
</table>

The Scheffe Multiple Comparison Test shows statistically significant differences between non-tenured and volunteer tenured teachers (p < .05) and between non-tenured and regular tenured teachers (p < .05). The data show that the mean impact score for the non-tenured teachers (7.40) was significantly higher than the scores of both the
volunteer tenured group (5.12) and the regular tenured group (5.46).

**Summary of findings related to research question #2.**

Overall, the three groups in this study had a favorable rating of the overall impact of the evaluation experience with mean scores varying from 5.12 to 7.40 on a 0 to 9-point scale. There was a statistically significant difference (p<.05) between the mean scores of the non-tenured teachers and both the volunteer tenured and regular tenured teachers. The data show that the mean impact score for the non-tenured (7.40) teachers was significantly higher than the scores of both the volunteer tenured group (5.12) and the regular tenured group (5.46). There was no significant difference between the two tenured groups.

**Presentation of the Data Related to Research Question #3**

**Research question #3.**

What relationships did teachers in each of the three groups perceive between specific attributes of teacher evaluation (attributes of the teacher, attributes of the evaluator, attributes of the evaluation procedures, attributes of the feedback on teaching performance and attributes of the evaluation context) and the overall quality of the teacher evaluation experience?
A correlational analysis was used to examine the relationships between the teachers' ratings of overall quality and the subscales related to specific attributes of the teacher evaluation process.

Table 6

**Correlations Between the Overall Quality of the Evaluation Experience and the Specific Attributes of Teacher Evaluation**

<table>
<thead>
<tr>
<th>Group</th>
<th>TEACHERS</th>
<th>EVALS</th>
<th>PROC</th>
<th>FEEDBACK</th>
<th>CONTEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-tenured Quality Pearson Corr</td>
<td>.404</td>
<td>.565</td>
<td>.233</td>
<td>.999</td>
<td>.311</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tail)</td>
<td>.008</td>
<td>.000</td>
<td>.031</td>
<td>.000</td>
</tr>
<tr>
<td>N=42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2 tail)</td>
<td>.328</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>N=57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2 tail)</td>
<td>.058</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>N=128</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Correlations at the p<.05 level are in bold type.

Table 6 shows a positive, statistically significant (p<.05) correlation between the non-tenured teachers' rating of the overall quality of their most recent evaluation experience and the following subscale areas:

1. **TEACHERS** (attributes of the teacher) \( r = .40 \) \( p < .01 \)
2. **EVALS** (attributes of the evaluator) \( r = .56 \) \( p < .01 \)
3. **PROCS** (attributes of the evaluation procedures) \( r = .33 \) \( p < .05 \)
4. FDBACK (attributes of the feedback on teaching performance) $r = .59 \ p < .01$

5. CONTEXT (attributes of the evaluation context) $r = .31 \ p < .05$

The data show a positive, statistically significant ($p < .05$) correlation between volunteer tenured teachers' ratings of the overall quality of their most recent evaluation experience and EVALS ($r = .47 \ p < .01$), PROCS ($r = .51 \ p < .01$), FDBACK ($r = .52 \ p < .01$), and CONTEXT ($r = .45 \ p < .01$).

The data show a positive, statistically significant ($p < .05$) correlation between regular tenured teachers' ratings of the overall quality of their most recent evaluation experience and EVALS ($r = .50 \ p < .01$), PROCS ($r = .54 \ p < .01$), FDBACK ($r = .49 \ p < .01$) and CONTEXT ($r = .31 \ p < .01$). The only difference between non-tenured and tenured teachers occurred in the TEACHERS subscale.

For the volunteer tenured teachers and the regular tenured teacher groups there was no statistically significant correlation between their overall rating of the quality of their most recent evaluation experience and TEACHERS.

In addition to the correlational analysis of the relationship between overall quality and each of the 5 subscales as a whole, this researcher examined the correlation between overall quality and each of the
individual attributes in each subscale. The data in Tables 7 through 11 depict a breakdown of those attributes in each subscale that had a positive, statistically significant (p<.01) correlation to the teachers’ rating of the overall quality of their most recent evaluation experience.

Table 7

Correlations Between TEACHERS and Overall Quality

<table>
<thead>
<tr>
<th>TEACHERS Attributes</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-tenured</td>
</tr>
<tr>
<td>1. Rate the strength of your professional expectations of yourself</td>
<td>-</td>
</tr>
<tr>
<td>2. Orientation to risk taking</td>
<td>-</td>
</tr>
<tr>
<td>3. Orientation to change</td>
<td>-</td>
</tr>
<tr>
<td>4. Orientation to experimentation in your classroom</td>
<td>.39</td>
</tr>
<tr>
<td>5. Openness to criticism</td>
<td>-</td>
</tr>
<tr>
<td>6. Knowledge of technical aspects of teaching</td>
<td>-</td>
</tr>
<tr>
<td>7. Knowledge of subject matter</td>
<td>-</td>
</tr>
<tr>
<td>8. Years of experience in position</td>
<td>-</td>
</tr>
<tr>
<td>9. Experience with teacher evaluation prior to most recent experience</td>
<td>.43</td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the p<.01 level have been included in this table.

Of the 9 attributes in the TEACHERS subscale, 2 of the 9 (22%) correlated at the p<.01 level of significance to the rating of the overall quality of the most recent evaluation experience in at least one of the groups. Those attributes
were orientation to experimentation in your class (r=.39 for non-tenured), and experience with teacher evaluation prior to most recent experience (r= .43 for non-tenured and r= .46 for regular tenured).

Table 8

Correlations Between EVALS Subscale Attributes and the Rating of Overall Quality

<table>
<thead>
<tr>
<th>EVALS Attributes</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-tenured</td>
</tr>
<tr>
<td>10. Credibility as a source of feedback</td>
<td>.41</td>
</tr>
<tr>
<td>11. Working relationship with you</td>
<td>.44</td>
</tr>
<tr>
<td>12. Level of trust</td>
<td>.41</td>
</tr>
<tr>
<td>13. Interpersonal manner</td>
<td>-</td>
</tr>
<tr>
<td>14. Temperament</td>
<td>-</td>
</tr>
<tr>
<td>15. Flexibility</td>
<td>-</td>
</tr>
<tr>
<td>16. Knowledge of technical aspects of teaching</td>
<td>-</td>
</tr>
<tr>
<td>17. Capacity to demonstrate or model needed improvements</td>
<td>.44</td>
</tr>
<tr>
<td>18. Familiarity with your classroom</td>
<td>.44</td>
</tr>
<tr>
<td>19. Familiarity with classrooms in general</td>
<td>.58</td>
</tr>
<tr>
<td>20. Usefulness of suggestions for improvements</td>
<td>.59</td>
</tr>
<tr>
<td>21. Persuasiveness of rational for suggestions</td>
<td>.47</td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the p<.01 level have been included in this table.
Of the 12 items in the EVALS subscale the same 6 items (50%) were positively correlated at the p<.01 level of significance to the rating of the overall quality of their most recent evaluation experience for all 3 groups. Those attributes were: (a) credibility as a source of feedback (r = .41 for non-tenured, r = .43 for volunteer tenured and r = .51 for regular tenured), (b) working relationship with you (r = .44 for non-tenured, r = .36 for volunteer tenured, r = .46 for regular tenured), (c) level of trust (r = .41 for non-tenured, r = .54 for volunteer tenured, r = .45 for regular tenured), (d) familiarity with classrooms in general (r = .58 for non-tenured, r = .34 for volunteer tenured, r = .35 for regular tenured), (e) usefulness of suggestions for improvements (r = .59 for non-tenured, r = .37 for volunteer tenured, r = .42 for regular tenured), and (f) persuasiveness of rational for suggestions (r = .47 for non-tenured, r = .49 for volunteer tenured, r = .42 for regular tenured).

The regular tenured teachers group had 11 of the 12 items (92%) on the EVALS subscale correlated at the p<.01 level to the rating of the overall quality of their most recent evaluation experience. Volunteer tenured teachers had 9 of the 12 items (75%) correlated at the p<.01 level,
while the non-tenured teachers had 8 of the 12 items (66%) of the items correlated at the p<.01 level of significance.

Table 9

Correlations between the PROCS Subscale Attributes and the Rating of Overall Quality

<table>
<thead>
<tr>
<th>PROCS Attributes</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-tenured</td>
</tr>
<tr>
<td>22. Were standards communicated to you?</td>
<td>-</td>
</tr>
<tr>
<td>23. Were the standards clear to you?</td>
<td>-</td>
</tr>
<tr>
<td>24. Were standards endorsed by you as appropriate for your classroom?</td>
<td>-</td>
</tr>
<tr>
<td>25. Were the standards...All the same for all teachers...Tailored somewhat for your unique needs?</td>
<td>-</td>
</tr>
<tr>
<td>26. Observation of your classroom performance</td>
<td>-</td>
</tr>
<tr>
<td>27. Examination of classrooms or school records (lesson plans, etc.)</td>
<td>-</td>
</tr>
<tr>
<td>28. Examination of student achievement</td>
<td>-</td>
</tr>
<tr>
<td>29. Number of formal (prescheduled) observations per year (0 to 4 or more)</td>
<td>-</td>
</tr>
<tr>
<td>30. Approximate frequency of informal (unannounced drop-in) observations (0 to daily)</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the p<.01 level have been included in this table.

