A comparative study of the perceptions of novice Iowa public school teachers towards the effects of the Iowa Evaluator Approval Training Program

by

Bradley Allen Buck

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Education (Educational Administration)

Program of Study Committee:
Thomas L. Alsbury, Major Professor
Joanne Marshall
Gregory S. Miller
Mack C. Shelley II
Veronica Stalker

Iowa State University

Ames, Iowa

2006

Graduate College

Iowa State University

This is to certify that the doctoral dissertation of

Bradley Allen Buck

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Teacher evaluators in the state of Iowa were required to take part in the newly created Iowa Evaluator Approval Training Program (IEATP) in the 2002-03 school year. The IEATP was designed to train evaluators in teacher evaluation based on the Iowa Teaching Standards and Criteria. This research surveyed novice teachers from Iowa public schools regarding the five attributes of teacher evaluation using a modified Teacher Evaluation Profile (TEP) originally designed by Stiggins and Duke (1988). The attributes were defined through a number of different traits: (a) self as a teacher, (b) evaluator, (c) evaluation processes, (d) attributes of the feedback, and (e) context in which the evaluations occurred.

Data were also collected regarding the concept of potential bias in teacher evaluation based on the gender of the teacher and/or the gender of the evaluator. A comparison was made from year one of the study, before the implementation of IEATP (2001-02), to year two of the study, after the implementation of IEATP (2002-03).

Descriptive statistics included t-tests, pairwise comparisons, multivariate tests, within-subjects tests, and multiple analysis of variance [MANOVA]. Following Bonferroni correction, significant differences were observed in the study. Traits and non-traits of each of the five attributes of teacher evaluation emerged through an analysis of the data. In addition, a statistically significant gender interaction favoring female evaluators regarding a number of traits was also observed in the study. A comparison of changes in novice teacher perceptions from 2001-02 to 2002-03 revealed changes in perceptions in the attributes of self as a teacher and in the evaluation processes.

Suggestions for future research include the need for emphasis on training evaluators to identify specific behaviors that are a part of quality instruction. This is currently a knowledge and skill expectation in training Module One that needs additional attention in follow-up training. In addition, to increase training efficacy, real world application to the learning elements is needed to meet the needs of adult learning theory.

CHAPTER 1. INTRODUCTION

Reform of public education has become a mantra for political leaders at both the state and federal levels. Federal interventions into public education are not a new phenomenon. The first version of the Elementary and Secondary Education Act came about in 1965. At that time, the act established five *titles* that influenced public schools and their governing structures more than any legislation before that time (Cunningham, 1971).

A nation at risk (1983), written by the Members of the National Commission on Excellence in Education, was a seminal document in the call for educational reform. This Commission was created by then Secretary of Education, T. H. Bell, and given the charge to examine the quality of education in the United States and to report back to the Secretary within eighteen months of their first meeting. The commission was directed to pay particular attention to assessing the quality of teaching and learning, to comparing the United States to other industrialized countries, to identifying educational programs which result in notable student success in college, and to assessing the degree to which major social and educational changes in the last quarter century had affected student achievement.

The commission, relying on information from papers commissioned by experts, testimonies by those in the field, letters from concerned constituents, existing analyses of problems in education and descriptions of notable programs and promising practices, defined a variety of issues facing public education. With its findings of diluted curricula across the country, lowered expectations for student learning, ineffective uses of student learning time, poor teacher candidates and an unappealing teaching lifestyle, the report led to a call for action.

From that starting point, waves of reform have washed through American public education. It is clear that the efforts from each previous wave impacted the structure of the wave(s) that followed. The first wave of reform, in the 1980s, focused on student performance requirements and teacher quality, and it was largely characterized by top-down expectations that asked educators to do more of the same but to do it better (Farrar, 1990). The second wave of reform was characterized by recognition of the systemic nature of the educational system and the importance of putting the teacher at the center of educational reform. In this model, the teacher was viewed as the cause of learning and an emphasis was placed on empowering teachers in the context of the work environment (Petrie, 1990). The current wave of reform calls for the reform of administrator preparation (Jacobson, 1990).

An essential component of each wave of reform has centered on increased accountability for public schools. As policy makers and the public have become more concerned with the quality of American public education, there has been an increased emphasis on the outcomes of education and holding educators accountable for those outcomes (Brown, 1990).