Of the 9 attributes in the PROCS subscale, the same 4 items (44%) were positively correlated at the p<.01 level of significance to the volunteer tenured teachers' and the
regular tenured teachers' rating of the overall quality of their most recent evaluation experience. Those attributes were: (a) Were standards communicated to you? \( r = 0.49 \) for the volunteer tenured, \( r = 0.37 \) for the regular tenured), (b) were the standards clear to you? \( r = 0.50 \) for the volunteer tenured, \( r = 0.32 \) for the regular tenured), (c) were standards endorsed by you as appropriate for your classroom? \( r = 0.47 \) for the volunteer tenured, \( r = 0.40 \) for the regular tenured), and (d) observation of your classroom performance \( r = 0.57 \) for the volunteer tenured, \( r = 0.42 \) for the regular tenured.

The regular tenured teachers had 7 of the 9 items (77%) on the PROCS subscale correlated at the \( p < 0.01 \) level of significance to the rating of overall quality of their most recent evaluation experience. The volunteer tenured teachers had 4 of the 9 items (44%) correlated at the \( p < 0.01 \) level of significance, while the non-tenured teachers had no items correlated at the \( p < 0.01 \) level of significance.
Table 10

Correlations between the Feedback Subscales and the Rating of Overall Quality

<table>
<thead>
<tr>
<th>Feedback Attributes</th>
<th>Non-tenured</th>
<th>Vol-tenured</th>
<th>Reg-tenured</th>
</tr>
</thead>
<tbody>
<tr>
<td>31. Amount of information received</td>
<td>.49</td>
<td>.37</td>
<td>.28</td>
</tr>
<tr>
<td>32. Frequency of formal feedback</td>
<td>-</td>
<td>-</td>
<td>.36</td>
</tr>
<tr>
<td>33. Frequency of informal feedback</td>
<td>.39</td>
<td>.40</td>
<td>.34</td>
</tr>
<tr>
<td>34. Depth of information provided</td>
<td>.59</td>
<td>.40</td>
<td>.34</td>
</tr>
<tr>
<td>35. Quality of the ideas and suggestions contained in the feedback</td>
<td>.63</td>
<td>.55</td>
<td>.51</td>
</tr>
<tr>
<td>36. Specificity of information provided</td>
<td>.54</td>
<td>.45</td>
<td>.36</td>
</tr>
<tr>
<td>37. Nature of information provided</td>
<td>.47</td>
<td>.50</td>
<td>.48</td>
</tr>
<tr>
<td>38. Timing of feedback</td>
<td>-</td>
<td>-</td>
<td>.34</td>
</tr>
<tr>
<td>39. Feedback focused on district teaching standards</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the p<.01 level have been included in this table.

Of the 9 attributes in the Feedback subscale, all 3 groups had the same 6 items (66%) correlated at the p<.01 level of significance to their rating of the overall quality of their most recent evaluation experience. Those attributes were: (a) amount of information received \( (r = .49 \) for non-tenured, \( r = .37 \) for volunteer tenured, \( r = .28 \) for regular tenured), (b) frequency of informal feedback \( (r = \)
.39 for non-tenured, r = .40 for volunteer tenured, r = .34 for regular tenured), (c) depth of information provided (r = .59 for non-tenured, r = .40 for volunteer tenured, r = .34 for regular tenured), (d) quality of the ideas and suggestions contained in the feedback (r = .63 for non-tenured, r = .55 for volunteer tenured, r = .51 for regular tenured), (e) specificity of information provided (r = .54 for non-tenured, r = .45 for volunteer tenured, r = .36 for regular tenured), and (f) nature of information provided (r = .47 for non-tenured, r = .50 for volunteer tenured, r = .48 for regular tenured).

The regular tenured teachers had 8 of the 9 items (88%) on the FDBACK subscale correlated at the p < .01 level to the rating of the overall quality of their most recent evaluation experience. The volunteer tenured teachers and the non-tenured teachers both had the same 6 of the 9 items (66%) correlated at the p < .01 level.
Table 11

Correlations between the CONTEXT Subscale and the Rating of Overall Quality

<table>
<thead>
<tr>
<th>CONTEXT Attributes</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-tenured</td>
</tr>
<tr>
<td>40. Amount of time spent on the evaluation process, including your time and that of all participants</td>
<td>-</td>
</tr>
<tr>
<td>41. Time allotted during the teaching day for professional development</td>
<td>-</td>
</tr>
<tr>
<td>42. Availability of training programs and models of good practice</td>
<td>-</td>
</tr>
<tr>
<td>43. Clarity of policy statements</td>
<td>-</td>
</tr>
<tr>
<td>44. Intended role of evaluation</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note.* Only those correlations significant at the $p<.01$ level have been included in this table.

Of the 5 items in the CONTEXT subscale, the volunteer tenured teacher and the tenured teacher groups had the same 2 items (40%) correlated at the $p<.01$ level of significance to their rating of the overall quality of their most recent evaluation experience. Those attributes were: (a) amount of time spent on the evaluation process, including your time and that of all other participants ($r = .49$ for volunteer tenured, $r = .24$ for regular tenured), and (b) clarity of policy statements regarding purpose for
evaluation (r = .38 for volunteer tenured, r = .30 for regular tenured)

The regular tenured teachers had 3 of the 5 CONTEXT subscale items (60%) correlated at the p<.01 level to the rating of the overall quality of their most recent evaluation experience. The volunteer tenured teachers had 2 of the 5 items (40%), while the non-tenured teachers had no items correlated.

Summary of findings for research question #3.

There were positive, statistically significant (p<.05) correlations for all three groups between their rating of the overall quality of their most recent evaluation experience and 4 (EVALS, PROCS, FDBACK and CONTEXT) out of the 5 subscales. The non-tenured teacher group was the only group that had a positive, statistically significant (p<.05) correlation between their rating of the overall quality of their most recent evaluation experience and the TEACHERS subscale.

In addition, the subscales of EVALS, FDBACK had the highest percentages (from 50% to 92%) of items that were correlated at a statistically significant level (p<.01) to teachers' rating of the overall quality of their most recent evaluation experience for all 3 groups. The regular tenured group had the highest number of items correlated to
the rating of overall quality in all subscales except TEACHERS. On the TEACHERS subscale, the non-tenured group had the highest percentage of items (22%) correlated at the p<.01 level to overall quality.

Presentation of Data Related to Research Question #4

Research question #4.

What relationships did teachers in each of the three groups perceive between specific attributes of teacher evaluation (attributes of the teacher, attributes of the evaluator, attributes of the evaluation procedures, attributes of the feedback on teaching performance and attributes of the evaluation context) and the promotion of teacher growth (overall impact in terms of changes in teaching practices, attitudes about teaching, and/or understanding of the teaching profession)?
Table 12

Correlations Between the Overall Impact of the Evaluation Experience and the Specific Attributes of Teacher Evaluation

<table>
<thead>
<tr>
<th>Group</th>
<th>TEACHERS</th>
<th>EVALS</th>
<th>PROCs</th>
<th>FDBACK</th>
<th>CONTEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-tenured</td>
<td>.162</td>
<td>.185</td>
<td>.326</td>
<td>.338</td>
<td>.347</td>
</tr>
<tr>
<td>Sig. (2-tail)</td>
<td>.305</td>
<td>.241</td>
<td>.035</td>
<td>.033</td>
<td>.024</td>
</tr>
<tr>
<td>N=42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vol-tenured</td>
<td>-.223</td>
<td>.097</td>
<td>.097</td>
<td>.163</td>
<td>.051</td>
</tr>
<tr>
<td>Sig. (2 tail)</td>
<td>.096</td>
<td>.427</td>
<td>.474</td>
<td>.226</td>
<td>.708</td>
</tr>
<tr>
<td>N=57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reg-tenured</td>
<td>.130</td>
<td>.264</td>
<td>.311</td>
<td>.436</td>
<td>.185</td>
</tr>
<tr>
<td>Sig. (2 tail)</td>
<td>.144</td>
<td>.003</td>
<td>.000</td>
<td>.000</td>
<td>.036</td>
</tr>
<tr>
<td>N=128</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Correlations at the p<.05 level are in bold type.

The data show that none of the three groups found a statistically significant correlation between the overall impact of their most recent evaluation experience and the TEACHERS subscale.

The data show a positive, statistically significant correlation between the non-tenured teachers’ ratings of the overall impact of their most recent evaluation experience and the following: PROCs (r=.32 p<.05), FDBACK (r=.33 p<.05) and CONTEXT (r=.34 p<.05).
Table 12 shows that there are no statistically significant correlations between the volunteer tenured teacher group and the 5 subscale areas.

The data show a positive, statistically significant correlation between the regular tenured teachers' ratings of the overall impact of their most recent evaluation experience and EVALS ($r=.26 \ p<.01$), PROCS ($r=.31 \ p<.01$), FDBACK ($r=.43 \ p<.01$) and CONTEXT ($r=.18 \ p<.05$).

In addition to the correlational analysis of the relationship between overall impact and each of the 5 subscales as a whole, this researcher examined the correlation between overall impact and each of the individual attributes in each subscale. The data in Tables 13 through 17 depict a breakdown of those attributes in each subscale that had a positive, statistically significant ($p<.01$) correlation to the teachers' rating of the overall impact of their most recent evaluation experience.
Table 13

Correlations between the TEACHERS Subscale Attributes and the Rating of Overall Impact

<table>
<thead>
<tr>
<th>TEACHERS Attributes</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-tenured</td>
</tr>
<tr>
<td>1. Rate the strength of your professional expectations of yourself</td>
<td>-</td>
</tr>
<tr>
<td>2. Orientation to risk taking</td>
<td>-</td>
</tr>
<tr>
<td>3. Orientation to change</td>
<td>-</td>
</tr>
<tr>
<td>4. Orientation to experimentation in your classroom</td>
<td>-</td>
</tr>
<tr>
<td>5. Openness to criticism</td>
<td>-</td>
</tr>
<tr>
<td>6. Knowledge of technical aspects of teaching</td>
<td>-</td>
</tr>
<tr>
<td>7. Knowledge of subject matter</td>
<td>-</td>
</tr>
<tr>
<td>8. Years of experience in position</td>
<td>-.39</td>
</tr>
<tr>
<td>9. Experience with teacher evaluation prior to most recent experience</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the p<.01 level have been included in this table.

Of the 9 attributes in the TEACHERS subscale, no items were correlated at the p<.01 level to the rating of overall impact of their most recent evaluation experience for the no-tenured teacher group. Attribute #8 (years of experience in position) was negatively correlated (r= -.39) at the p<.01 level to overall impact for the volunteer tenured teacher group and attribute #9 (experience with teacher evaluation prior to most recent experience) was positively
correlated \((r = .42)\) at the \(p < .01\) level for the regular tenured group.

Table 14

Correlations between the EVALS Subscale Attributes and the Rating of Overall Impact

<table>
<thead>
<tr>
<th>EVALS Attributes</th>
<th>Groups</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-tenured</td>
<td>Vol-tenured</td>
<td>Reg-tenured</td>
<td></td>
</tr>
<tr>
<td>10. Credibility as a source of feedback</td>
<td>-</td>
<td>-</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>11. Working relationship with you</td>
<td>-</td>
<td>-</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>12. Level of trust</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>13. Interpersonal manner</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>14. Temperament</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>15. Flexibility</td>
<td>-</td>
<td>-</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>16. Knowledge of technical aspects of teaching</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>17. Capacity to demonstrate or model needed improvements</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>18. Familiarity with your classroom</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>19. Familiarity with classrooms in general</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>20. Usefulness of suggestions for improvements</td>
<td>-</td>
<td>-</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>21. Persuasiveness of rational for suggestions</td>
<td>-</td>
<td>-</td>
<td>.23</td>
<td></td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the \(p < .01\) level have been included in this table.

Of the 12 items on the EVALS subscale, none of them were correlated to the rating of overall impact of their most recent evaluation experience for the non-tenured and volunteer tenured groups.
Of the 12 items in the EVALS subscale, 5 of the 12 (42%) were positively correlated at the p<.01 level of significance to the regular teachers' ratings of the overall impact of their most recent evaluation experience. Those 5 attributes were the following: (a) credibility as a source of feedback ($r = .33$), (b) working relationship with you ($r = .23$), (c) flexibility ($r = .33$), (d) usefulness of suggestions for improvements ($r = .33$), and (e) persuasiveness of rational for suggestions ($r = .23$).
Table 15

Correlations between the PROCS Subscale Attributes and the Rating of Overall Impact

<table>
<thead>
<tr>
<th>PROCS Attributes</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-tenured</td>
</tr>
<tr>
<td>22. Were standards communicated to you?</td>
<td>-</td>
</tr>
<tr>
<td>23. Were the standards clear to you?</td>
<td>-</td>
</tr>
<tr>
<td>24. Were standards endorsed by you as appropriate for your classroom?</td>
<td>-</td>
</tr>
<tr>
<td>25. Were the standards...All the same for all teachers...Tailored somewhat for your unique needs?</td>
<td>-</td>
</tr>
<tr>
<td>26. Observation of your classroom performance</td>
<td>-</td>
</tr>
<tr>
<td>27. Examination of classrooms or school records (lesson plans, etc.)</td>
<td>-</td>
</tr>
<tr>
<td>28. Examination of student achievement</td>
<td>-</td>
</tr>
<tr>
<td>29. Number of formal (prescheduled) observations per year (0 to 4 or more)</td>
<td>-</td>
</tr>
<tr>
<td>30. Approximate frequency of informal (unannounced drop-in observations 0 to daily)</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the p<.01 level have been included in this table.

Table 15 shows that no attributes in the PROCS subscale were correlated at the p<.01 level to the rating of the overall impact of the evaluation experience for the non-
tenured and volunteer tenured groups. For the regular tenured teachers, 3 of the 9 attributes (33%) were positively correlated at the p<.01 level. Those three attributes were the following: (a) Were the standards communicated to you? (r= .25), (b) Were the standards endorsed by you as appropriate for your classroom? (r= .25), and (c) Examination of classroom or school records (lesson plans, etc.) (r= .28).
Table 16

**Correlations between the FDBACK Subscales and the Rating of Overall Impact**

<table>
<thead>
<tr>
<th>FDBACK Attributes</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-tenured</td>
</tr>
<tr>
<td>31. Amount of information received</td>
<td>-</td>
</tr>
<tr>
<td>32. Frequency of formal feedback</td>
<td>-</td>
</tr>
<tr>
<td>33. Frequency of informal feedback</td>
<td>-</td>
</tr>
<tr>
<td>34. Depth of information provided</td>
<td>-</td>
</tr>
<tr>
<td>35. Quality of the ideas and suggestions contained in the feedback</td>
<td>-</td>
</tr>
<tr>
<td>36. Specificity of information provided</td>
<td>-</td>
</tr>
<tr>
<td>37. Nature of information provided</td>
<td>-</td>
</tr>
<tr>
<td>38. Timing of feedback</td>
<td>-</td>
</tr>
<tr>
<td>39. Feedback focused on district teaching standards</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the p<.01 level have been included in this table.

Of the 9 attributes in the FDBACK subscale, none of the attributes correlate at the p<.01 level of significance to the rating of the overall impact of their most recent evaluation experience for the non-tenured and volunteer tenured teacher groups.

Of the 9 attributes in the FDBACK subscale, 100% of the attributes were positively correlated at the p<.01 level to their rating of the overall impact of their most
recent evaluation experience for the regular tenured group. Those attributes were the following: (a) amount of information received \((r = .22)\), (b) frequency of formal feedback \((r = .44)\), (c) frequency of informal feedback \((r = .31)\), (d) depth of information provided \((r = .25)\), (e) quality of the ideas and suggestions contained in the feedback \((r = .40)\), (f) specificity of information provided \((r = .22)\), (g) nature of information provided \((r = .33)\), (h) timing of feedback \((r = .38)\), and (i) feedback focused on district teaching standards \((r = .25)\).
Table 17

Correlations between the CONTEXT Subscale and the Rating of Overall Impact

<table>
<thead>
<tr>
<th>CONTEXT Attributes</th>
<th>Non-tenured</th>
<th>Vol-tenured</th>
<th>Reg-tenured</th>
</tr>
</thead>
<tbody>
<tr>
<td>40. Amount of time spent on the evaluation process, including your time and that of all participants</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>41. Time allotted during the teaching day for professional development</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>42. Availability of training programs and models of good practice</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>43. Clarity of policy statements</td>
<td>.39</td>
<td>-</td>
<td>.29</td>
</tr>
<tr>
<td>44. Intended role of evaluation</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Only those correlations significant at the p<.01 level have been included in this table.

Of the 5 items in the CONTEXT subscale 1 of the 5 attributes (20%) was correlated at the p<.01 level of significance to both the non-tenured teachers' and the regular tenured teachers' ratings of the overall impact of their most recent evaluation experience. That attribute was: clarity of policy statements regarding purpose for evaluation (r= .39 for non-tenured, r= .29 for regular tenured). No CONTEXT subscale attributes were correlated at the p<.01 level to the volunteer tenured teachers'
overall ratings of the impact of their most recent evaluation experience.

**Summary of findings for question #4.**

The data show that none of the three teacher groups found a statistically significant correlation (p<.05) between the overall impact of their most recent evaluation experience and the TEACHERS subscale. The non-tenured teachers group had statistically significant correlations (p<.05) between the PROCS (r= .32 p<.05), FDBACK (r= .33 p<.05) and CONTEXT (r= .34 p< .05) subscales and their overall rating of the impact of their most recent evaluation experience. The group with the most subscales correlated at the p<.05 level of significance to the rating of the overall impact of their most recent evaluation experience was the regular tenured group with EVALS (r= .26), PROCS (r= .31), FDBACK (r= .43) and CONTEXT (r= .18) showing correlations at p<.05.

In addition, for the regular tenured teacher group the subscales of EVALS, PROCS and FDBACK had the highest percentages of individual attribute items correlated at the p<.01 level of significance to their rating of the overall impact of their most recent evaluation experience when compared to the non-tenured and volunteer tenured groups. The volunteer tenured group had only one attribute (years
of experience in position) negatively correlated \((r = -0.39)\) at the \(p < 0.01\) level to the rating of the overall impact of their most recent evaluation experience. The non-tenured teacher group had only one attribute (clarity of policy statements \(r = 0.39\)) positively correlated at the \(p < 0.01\) level to their rating of the overall impact of their most recent evaluation experience.

**Presentation of the Data for Research Question #5**

**Research question #5.**

Are there differences among the three groups of teachers in the study regarding research questions 1-4?