Much like school reform, accountability for public education is not a new concept. Accountability was a recommendation from *A nation at risk* (1983):

> We recommend that citizens across the Nation hold educators and elected officials responsible for providing the leadership necessary to achieve these reforms, and that citizens provide the fiscal support and stability required to bring about the reforms we propose. (p. 32)

Dawson and Acker-Hocevar (1998) noted that educational accountability is a continuous expectation of the supporters of public education. At times, the outcry for quality education is loud; at other times, it is audible, but soft. Yet, at almost no time is it inaudible.

At the federal level, the current evidence for reform and accountability is contained within the reauthorization of the Elementary and Secondary Schools Act for 2002, perhaps better known as the No Child Left Behind (NCLB) Legislation. The legislation calls for, among other things, "highly qualified teachers", applying teaching strategies in the classroom that are "scientifically research-based", and the achievement of "proficiency" in reading, math, and science for all students by the year 2014.

The belief and action statements of NCLB include outstanding teaching as one of the core foundations for high-quality student success. Stronge and Tucker (2000), commenting on a study done by Bill Sanders, noted that when third graders were placed in high-performance teacher classrooms for three years in a row, they scored, on average, at the 96th percentile on Tennessee's statewide mathematics test at the end of fifth grade. Those students who were placed with low performing teachers for three years in a row scored an average of the 44th percentile on the same fifth grade test.

In an effort to improve teaching, many states have established state-mandated teacher evaluation systems. Included in that group would be states such as Tennessee, Missouri, Texas, Georgia, and Arizona. Tennessee and Missouri have state-mandated teacher evaluation programs that date to the early to mid 1980s. In Iowa, this push for reform and accountability is evidenced in the establishment of the Student Achievement and Teacher Quality Program.

The Student Achievement and Teacher Quality Program is quite overt in the belief that teacher quality is the key to student achievement gains. Senate File 476 (2001) noted:

> The legislature acknowledges that outstanding teachers are a key component in student success. The program's goals are to enhance student achievement and to re-design compensation strategies and teachers' professional

> development. Such compensation strategies are designed to attract and retain high performing teachers, to reward teachers for improving their skills and knowledge in a manner that translates into better student learning, and to reward the staff of school attendance centers for improvement in student achievement. (p. 1)

In tandem with this state legislation, Iowa is entering into a new phase of state-mandated teacher evaluation through the use of a standards-based model, the foundation of which is the Iowa Teaching Standards and the Iowa Professional Development Standards. The approach in Iowa is intended to be an integrated model infusing the Teacher Quality Program and Evaluator Approval. There is mentoring and induction programming to support teachers in the extremely important first years of teaching, teacher evaluation systems that document skills related to the Iowa Teaching Standards and Criteria designed to provide school districts with necessary information regarding teachers' skills, and an evaluation process that supports the development of individual career development plans for career teachers (Berger, 2002).

An additional component of the improved accountability system is the Data-Driven Leadership (DDL) training as well as the state of Iowa Teacher Evaluator Approval Program (ITEAP). The adoption of ITEAP is based on the premise that improved supervision of teaching is a necessary precursor to improved teaching. The ITEAP legislation calls for the following:

> The department shall establish an evaluator training program to improve the skills of school district evaluators in making employment decisions, making recommendations for licensure, and moving teachers through a career path as established under this chapter. The department shall consult with persons representing teachers, national board-certified teachers, administrators, school boards, higher education institutions with approved practitioner and administrator preparation programs, and with persons from the private sector knowledgeable in employment evaluation and evaluator training in order to develop standards and requirements for the program. Evaluator training

> programs offered pursuant to this chapter may be provided by a public or private entity. The department shall distribute a list of evaluator training program providers to each school district. (284.10 paragraph 1)

State mandates that teachers be evaluated and evaluator training programs be implemented are not new to Iowa. In 1976, the Iowa General Assembly modified procedures for terminating the contract of a teacher (Section 279.13 Code of Iowa) by passing H 6559 and amending SF 205. At that time, it was confirmed that the school boards had the power to establish the criteria or standards of the evaluation, but the procedures for evaluation were to be negotiated with the bargaining representative(s) for the district. Manatt (1976) noted that, while not all districts had to develop a new teacher evaluation system to satisfy SF 205, it was a good bet that all Iowa districts could enhance what existed… and could improve by giving skills training for those designated to be evaluators.

A revision of expectations for teacher evaluators was contained in the mandate of Senate Bill 2175 (Iowa Legislative Assembly, 1986) which required that Iowa administrators receive approval to evaluate personnel. I-LEAD became the instrument through which that evaluator approval training was delivered (Dowdle, 1991). As a part of that work, in SB 2175, the Iowa Department of Education (DOE) was required to establish "competencies" that evaluators should possess upon completion of the evaluator training.