The differences among the three groups have been discussed in the presentation of the data for research questions 1-4. However, Table 18 presents additional data about the mean scores and standard deviations for the three groups of teachers for the 5 subscale areas. Table 19 presents a one-way ANOVA comparing the differences between the mean scores of the three groups.
Table 18

Summary of Mean Scores and Standard Deviations Related to the 5 Subscale Areas

<table>
<thead>
<tr>
<th>Group</th>
<th>TEACHERS</th>
<th>EVALS</th>
<th>PROC</th>
<th>FEEDBACK</th>
<th>CONTEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-tenured</td>
<td>N Valid</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>36.40</td>
<td>48.57</td>
<td>32.98</td>
<td>35.93</td>
<td>17.67</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>2.72</td>
<td>10.23</td>
<td>4.11</td>
<td>5.88</td>
<td>4.11</td>
</tr>
<tr>
<td>Minimum</td>
<td>29</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>Maximum</td>
<td>60</td>
<td>60</td>
<td>41</td>
<td>45</td>
<td>25</td>
</tr>
<tr>
<td>Vol-tenured</td>
<td>N Valid</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>38.51</td>
<td>49.84</td>
<td>30.58</td>
<td>34.28</td>
<td>16.65</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>3.52</td>
<td>10.32</td>
<td>6.94</td>
<td>7.53</td>
<td>3.83</td>
</tr>
<tr>
<td>Minimum</td>
<td>28</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Maximum</td>
<td>45</td>
<td>60</td>
<td>42</td>
<td>45</td>
<td>25</td>
</tr>
<tr>
<td>Reg-tenured</td>
<td>N Valid</td>
<td>128</td>
<td>128</td>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>38.20</td>
<td>45.72</td>
<td>28.52</td>
<td>31.42</td>
<td>15.23</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>3.54</td>
<td>10.21</td>
<td>6.28</td>
<td>7.13</td>
<td>4.04</td>
</tr>
<tr>
<td>Minimum</td>
<td>29</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>Maximum</td>
<td>45</td>
<td>60</td>
<td>42</td>
<td>45</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 18 shows that on all subscales except TEACHERS the mean scores for the regular tenured teachers evaluated using the traditional process are lower than the scores of the non-tenured and volunteer tenured teachers who were evaluated using the new process.
Table 19

One-Way Analysis of Variance Related to the Mean Scores for the Subscales

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEACHERS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>125.774</td>
<td>2</td>
<td>62.887</td>
<td>5.438</td>
<td>.005</td>
</tr>
<tr>
<td>Within</td>
<td>2590.482</td>
<td>224</td>
<td>11.565</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2716.256</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVALS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>756.216</td>
<td>2</td>
<td>378.108</td>
<td>3.602</td>
<td>.029</td>
</tr>
<tr>
<td>Within</td>
<td>23513.740</td>
<td>224</td>
<td>104.972</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24269.956</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROCS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>668.649</td>
<td>2</td>
<td>334.325</td>
<td>8.919</td>
<td>.000</td>
</tr>
<tr>
<td>Within</td>
<td>8396.849</td>
<td>224</td>
<td>37.486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9065.489</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEEDBACK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>772.328</td>
<td>2</td>
<td>386.164</td>
<td>7.826</td>
<td>.001</td>
</tr>
<tr>
<td>Within</td>
<td>11053.513</td>
<td>224</td>
<td>49.346</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11825.841</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTEXT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>215.359</td>
<td>2</td>
<td>107.679</td>
<td>6.713</td>
<td>.001</td>
</tr>
<tr>
<td>Within</td>
<td>3593.285</td>
<td>224</td>
<td>16.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3808.643</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. F significant at the p < .05 level are in bold type.

The one-way ANOVA shows that there were statistically significant mean differences in the scores of the three groups of teachers on all 5 of the subscales. The Scheffe Multiple Comparisons Test was used to test each of all possible comparisons for each of the subscales.
Table 20

Scheffe Test Related to the Mean Differences Between Groups on the 5 Subscales

<table>
<thead>
<tr>
<th>Dependent Variable (I) Group</th>
<th>(J) Group</th>
<th>Mean Diff. (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-tenured</td>
<td>vol-tenured</td>
<td>-2.10</td>
<td>.692</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>-1.79</td>
<td>.605</td>
<td>.014</td>
</tr>
<tr>
<td>vol-tenured</td>
<td>non-tenured</td>
<td>2.10</td>
<td>.692</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>.31</td>
<td>.542</td>
<td>.846</td>
</tr>
<tr>
<td>reg-tenured</td>
<td>non-tenured</td>
<td>1.79</td>
<td>.605</td>
<td>.014</td>
</tr>
<tr>
<td></td>
<td>vol-tenured</td>
<td>-.31</td>
<td>.542</td>
<td>.846</td>
</tr>
</tbody>
</table>

Note. Mean differences at the p<.05 level are in bold type.
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(I) Group</th>
<th>(J) Group</th>
<th>Mean Diff. (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EVALS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-tenured</td>
<td>vol-tenured</td>
<td>1.27</td>
<td>2.083</td>
<td>.830</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>2.85</td>
<td>1.822</td>
<td>.295</td>
<td></td>
</tr>
<tr>
<td>vol-tenured</td>
<td>non-tenured</td>
<td>4.12</td>
<td>1.631</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>.61</td>
<td>1.631</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td>reg-tenured</td>
<td>non-tenured</td>
<td>-2.85</td>
<td>1.822</td>
<td>.295</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vol-tenured</td>
<td>-4.12</td>
<td>1.631</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td><strong>PROCS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-tenured</td>
<td>vol-tenured</td>
<td>2.40</td>
<td>1.245</td>
<td>.159</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>4.46</td>
<td>1.089</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>vol-tenured</td>
<td>non-tenured</td>
<td>-2.40</td>
<td>1.245</td>
<td>.159</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>2.66</td>
<td>.975</td>
<td>.109</td>
<td></td>
</tr>
<tr>
<td>reg-tenured</td>
<td>non-tenured</td>
<td>-4.46</td>
<td>1.089</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>vol-tenured</td>
<td>-2.66</td>
<td>.975</td>
<td>.109</td>
<td></td>
</tr>
<tr>
<td>Dependent Variable</td>
<td>(I) Group</td>
<td>(J) Group</td>
<td>Mean Diff. (I-J)</td>
<td>Std. Error</td>
<td>Sig.</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>------------------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>FDBACK</td>
<td>non-tenured</td>
<td>vol-tenured</td>
<td>1.65</td>
<td>1.429</td>
<td>.515</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reg-tenured</td>
<td>4.51</td>
<td>1.249</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>vol-tenured</td>
<td>non-tenured</td>
<td>-1.65</td>
<td>1.429</td>
<td>.515</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reg-tenured</td>
<td>2.86</td>
<td>1.119</td>
<td>.040</td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>non-tenured</td>
<td>-4.51</td>
<td>1.249</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>vol-tenured</td>
<td>-2.86</td>
<td>1.119</td>
<td>.040</td>
</tr>
<tr>
<td>CONTEXT</td>
<td>non-tenured</td>
<td>vol-tenured</td>
<td>1.02</td>
<td>.814</td>
<td>.459</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reg-tenured</td>
<td>2.43</td>
<td>.712</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>vol-tenured</td>
<td>non-tenured</td>
<td>-1.02</td>
<td>.814</td>
<td>.459</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reg-tenured</td>
<td>1.41</td>
<td>.638</td>
<td>.088</td>
</tr>
<tr>
<td></td>
<td>reg-tenured</td>
<td>non-tenured</td>
<td>-2.43</td>
<td>.712</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>vol-tenured</td>
<td>-1.41</td>
<td>.638</td>
<td>.088</td>
</tr>
</tbody>
</table>

**Note.** Mean differences at the p < .05 level are in bold type.

There were no statistically significant differences (p < .05) between the mean scores of the volunteer tenured teachers (38.51) and the regular tenured teachers (38.20) on the TEACHERS subscale. There was a statistically significant difference (p < .05) between the non-tenured
teachers' mean score (36.40) and the volunteer teachers' mean score (38.51) and between the non-tenured teachers' mean score (36.40) and the regular tenured teachers mean score (38.20) on the TEACHERS subscale. The non-tenured teachers had the lowest mean score (36.40) of the three groups on the TEACHERS subscale.

On the EVALS subscale there was no statistical difference (p<.05) between the mean scores of the non-tenured (48.57) and volunteer tenured (49.84) teachers' scores. There was a statistically significant (p<.05) difference between the mean scores of the volunteer tenured teachers (49.84) and the regular tenured teachers (45.72). The volunteer tenured teachers had a statistically higher mean score for the EVALS subscale than the regular tenured teachers.

On the PROCS subscale there was no statistically significant difference between the mean scores of the non-tenured (32.98) and volunteer tenured teachers (30.58). There was a statistically significant difference (p<.05) between the non-tenured teachers mean score (32.98) and the regular tenured teachers mean score (28.52). The non-tenured teachers had a statistically higher mean score for the PROCS subscale than the regular tenured teachers.
On the FDBACK subscale there was no significant difference (p<.05) between the mean scores for the non-tenured (35.93) and the volunteer tenured (34.28) teachers. There was a statistically significant (p<.05) difference between the mean scores of the non-tenured teachers (35.93) and the regular tenured teachers (31.42) and between the volunteer tenured teachers (34.28) and the regular tenured teachers (31.42). The regular tenured teachers had the lowest mean score on the FDBACK subscale of the three groups.

On the CONTEXT subscale there were no statistically significant differences (p<.05) both between the mean scores of the non-tenured teachers (17.67) and the volunteer tenured teachers (16.65) and the mean scores of the volunteer tenured teachers (16.65) and the regular tenured teachers (15.23). There was a statistically significant (p<.05) difference between the non-tenured teachers' mean score (17.67) and the regular tenured teachers' mean score (15.23). The non-tenured teachers' mean score was significantly higher than the regular tenured teachers' mean score.

Summary of findings for research question #5.

On the TEACHERS subscale the non-tenured teachers' mean score (36.40) was statistically (p<.05) lower than
both the volunteer tenured teachers’ (38.51) mean score and the regular tenured teachers’ (38.20) mean score.

On the EVALS subscale, the volunteer tenured teachers had a statistically higher (p<.05) mean score (49.84) than the regular tenured teachers (45.72).

On the PROCS subscale, the non-tenured teachers had a statistically significant (p<.05) mean score (32.98) higher than that of the regular tenured teachers (28.52).

On the FDBACK subscale the regular tenured teachers had a statistically significant (p<.05) mean score lower than that of either of the other two groups.

On the CONTEXT subscale the non-tenured teachers’ had a statistically significant (p<.05) mean score (17.67) higher than that of the regular tenured teachers (15.23).
CHAPTER V

Summary, Discussion, Conclusions, and Recommendations

Introduction

This study examined the perceptions of three groups of teachers who were evaluated with two different evaluation processes and their perceptions regarding the overall quality and impact of their most recent evaluation experience and the attributes of a teacher evaluation system designed to promote professional growth. This chapter includes 4 sections: Introduction, Summary, Discussion and Implications, and Recommendations.

Summary

Purpose of the study.

The purpose of this study was to increase the knowledge base regarding the key attributes of teacher evaluation systems that foster professional growth in terms of effective instructional practices.

Statement of the problem.

The goal of teacher evaluations, both formative and summative, has historically served two purposes: establishing minimum competency for the purposes of
personnel decisions and the promotion of professional growth (Scriven, 1973; Lortie, 1975; Owens, 1991; Bridges, 1992; Wang, et al., 1993). A review of the literature indicated that the process of teacher evaluation has had very little success with the goal of promoting professional growth (Duke and Stiggins, 1990; Guskey, 1988; Peterson, 1995). Most recently, a series of alternative methods of teacher evaluation has been developed (e.g., professional growth portfolio, self-assessment, student assessment) and implemented in some public schools. However, the large majority of schools continue to rely on the traditional model of principal as evaluator. In this traditional model, the principal observes teachers and their performance is evaluated based on a set of predetermined criteria (usually a checklist format). A conference is usually held in which the evaluator reviews the completed checklist with the teacher. This model is widely used in New Jersey because such a model is spelled out in the New Jersey Administrative Code. Districts vary in their attempts to make this a meaningful process for professional growth (Capasso, et al., 1996).
Research questions.

This study addressed the five research questions. These questions are listed in Chapter IV on pages 71 and 72.

Description of the sample.

The school district involved in the study was a suburban, public school district located in central New Jersey. At the time of the study the district served a student population of 3500 students in kindergarten through grade 8 in 7 school buildings (5 k-5 elementary schools and two 6-8 middle schools). Each school building was administered by a building principal who had direct responsibility for the observation and evaluation of the building’s professional staff.

The professional staff in the district were evaluated according to the procedures required in The New Jersey Administrative Code, Title 6 (N.J.A.C. 6:3-4.1 and 6:3-3.3). The Code requires non-tenured teachers to be observed three times a year. Tenured teachers are to be observed one time a year. Written annual evaluations, which include individual improvement plans, are to be completed for all staff at the end of each school year. In the district under study, all observations are preceded by a conference between the teacher and the building principal.
to discuss the goals and objectives of the lesson to be observed. Following the observation, postconferences are held to discuss the observation, to set goals for improvement and to review the written observation report.

The school district involved in this study began a revision of its teacher evaluation system in 1996-1997. A 15 member committee was charged with developing a revised teacher evaluation system that would reflect both a growth model and the district’s commitment to excellence. The ultimate aim of the new process was to be the realization of effective teaching and learning. The committee developed a new evaluation instrument to be used for the first time during the 1997-98 school year (Appendix B).

During the summer of 1997 district administrators received training in the use of the newly developed evaluation tool. Twelve additional training sessions for administrators were conducted during 1997-98 school year. These training sessions focused on instructional skills, inter-rater reliability and the writing of observation reports and annual summaries.

The proposed plan for implementing the new evaluation instrument stated it should be used with all non-tenured teachers and volunteer tenured teachers during the 1997-98 school year. All non-tenured teachers and all volunteer
tenured teachers who were evaluated using the new tool received training on the new instrument in the fall of 1997. During the 1997-98 school year, all remaining staff were evaluated using the old evaluation tool (see Appendix C). For the purpose of this study all teachers in the school district were surveyed. Such a sample provided data that were useful in comparing results within and among the three groups included in the study.

**Methods of research.**

The research methodology in this study was quantitative, as reflected in the reporting of the data gathered from the survey instrument, The Teacher Evaluation Profile (TEP). The survey data was summarized in three forms for each of the three groups: a frequency distribution summary, a profile of means and standard deviations and a correlational analysis. The correlational analysis examined the relationship between the ratings of the teachers in each of the three groups for each of the items on the survey and their ratings of the overall quality and overall impact of their most recent evaluation experience, and the relationship between the ratings of the teachers in each of the three groups in each of the survey subscale areas (attributes of the teacher, attributes of the evaluator, attributes of the evaluation
procedures, attributes of the feedback on teaching performance and attributes of the evaluation context) on the survey and their ratings of the overall quality and impact of their most recent evaluation experiences. In addition a one-way analysis of variance was used to determine any statistically significant differences among the mean scores of the three groups of teachers surveyed in terms of the ratings of the overall quality and impact of their most recent evaluation experience and in terms of their mean scores on each of the TEP subscales.

A survey research method was used to answer the research questions posed in the study. This method allowed this researcher to gather data from a relatively large sample.

Summary of the findings.

All three teacher groups had favorable ratings for both the quality and impact of their most recent evaluation experiences. However, the only statistically significant difference between the mean scores of the three groups was the mean score related to overall impact for the non-tenured group which was higher than that of the two other groups (see Table 5 p. 78).

All three groups had positive, statistically significant correlations between their rating of overall
quality of the evaluation experience and the attributes of
the evaluator, the attributes of the procedures used during
the evaluation experience, the attributes of the feedback
they received and the attributes of the evaluation context
(see Table 6 p.81).

Once again, the non-tenured teachers differed from the
two tenured groups because the non-tenured teachers were
the only group for whom there was a statistically
significant correlation between overall quality and the
attributes of the teacher.

Overall, the subscale for the attributes of the
evaluator and the subscale for the attributes of the
feedback the teachers received had the highest percentages
of individual items significantly correlated to the rating
of overall quality (see Tables 7-11 pp. 82-89).

In terms of the correlation between overall impact and
the 5 subscales there were no statistically significant
correlations for the volunteer tenured group (see Table 12
p.94). However, for the non-tenured and regular tenured
groups there were statistically significant correlations
between overall impact and the attributes of the
procedures, the attributes of the feedback and the
attributes of the evaluation context (see Tables 13-17
pp.96-103). The tenured teacher group had the highest
percentage of individual items on the subscales significantly correlated to overall impact.

On all subscales, except the attributes of the teacher, the mean scores for the regular tenured teachers were lower than the mean scores of the other two groups (see Table 18 p. 104).

Discussion and Implications

The findings of this study indicated that the three groups of teachers did not differ significantly regarding their rating of the overall quality of their most recent evaluation experience. Since all three groups of teachers had mean scores for overall quality above the midpoint on the 9-point scale, the data seems to support a conclusion that these teachers had a positive attitude about the quality of their most recent evaluation experience. Since there were no statistically significant differences between the three groups' mean scores for the rating of overall quality of the evaluation experience, the data supported the conclusion that the rating was not dependent on whether or not the groups were evaluated using the new or traditional evaluation processes.

These findings were not consistent with the literature on teachers' attitudes towards evaluation because historically, teachers have mistrusted evaluation (Wolf,
1991 and Johnson, 1990) and have not been enthusiastic about the quality of the feedback they received (Capasso, Monahan, and White, 1996). The groups of teachers in this study did reflect a positive attitude toward their evaluation experience. The literature indicated that in those cases in which teachers rated the quality of their evaluations highly, this rating was dependent on the perceived skill, integrity and caring of the evaluator (Duke and Stiggins, 1990; McGreal, 1988; McLaughlin and Pfeifer, 1986). The literature also indicated that teachers viewed the quality of their evaluation experience more positively when the purpose of evaluation was clearly stated, when procedures were in place that were in alignment with this purpose, and when both the purpose and procedures were sensitive to the uniqueness of the working environment of the teachers (Wise et al., 1984; Rosenholtz, 1991). Data from this study related to research questions three and four support the conclusion that some of the factors reported in the literature (McLaughlin and Pfeifer, 1986; Guskey, 1988, McGreal, 1988; Duke and Stiggins, 1990) as contributing to a positive rating of teacher evaluation were present in the evaluation process of the district in this study.
Overall, the findings seem to indicate that the three groups of teachers held a favorable rating in regard to the overall impact of their most recent evaluation experience, once again supporting the conclusion that factors other than the evaluation instrument influenced these teachers’ views of the impact of their evaluation experience (McLaughlin and Pfeiffer, 1986; Guskey, 1988; McGreal, 1988; Duke and Stiggins, 1990).

The data indicate that there were statistically significant differences between the mean scores of the non-tenured teachers and both the volunteer and regular tenured teacher groups. Since there were no statistically significant differences found between the two tenured groups of teachers, the data may support the conclusion that the differences in impact rating were not dependent on the use of the new model or the traditional model but on whether the teachers were tenured or not. The non-tenured teachers may have viewed themselves as having more need for professional growth.