The "new" evaluator approval legislation, as previously described in Iowa Code 284.10, is the third incarnation of these efforts to improve instruction through improved teacher supervision and evaluation in less than 30 years. As has been noted, this legislation indicates "the department shall establish an evaluator training program to improve the skills of school district evaluators in making employment decisions, making recommendations for

licensure, and moving teachers through a career path as established under this chapter" (284.10 paragraph 1).

This directive is further clarified in an Iowa Department of Education document listing the types of skills to be developed in evaluators as a result of the training. Those skills include the following: support for teaching standards, data collection skills, feedback techniques, and leadership in data driven decision-making. The training will also help those who have been trained to document the development of teaching skills as defined by the Iowa Teaching Standards and Criteria, support the locally developed evaluation process, and provide support for the on-going individual development of beginning and career teachers.

Under the new IEATP, certified evaluators will know and be able to do the following (Berger, 2002):

1. Understand the theory behind best practice teacher evaluation systems (and quality professional development models).
2. Demonstrate the ability to provide data-based leadership.
3. Demonstrate the ability to manage data relating to teacher performance.
4. Identify quality instruction in the classroom.
5. Validate quality teaching based on the Iowa Teaching Standards.
6. Provide coaching/feedback in a professional growth environment.

These competencies for evaluators are different in that there is a focus on data in decision-making, professional growth in the form of individual career development plans, and a common language for expectations for teacher behaviors in the form of the Iowa Teaching Standards and Criteria. Unfortunately, no study has been conducted to see if the

teacher evaluation and evaluator training reform has been successful in changing teacher instructional practices or perceptions based on the perspective of the novice teacher.

Statement of the Problem

Effects of implementing the IEATP need to be examined from the perspective of the novice teachers who completed their second year of teaching in 2002-03. This is the first group of teachers that will have had their permanent licensure granted contingent upon the new evaluation requirements. The goal of the Iowa Teacher Quality legislation was to improve student achievement. The belief was that improved teacher quality would lead to student achievement gains, and that improved teacher quality would occur through a change in evaluator behaviors and knowledge and in the evaluation processes. Thus it is important to assess whether substantive changes have occurred in perceptions of evaluator behaviors and knowledge and in the evaluation process at the building level. This can be measured by comparing perceptions of the evaluation methods in the 2001-02 school year, before the IEATP, and in 2002-03, after the implementation of the IEATP. While this study will not determine directly whether the legislation has lead to student achievement gains, it provides a measure of a seminal step in the process. Clearly, if no change in evaluation has occurred it is unlikely that any change in student achievement over this time could be attributed to the teacher evaluation requirements.

Purpose of the Study

The purpose of this study was to investigate the quality and effectiveness of the newly mandated teacher evaluation process in Iowa from the vantage point of novice teachers, and to determine their perceptions of the attributes of teachers, their supervisor, the procedures of

evaluation, and the feedback. It also provided information on the differences that may exist in perceptions of novice teachers based on their own gender and/or the gender of their evaluator. The study may provide a model for the evaluation of reform programs such as this one, by focusing measures of success on variables such as teacher perceptual and behavior change in addition to comparing student test results in isolation. Additionally, the information from the study may impact the training programs for evaluators as well as the framework for the work being done across the state with mentoring and induction programming. The study may also impact thinking on the manner in which training occurs related to the gender of the evaluator and/or the gender of the novice teacher.

This information will be shared with members of the Iowa Department of Education, School Administrators of Iowa, university faculty and all trainers and evaluators of the IEATP, to improve the implementation of the newly created and mandated teacher evaluation process in Iowa.

Research Questions

The following research questions guided this study:

1. What traits do novice teachers report concerning their perceptions of self as a teacher, perceptions of his/her evaluator, perceptions of the attributes of the evaluation processes, perceptions of the attributes of the feedback, and perceptions of the context in which the evaluations occurred?
2. Do factors such as gender of the teacher, gender of the evaluator, or gender interactions between the teacher and the evaluator impact perceptions about teacher evaluation?

3. Are there differences in perceiver data from the 2001-02 school year to the 2002-03 school year (before and after the IEATP) around attributes of self as a teacher, perceptions of his/her evaluator, perceptions of the attributes of the evaluation processes, perceptions of the attributes of the feedback, and perceptions of the context in which the evaluations occurred?