The findings indicate that the non-tenured teachers had positive, statistically significant correlations between their rating of overall quality and each of the 5 subscales: attributes of the teacher, attributes of the evaluator, attributes of the procedures used during the
evaluation process, attributes of the feedback received and attributes of the evaluation context. Both the volunteer tenured teachers and the regular tenured teachers had positive, statistically significant correlations between their rating of quality and the attributes of the evaluator, the attributes of the procedures used during the evaluation process, the attributes of the feedback received and the attributes of the evaluation context. Those more experienced teachers did not seem to credit their positive rating of the quality of their evaluation experience to their attributes as teachers. This finding was not consistent with Duke and Stiggins (1990) finding that teacher attributes contributed to a positive rating of the evaluation experience. The non-tenured teachers reported that they felt they were oriented to experimentation in their classrooms and that their previous experiences with evaluation had been helpful. Even though the attributes of the teacher, as a whole subscale, was not significantly correlated to their rating of quality, the regular tenured teacher group had a statistically significant relationship between their quality rating and their report that their previous experience with evaluation was helpful to them. This finding provided additional support to the conclusion that some of the factors reported
in the research as contributing to a positive rating of teacher evaluation were present in the evaluation process of the district in this study. The fact that the subscales dealing with the attributes of the evaluator, the attributes of the evaluation procedures, the attributes of the feedback on teaching performance and the attributes of the evaluation context all had positive, statistically significant correlations to the favorable ratings of the quality of the most recent evaluation experience for all three groups of teachers was consistent with the findings of the research in teacher evaluation (McLaughlin and Pfeiffer, 1986; Guskey, 1988; McGreal, 1988; Duke and Stiggins, 1990).

The finding that the subscales of the attributes of the evaluator, and the attributes of the feedback received had the highest percentages of individual attributes significantly correlated to the favorable ratings of quality for all three groups is additional evidence consistent with the research in teacher evaluation. This finding also seems to support the conclusion that the evaluation instrument itself was not the determining factor in the positive rating for the quality of the evaluation experience in all three groups.
The non-tenured teachers reported statistically significant relationships between the attributes of the procedures used during the evaluation, the attributes of the feedback received and the attributes of the evaluation context subscales and their favorable rating of impact on their professional growth.

None of the three groups reported a significant correlation between their attributes as teachers and the impact of their most recent evaluation experience. The fact that all three groups had rated the impact of their most recent evaluation as favorable, this proved to be inconsistent with the research (Duke and Stiggins, 1990) that teacher attributes play a role in evaluation processes that foster professional growth.

Although the volunteer tenured teachers had a mean score that was favorable regarding the overall impact of their most recent evaluation experience, they did not report any of the subscale areas as statistically significant in relationship to impact.

The largest of the three groups, the regular tenured teachers (n= 128), reported statistically significant relationships between their rating of impact and the attributes of the evaluator, the attributes of the procedures used in the evaluation, the attributes of the
feedback received and the attributes of the evaluation context. This finding is consistent with the research on the characteristics that appeared to be present in teacher evaluation models that were effective in promoting professional growth (McLaughlin and Pfeiffer, 1986; Guskey, 1988; McGreal, 1988; Duke and Stiggins, 1990). The regular tenured teachers also reported the largest percentage of individual items on these subscales significantly correlated to their rating of overall impact. Once again, this finding supports the conclusion that the evaluation instrument itself seemed not to be the determining factor in the rating of the quality and impact of the evaluation experience. In this study the attributes of the feedback (e.g. amount of information received, quality of the ideas and suggestions contained in the feedback, feedback focused on teaching strategies) were reported to have had the greatest influence on the rating of the impact of the evaluation experience for the largest number of teachers.

In addition to the differences discussed above, this study also investigated if there were statistically significant differences among the groups in terms of their mean scores on the 5 subscales.

The mean scores on the attributes of the teacher subscale reflect the non-tenured teachers' significantly
lower scores as might be expected due to their inexperience (item #8 on TEP) and possible lack of confidence in their knowledge about the technical aspects of teaching (item #6 on the TEP). The attributes of the evaluator, the attributes of the procedures used in the evaluation, the attributes of the feedback received and the attributes of the evaluation context subscales represent aspects of the evaluation process other than the teachers themselves. On all 4 of these subscales the teachers participating in the new evaluation process had mean scores higher than those teachers participating in the traditional evaluation process. These data support the conclusion that the new evaluation process may have more of the attributes that the literature identifies in evaluation models that promote professional growth than the traditional process.

In general all three teacher groups had favorable ratings on the overall quality and impact of their most recent evaluation experience. This researcher believes that the district in this study probably had in place a number of the attributes related to teacher satisfaction with the evaluation process. There were significant differences between non-tenured teachers and the two tenured groups in both the mean scores on impact and the mean scores on the attributes of the teacher subscale. This researcher
believes that these differences were probably attributable to their inexperience and lack of knowledge about the technical aspects of teaching. Statistically significant differences were found between the mean scores of both groups evaluated using the new process and the group evaluated using the traditional process only on the attributes of the feedback received subscale. However, the mean scores on all of the 4 subscales reflecting attributes of the evaluation process (attributes of the evaluator, attributes of the procedures used in the evaluation, attributes of the feedback received, attributes of the evaluation context) were lower for the regular tenured teacher group. These data suggest that the new evaluation process had more attributes reflective of an evaluation model that promotes professional growth than the traditional model.

Recommendations

Recommendations for the district.

1. The district in this study began a revision of its teacher evaluation system in 1996-1997 that resulted in the development of a new instrument for evaluation as well as a staff development component for both teachers and evaluators. As part of the revision process a representative committee of teachers and administrators
worked to develop a set of objectives for the realization of an evaluation process that would impact effective teaching and learning. The largest group of teachers in this study reported that the standards for teacher evaluation were clear to them and that they considered these standards to be appropriate for their classrooms. Even though these teachers were not evaluated with the new instrument that clearly spells out in detail the characteristics of effective instruction, they expressed clarity about the overall goals of teacher evaluation in this district. Research supports the need to establish criteria for effective teaching that match the educational goals and values of the school community (Wise et al., 1984). The district should continue to provide this kind of clarity in regard to the objectives of the teacher evaluation process.

2. The staff development component that involved the training of both teachers and administrators should also be continued. It provides an ongoing vehicle to clarify the objectives for staff evaluation as well as a means to build the attributes of the evaluators and the attributes of the kind of feedback they offer to the teachers in the district. The research (McLaughlin and Pfeiffer, 1986; Guskey, 1988; McGreal, 1988; Duke and Stiggins, 1990)
clearly states that both the attributes of the evaluator and the attributes of the feedback received by the teachers have an impact on effective teacher evaluation models.

3. The district should consider more ways to involve the teachers in setting their own learning goals. For the most part, teachers in this district do not see their own attributes as contributing to the overall quality or impact of the evaluation process. Both the research on adult learning (Knowles, 1980) and effective evaluation practices point to involving teachers in setting professional improvement goals. The development of the annual individual improvement plan could be an effective tool to build a collaborative process in which teachers play an active role in setting and self-evaluating their professional goals.

4. The district has identified the development of a comprehensive supervision plan that includes the opportunity to use multiple measures of evaluation as one of its objectives (see Chapter Three Description of the Sample). The district should continue to pursue this objective as it may assist in addressing the above mentioned third recommendation. It is assumed that the tenured teachers in this district have achieved at least
a minimum level of competency. The primary purpose of the evaluation process for this group of teachers is not focused on making hiring decisions but on the improvement of instruction. A multiple options approach for the more experienced teacher seems more appropriate.

5. Since the data in this study support the conclusion that the attributes of the evaluator and the attributes of the feedback are keys to the effectiveness of the evaluation process, the district needs to consider whether or not adequate staff development, personnel, and time resources are committed to the evaluation process so that these attributes can best be developed and nurtured. Wise and associates (1984) found that a top-level commitment to and resources for evaluation outweighed checklists and procedures.

6. The district should continue its practice of interviewing only those candidates with a depth and breadth of instructional experience. As part of its interviewing process the district should continue to include a lesson observation component that would involve both the candidates' observations of a lesson and a written summary.
Recommendations for further research.

1. The field of teacher evaluation has expanded over the years to include a variety of models. Although New Jersey public schools are bound by Code to follow certain procedures in the teacher evaluation process, that does not preclude districts from investigating alternative means of evaluation. Further research is needed in such evaluation models as teacher portfolios, peer evaluation and self-evaluation to identify characteristics of these models that can contribute to the body of knowledge about the characteristics of effective evaluation models.

2. Further research is needed in models of teacher evaluation that are designed for different stages of a teacher’s career. There is some evidence that the shape of evaluation should be different for the more experienced teacher (Mezirow, 1990; Glickman et al., 1995).

3. Research indicates (McLaughlin and Pfeiffer, 1986; Guskey, 1988; McGreal, 1988; Duke and Stiggins, 1990) that the characteristics of the evaluator are key to the effective evaluation model. Interpersonal skills, as well as, those skills related to knowledge of instruction, observation, and the provision of feedback
appear to be more important in the evaluation process that the specific instrument used in the process.

Districts tend to spend time and resources developing forms and checklists rather than developing support and training programs for evaluators. Evaluators need to be provided with training in human relations skills and in observing teachers, recording observation data, and in providing specific feedback. More research is needed regarding training and support programs that effectively develop these characteristics. In addition, more research is needed in regard to effective evaluation programs for evaluators. Superintendents and assistant superintendents, charged with the evaluation of principals and supervisors who are in turn evaluating teachers, need more information about best practices in this area.

4. The results of this study indicate support for the conclusion that the overall impact rating of the evaluation process was effected by whether the teachers completing the TEP were tenured or not. The non-tenured teachers had a statistically significant higher impact rating. In addition, on the TEACHERS subscale the non-tenured teachers had the lowest mean score for knowledge of subject matter and the highest mean scores for
orientation to change and openness to criticism. More research is needed in the area of new teacher induction programs that can tap into the inexperienced teachers' awareness of the gaps in their knowledge and their willingness to respond to change and criticism.

5. The results of this study indicate that the TEACHERS subscale was perceived by none of the three teacher groups as related to impact on teacher professional growth. Research may be needed to further refine the attributes in this subscale.

6. If the major focus of teacher evaluation is on professional growth in terms of effective instructional practices, attention needs to be paid to whether or not there is a link to student achievement. If professional growth leads to improved instruction, does increased student achievement follow? Further research, similar to that underway by the School Improvement Model at Iowa State University (Manatt and Benway, 1998), into the inclusion of student achievement data as part of the evaluation process is needed.

This researcher believes that one of the most important recommendations resulting from this study is the need for further research into the characteristics of programs that develop effective skills for teacher evaluators.
REFERENCES


Capasso, R. L., Monahan, T. C., & White, E. H. (1996,


teachers’ preferences concerning their evaluation. Educational Administration Quarterly, 20, 70-106.


APPENDIX A

THE TEACHER EVALUATION PROFILE
THE TEACHER EVALUATION PROFILE

A Questionnaire Reviewing Your Most Recent Teacher Evaluation Experience
INSTRUCTIONS

This form has been designed to allow you to describe your experience with the district's teacher evaluation process. Your responses will be combined with those of other teachers to yield a picture of the key ingredients in an effective teacher evaluation experience. The goal of this research is to determine how the evaluation process can be revised to help it serve relevant and useful purposes. If we are to reach this goal, it will be important for you to provide frank and honest responses. This is why your answers will remain anonymous.

This questionnaire has been designed to be comprehensive in scope. It will take approximately 15 minutes to complete. It is crucial that you read and follow these instructions very carefully.

The Definition of Teacher Evaluation

The process leading to the annual evaluation summary usually consists of a professional improvement plan, classroom observation, and conferencing between teacher and supervisor before and after the observation. When reference is made in this questionnaire to teacher evaluation, it should be understood to encompass all these elements.

IMPORTANT: PLEASE READ BEFORE FILLING OUT TEACHER EVALUATION PROFILE

SPECIFIC INSTRUCTIONS:

Please enter your responses using a No. 2 pencil on the NCS Response Form No. 16432 which you have been given.

• Find the side of the form that is printed in brown.

• Then find the Identification Number Box in the Upper Left Corner. In the far right column (the last column in the ID Number Box) fill in the column accordingly:

1 = if you are a nontenured teacher

2 = if you are a tenured teacher evaluated with the new/pilot evaluation instrument during the 1997-98 school year

3 = tenured teacher evaluated using the old/traditional instrument during the 1997-98 school year

• Be sure to blacken the corresponding circle.

COLUMN A: Special Codes – Rating Quality of Evaluation

• Find the Special Codes Box to the right of the ID Box.
Given the above definition of teacher evaluation, please reflect on the last time you were evaluated; that is, your most recent experience with your teacher evaluation system. Regard the entire evaluation process, including planning for evaluation, classroom observations, and feedback. As you think about this experience, how would you rate the overall quality of the evaluation process? Use a scale of 0 to 9, with 0 representing very poor quality and 9 very high quality.

- Using a No. 2 pencil, please write your Rating of Overall Quality in Column A of the SPECIAL CODES BOX. Then blacken the corresponding circle for that rating in Column A below.

<table>
<thead>
<tr>
<th>IDENTIFICATION NUMBER</th>
<th>DATE (DAY YEAR)</th>
<th>SPECIAL CODES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A  B  C  D  E  F  G</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>1  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>2  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>3  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>4  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>5  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>6  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>7  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>8  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0</td>
<td>9  0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

COLUMN B: Special Codes — Rating Impact of Evaluation

In the next column of the SPECIAL CODES BOX, Column B, rate the Overall Impact of your last evaluation on your professional practices. A low rating of 0 would reflect no impact at all and no changes in your practices, attitudes and/or understanding. A high rating of 9 would reflect a strong impact leading to profound changes in your teaching practices, attitudes, about teaching, and/or understanding of the teaching profession.

DO NOT RECORD YOUR NAME

GENERAL INSTRUCTIONS: Rating Attributes of Evaluation

Now, please use the scales provided on the following pages to describe yourself and the nature of your most recent teacher evaluation experience. Do this by:

- Considering each of the 44 attributes to be described,
- Studying the scale to be used to describe each,
- Selecting the letter that represents the point you select on each scale, and
- Coding that letter on the NCS form.

BE SURE THAT NUMBER OF THE ATTRIBUTE YOU ARE DESCRIBING CORRESPONDS TO THE NUMBER ON THE RESPONSE SHEET WHERE YOU ENTER YOUR RESPONSE.
A. **Describe these attributes of you as a teacher:**

1. Rate the strength of your professional expectations of yourself
   - I demand little 1 2 3 4 5
   - I demand a great deal

2. Orientation to risk taking
   - I avoid risks 1 2 3 4 5
   - I'm relatively flexible

3. Orientation to change
   - I'm relatively slow to change 1 2 3 4 5
   - I'm relatively flexible

4. Orientation to experimentation in your classroom
   - I don't experiment 1 2 3 4 5
   - I experiment frequently

5. Openness to criticism
   - I'm relatively closed 1 2 3 4 5
   - I'm relatively open

6. Knowledge of technical aspects of teaching
   - I know a little 1 2 3 4 5
   - I know a great deal

7. Knowledge of subject matter
   - I know a little 1 2 3 4 5
   - I know a great deal

8. Years of experience in position
   - 1: 0 to 1 year
   - 2: 2 to 5 years
   - 3: 6 to 10 years
   - 4: 11 to 15 years
   - 5: 16 or more years

9. Experience with teacher evaluation prior to most recent experience
   - Waste of time 1 2 3 4 5
   - Very helpful

B. **Describe your perceptions of the person who evaluated your performance (most recently):**

10. Credibility as a source of feedback
    - Not credible 1 2 3 4 5
    - Very credible

11. Working relationship with you
    - Adversary 1 2 3 4 5
    - Helper

12. Level of trust
    - Not trustworthy 1 2 3 4 5
    - Trustworthy

13. Interpersonal manner
    - Threatening 1 2 3 4 5
    - Not Threatening

14. Temperament
    - Impatient 1 2 3 4 5
    - Patient
15. **Flexibility**

| Rigid | 1 2 3 4 5 | Flexible |

16. **Knowledge of technical aspects of teaching**

| Not Knowledgeable | 1 2 3 4 5 | Knowledgeable |

17. **Capacity to demonstrate or model needed improvements**

| Low | 1 2 3 4 5 | High |

18. **Familiarity with your particular classroom**

| Unfamiliar | 1 2 3 4 5 | Familiar |

19. **Familiarity with classrooms in general**

| Unfamiliar | 1 2 3 4 5 | Familiar |

20. **Usefulness of suggestions for improvements**

| Useless | 1 2 3 4 5 | Useful |

21. **Persuasiveness of rational for suggestions**

| Not Persuasive | 1 2 3 4 5 | Very Persuasive |

---

**C. DESCRIBE THESE ATTRIBUTES OF THE PROCEDURES USED DURING YOUR MOST RECENT EVALUATION:**

(1) **What procedures were used to address the dimensions of your TEACHING (standards) to be evaluated?**

22. Were standards communicated to you?

| Not At All | 1 2 3 4 5 | In Great Detail |

23. Were the standards clear to you?

| Vague | 1 2 3 4 5 | Clear |

24. Were standards endorsed by you as appropriate for your classroom?

| Not Endorsed | 1 2 3 4 5 | Endorsed |

25. Were the standards...

| All the Same For All Teachers | 1 2 3 4 5 | Tailored Somewhat For Your Unique Needs |

(2) **To what extent were the following sources of performance information tapped as part of the evaluation?**

26. **Observation of your classroom performance**

| Not Considered | 1 2 3 4 5 | Used Extensively |

27. **Examination of classrooms or school records (lesson plans, etc.)**

| Not Considered | 1 2 3 4 5 | Used Extensively |

28. **Examination of student achievement**

| Not Considered | 1 2 3 4 5 | Used Extensively |
(3) Extent of observation in your classroom, based on your most recent experience:

(Note: In these items, FORMAL refers to observations that were preannounced and were preceded and followed by a conference with the evaluator; INFORMAL refers to unannounced drop-in visits.)

29. Number of FORMAL (prescheduled) observations per year
   1: 0
   2: 1
   3: 2
   4: 3
   5: 4 or more

30. Approximate frequency of INFORMAL (unannounced drop-in) observations
   1: None
   2: Less than 1 per month
   3: Once per month
   4: Once per week
   5: Daily

D. Please describe these attributes of the feedback you received:

31. Amount of information received
    None 1 2 3 4 5 Great Deal

32. Frequency of formal feedback
    Infrequent 1 2 3 4 5 Frequent

33. Frequency of informal feedback
    Infrequent 1 2 3 4 5 Frequent

34. Depth of information provided
    Shallow 1 2 3 4 5 Indepth

35. Quality of the ideas and suggestions contained in the feedback
    Low 1 2 3 4 5 High

36. Specificity of information provided
    General 1 2 3 4 5 Specific

37. Nature of information provided
    Judgmental 1 2 3 4 5 Descriptive

38. Timing of feedback
    Delayed 1 2 3 4 5 Immediate

39. Feedback focused on district teaching standards
    Ignored Them 1 2 3 4 5 Reflected Them
E. Describe these attributes of the evaluation context:

40. Amount of time spent on the evaluation process, including your time and that of all other participants

<table>
<thead>
<tr>
<th>None</th>
<th>Great Deal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

Resources available for professional development:

41. Time allotted during the teaching day for professional development

<table>
<thead>
<tr>
<th>None</th>
<th>Great Deal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

42. Availability of training programs and models of good practice

<table>
<thead>
<tr>
<th>None</th>
<th>Great Deal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

District values and policies in evaluation:

43. Clarity of policy statements regarding purpose for evaluation

<table>
<thead>
<tr>
<th>Vague</th>
<th>Clear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

44. Intended role of evaluation

<table>
<thead>
<tr>
<th>Accountability</th>
<th>Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

Teacher Growth

THANK YOU FOR YOUR THOUGHTFUL RESPONSES!!
APPENDIX B

THE NEW TEACHER EVALUATION INSTRUMENT
**Vision Statement**

Supervision and Evaluation is a collaborative process of continuous professional growth and development that promotes the effectiveness of instruction and learning.