Assumptions of the Study

Several assumptions were made regarding the study:

1. If evaluators changed their evaluation methods teachers would notice it and report it accurately on the survey.
2. Teacher perception of evaluator change is an important precursor to teacher behavioral changes in instruction.
3. Teacher perceptions and behaviors change with additional experience in the classroom.
4. The respondents were honest in completing the surveys and in any potential follow-up contacts that were made.

Delimitations

Delimitations of this study included the following:

1. The respondents were limited to novice teachers, in an Iowa public school.
2. The evaluators will have met the expectation that the IEATP will have been completed.
3. The state did not require standardized forms for evaluation, so changes in perceptions are not related to variances in district forms or processes.

4. The respondents were limited to teachers who completed their second year of teaching in the 2002-03 school year, and were evaluated by the same person in both of the school years 2001-02 and 2002-03.
5. The respondents were limited to teachers who successfully transitioned their initial teaching license to a career teacher license.
6. The survey instrument is only measuring the perception of teachers and not the actual behavioral changes of principals or teachers.
7. The study was conducted over a two-year comparison period. Single year anomalies will still be possible errors, where multiple pre-IEATP years and multiple post years would be more powerful. Unfortunately, the program had only existed for one year at the time of the study, so this approach was not possible.
8. Grade levels were mixed in the study, as opposed to isolating for specific grade groupings such as elementary, middle school/junior high and high school teacher perceptions. This occurred largely due to potential small sample size for given grade level segments when combined with the gender of the teacher and the evaluator. For example, male teachers in the elementary level were paired with either gender of principal.
9. School and district sizes, based on the number of students attending, were of mixed sizes in the results. This same mixing occurred for areas in the state from which the results were derived, such as urban, suburban, and rural. Both delimitations were impacted by potential sample size concerns, as well.

10. The ages of the beginning teachers as well as the ages and experience levels of the evaluators were mixed in the results. Originally, attempt was made to gather data on experience level of evaluators, but those data were lost in the data transfer.
11. Controlling for the subject area(s) in which the teacher is certified and teaching was also a delimitation. For example, it is possible that there were perception differences among teaching areas based on perceptions of evaluator expertise in the content.

Definition of Terms

The following working definitions were provided for this study:

Beginning Teacher: An individual serving under an initial license, issued by the Board of Educational Examiners under Iowa Code Chapter 272, who is assuming a position as a classroom teacher, or as a preschool teacher.

Comprehensive Evaluation: A summative evaluation of a beginning teacher conducted by an evaluator for the purposes of determining a beginning teacher's level of competency relative to the Iowa teaching standards and for recommendation for licensure and to determine whether the teacher's practice meets the school district's expectations for a career teacher.

Formal Evaluation: Observations that were pre-announced, and were preceded and followed by a conference with the evaluator.

Formative Evaluation: Evaluation that focuses on teacher growth. This portion of evaluation is primarily focused on the specific behaviors of the teacher in the classroom. It is typically not concerned with the out of class or professional activities of the teacher.

Informal Evaluation: Unannounced drop-in visits that were accompanied by some type of feedback (i.e., written and/or oral).

Initial License: An initial license is valid for two years, and may be issued to an applicant who has a baccalaureate degree from a regionally-accredited institution, has human relations component; has completed the exceptional learner component; has completed the requirements for one of the basic teaching endorsements, and meets the regency requirement of 14.115 "3" (see 282-14.110).

Iowa Evaluator Approval Training Program: Within the State of Iowa, an evaluator training program designed to improve the skills of school district evaluators in making employment decisions, making recommendations for licensure, and moving teachers through a career path as established under this chapter (Iowa Code 284.10).

Probationary Period: The first three consecutive years of employment of a teacher in the same school district are a probationary period. However, if the teacher has successfully completed a probationary period of employment for another school district located in Iowa, the probationary period in the current district shall not exceed one year. A board of directors may waive the probationary period for any teacher who previously has served a probationary period in another school district and the board may extend the probationary period for an additional year with the consent of the teacher [IA Code 279.19].

Standard License: A standard license is valid for five years and may be issued to an applicant who completes the requirements for an initial license; shows evidence of successful completion of a state-approved induction program or two years' successful teaching experience based on a local evaluation process; and meets the regency requirement of 14.115 "3" (see 282-14.110).

Summative Evaluation: It is the evaluation that typically occurs at the end of an evaluation cycle. This form of evaluation takes into account both the in class and out of class factors which impact the responsibilities of the teacher.