**Mission Statement**

The focus of supervision and evaluation within the School District is the promotion of professional growth and excellence. This is achieved through a continuous, collaborative process that encourages knowledge of and competence in instructional skills. The ultimate aim is the realization of effective and rewarding teaching and learning.

**Goals and Objectives**

**Goals**

The Instructional Excellence Model is designed to promote:

- professional growth and excellence while meeting state mandates and district policies
- effective and rewarding teaching and learning
- a process that is professionally meaningful

**Objectives**

- To provide a professional profile that establishes criteria for effective teaching, reflective of acknowledged principles of learning and methods of instruction
- To develop and implement procedures that support collaboration, collegiality, and communication within the educational team
- To develop a model that encourages observation, analysis, questioning, and conferencing skills as a means of promoting reflection and self-analysis
- To provide training for the supervision/evaluation process
- To develop a comprehensive supervision plan that includes the opportunity to utilize multiple measures and divergent/alternative means of assessment
- To develop a comprehensive evaluation plan that specifies the procedures for the performance review of staff
- To design procedures to monitor and evaluate the supervision and evaluation process
### Key:
- **U**: Unsatisfactory indicates absence of requisite competence
- **B**: Basic indicates fundamental competence
- **P**: Proficient indicates thorough competence
- **E**: Expert indicates extraordinary proficiency
- **NO/A**: Not observed/Not applicable

### Description of Lesson (Focus, Instructional Organization):

#### Curriculum and Instruction
**Teaching Strategies/Techniques**
- Acted as a decision-maker who made appropriate instructional choices
- Provided a wide variety of instructional experiences and activities which addressed diverse learning styles
- Included questions, activities, and experiences which promoted creative, critical, and higher-level thinking skills
- Utilized strategies which promoted the rate and degree of learning
- Demonstrated enthusiasm for learning and communicated that to students
- Maintained and communicated high expectations to students
- Promoted desirable work habits and study skills
- Evaluated student performance for the improvement of instruction and assessment

#### Comments:

#### Knowledge of Subject Area/Content
- Was knowledgeable and current with regard to the curriculum
- Exhibited knowledge of subject matter

#### Comments:

#### Planning and Preparation
- Developed plans which reflected the goals, objectives and content of the curriculum
- Developed lessons and units which evidenced careful, thorough planning and preparation
CLASSROOM ENVIRONMENT AND MANAGEMENT

- Created a positive learning environment which was stimulating and purposeful
- Developed and maintained an atmosphere of mutual respect within the classroom
- Promoted students' feeling of self worth
- Demonstrated effective classroom management skills
- Maintained discipline that was fair and consistent
- Maintained an instructional setting conducive to safe and effective learning

Comments:

ADDITIONAL COMMENTS:

Summary of Post Observation Conference:

I have seen and read this report.

_________________________          ________________________
Teacher's Signature/Date          Supervisor's Signature/Date

Copies to be distributed to Teacher, Principal, and Supervisor of Personnel.
**SCHOOLS**

**ANNUAL PERFORMANCE SUMMARY**

Teacher's Name:  

School:  
Grade/Subject:  
School Year:  

Observation Dates:  

Attendance as of:  
Sick:  
Professional:  
Other:  

Key:  
U: Unsatisfactory indicates absence of requisite competence.  
B: Basic indicates fundamental competence.  
P: Proficient indicates thorough competence.  
E: Expert indicates extraordinary proficiency.  

<table>
<thead>
<tr>
<th>Curriculum and Instruction</th>
<th>U</th>
<th>B</th>
<th>P</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Teaching Strategies/Techniques</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Acts as a decision-maker who makes appropriate instructional choices</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Provides a wide variety of instructional experiences and activities which address diverse learning styles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Includes questions, activities, and experiences which promote creative, critical, and higher-level thinking skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Utilizes strategies which promote the rate and degree of learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Demonstrates enthusiasm for learning and communicates that to students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Maintains and communicates high expectations to students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Promotes desirable work habits and study skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Evaluates student performance for the improvement of instruction and assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments:

Knowledge of Subject Area/Content  

- Is knowledgeable and current with regard to the curriculum  
- Exhibits knowledge of subject matter  

Comments:

Planning and Preparation  

- Develops plans which reflect the goals, objectives and content of the curriculum  
- Develops lessons and units which evidence careful, thorough planning and preparation
CLASSROOM ENVIRONMENT AND MANAGEMENT

- Creates a positive learning environment which is stimulating and purposeful
- Develops and maintains an atmosphere of mutual respect within the classroom
- Promotes students' feeling of self worth
- Demonstrates effective classroom management skills
- Maintains discipline that is fair and consistent
- Maintains an instructional setting conducive to safe and effective learning

Comments:

PROFESSIONALISM

- Maintains the ethical standards of the profession
- Demonstrates a professional attitude in all areas of responsibility
- Projects a positive, professional image to students, colleagues and the community
- Demonstrates a commitment to continuing professional development

Comments:

☐ TEACHER'S SUPPLEMENTAL COMMENTS attached (optional)

PROFESSIONAL IMPROVEMENT PLAN

Focus:

☐ Unsatisfactory  ☐ In Progress  ☐ Satisfactorily Completed  ☐ Completed Beyond Expectation

Comments:

INDICATORS OF PUPIL PROGRESS AND GROWTH

Among the indicators of pupil progress and growth used to evaluate student performance are the following: teacher observation, parental or guardian interview, formal and informal evaluation.
techniques, cumulative pupil records, portfolios, state testing results, and visual, auditory and/or medical examinations.

In reviewing the teacher’s use of these indicators, it has been determined that the teacher □ has □ has not effectively incorporated them as a part of the program of instruction and evaluation.

SUMMARY

RECOMMENDATION

□ Award contract with Increment □ Award contract/Withhold Increment □ Deny Contract

I have seen and read this summary.

Teacher’s Signature/Date

Supervisor’s Signature/Date

Copies to be distributed to Teacher, Principal, and Supervisor of Personnel!
SCHOOLS
SUMMARY OF PROFESSIONAL IMPROVEMENT PLAN
(To be completed by Teacher)

Name:

School: Grade/Subject: School Year:

Objective:

Activities Completed:

Comments:
<table>
<thead>
<tr>
<th>Teacher's Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>School:</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
</tbody>
</table>

**Objective:**

**Staff Member's Activities:**

**Supervisor's Activities:**

**Interim Review date:**

**Final Review date:**

| Signature of Staff Member/Date | Signature of Administrator/Date |
APPENDIX C

THE TRADITIONAL EVALUATION FORM
TEACHER OBSERVATION AND APPRAISAL FOR INSTRUCTIONAL IMPROVEMENT

Teacher's Name:  
Observer:  
School:  
Grade/Subject:  
Date:  
Time: From:  To:  

KEY:  
1-A check in Box 1 indicates the observer was satisfied with the performance of the teacher.  
2-A check in Box 2 indicates the observer was not satisfied with the performance of the teacher.  
3-A check in Box 3 indicates the observer did not observe that particular area.  
4-A check in Box 4 indicates the area was not applicable in that particular setting.  
5-A check in Box 5 indicates that a comment has been made on the Supplemental Comment Sheet.

DESCRIPTION OF LESSON:

<table>
<thead>
<tr>
<th>TEACHER EFFECTIVENESS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Purpose of lesson was clearly understood by the class</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>B. Materials were being used appropriately</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>C. The atmosphere in the class was such that students felt free to ask questions and express opinions</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>E. Questions were clearly stated and thought provoking</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>F. The teacher provided prompt and positive feedback to students</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>G. The teacher demonstrated knowledge of the subject matter</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>H. The teacher demonstrated knowledge of the teaching process</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CLASSROOM SETTING</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Classroom was neat, attractive and conducive to learning</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>B. Rapport between teacher and students was evident</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>C. Classroom routines were well established</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PLANNING AND PREPARATION</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Plans were available and clearly written</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>B. Materials and supplies were readily available</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>C. Plans indicated a variety of appropriate activities</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STUDENT</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Students were being taught at their appropriate instructional levels</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>B. Students demonstrated respect for their peers and the teacher</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>C. Students functioned independently when appropriate</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>D. Students were courteous listeners</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>E. Students demonstrated respect for authority of the teacher</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PERSONAL CHARACTERISTICS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. The teacher exhibited good judgment, emotional maturity and a sense of humor</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>B. The teacher spoke in a properly modulated voice</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>C. The teacher used correct English in oral and written form</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>D. The teacher communicated information in a clear and concise manner</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
<tr>
<td>E. The teacher presented subject matter with enthusiasm and clarity</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
<td>[]</td>
</tr>
</tbody>
</table>

VI. COMMENTS: (Recommendations, suggestions for improvement)

VII. SUMMARY OF POST-OBSERVATION CONFERENCE
I have seen and read this evaluation:

<table>
<thead>
<tr>
<th>Teacher's Signature</th>
<th>Date</th>
<th>Observer's Signature</th>
<th>Date</th>
</tr>
</thead>
</table>

Please sign. Copies to be distributed to Teacher, Principal, and Assistant Superintendent.