Summary

Chapter 1 centered on the concept of accountability in public schools and the impact it has had on expectations for teacher evaluation in Iowa. This study focused on the traits of novice teachers and the possible changes in teacher perceptions based on implementation of the tenets of the IEATP. The next chapter is designed to define leadership in general as well as the role of principals as it relates to teacher evaluation. There will be an emphasis on novice teachers and gender interactions in teacher evaluation models.

CHAPTER 2. REVIEW OF LITERATURE

Perspectives of Leadership in Education

The history of leadership in education includes a wide variety of leadership styles and expectations. Bartky (1956) defined an educational leader as a person who influences people, is democratic, who derives his power from the majority, is loyal to his organization, and above all, educates. Achilles (1988) argues that the complete administrator knows what to do (administration as a science), how to do it (administration as a craft), and most important of all, when to do it (administration as an art).

In a summary of their study on educational leadership, Mazzarella and Grundy (1989) noted the following:

> According to this [assembled] research, typical educational leaders are a little more intelligent (but not too much more) than nonleaders. As children they were probably not firstborn and were probably allowed at an early age to make many of their own decisions… Effective educational leaders are outgoing, good at working with people, and have good communication abilities and skills. They take initiative, are aware of their goals, and feel secure. As proactive people, they are not afraid to stretch the rules, but also understand the compromises that must be made to get things done. (pp. 316-317)

More recently, definitions of educational leadership have included a variety of models: managerial leadership, political leadership, contingent leadership, participative leadership, moral leadership, constructivist leadership, facilitative leadership, instructional leadership, and transformational leadership. Some of these leadership types can be more readily generalized to all educational leaders, while others are more specific to the principalship.

Contingent leadership is designed to match the leader behaviors to the organizational context (Leithwood, Jantzi, & Steinbach, 1999). As a result, the leader must be regarded as an excellent problem-solver who is focused on responding to the challenges of the organization. In this leadership model, there is an emphasis on both achieving the formal goals of the organization and increasing the capacity of the organization to respond productively to internal and external demands for change.

Participative leadership is a form of leadership that emphasizes the group and is grounded in interpersonal communication (Leithwood et al., 1999). The purpose of this type of leadership is to increase the participation of all members in the group in the decision-making process. The intent is to make for a more democratic organization with an increased capacity to respond productively to internal and external demands for change.

Moral leadership has been described as leadership that believes in an expectation that a difference can be made in the lives of students, and there is an emphasis on a can-do spirit (Fullan, 2002; Johnson, 2002). In addition, this leadership style is participative with the formation of a leadership team, an emphasis is placed on forming a professional learning community and building leadership capacity, and there is an ongoing reference back to the mission of the school in the decision-making process (Glickman, 2002a; Lambert et. al., 2002; Uchiyama & Wolf, 2002). These leaders have a belief that long-term, sustained change comes through the involvement of many and that the group will move towards doing what is best to sustain clearly defined reform initiatives.

Constructivist leadership draws from the influences of Dewey, Piaget, Gardner, Marzano, and others. This type of leadership is viewed as a reciprocal process among the adults in the school (Lambert et. al., 2002). Purposes and goals develop from among the

participants, based upon values, beliefs, and individual and shared experiences. The school functions as a community that views the growth of its members as fundamental. There is an emphasis on language as a means for shaping the school culture, conveying commonality of experience, and articulating a joint vision. Shared inquiry is an important activity in problem identification and resolution.

Blase and Blase (1997) provide the principal strategies that influence teachers' sense of empowerment as it relates to facilitative leadership. They include: demonstrating trust in teachers; developing shared governance structures; encouraging/listening to individual input; encouraging individual teacher autonomy; encouraging innovation, creativity, risk-taking; giving rewards; and providing support. Related to those strategies are five personal characteristics of principals – caring, enthusiasm, optimism, honesty, and friendliness – that also contribute significantly to teachers' sense of empowerment. Additional findings from their research indicate that facilitative leadership is, in part, based on equitable (fair) exchanges between principals and teachers. They also found that facilitative leadership and teacher empowerment are based on value congruence – specifically, the strategies and characteristics that have already been described and viewing the principals' general goal of improving teaching and learning.

In a study of principals' perspectives on facilitative leadership, Blase and Blase (1999b) found that exemplary shared governance principals lacked the dominant presence of traditional principals precisely because they have rejected a preoccupation with self, the imposition of their personal vision on other, and the traditional authoritarian quest for power over others. They also found that principals should reflect on their readiness to enact a dramatically different leadership role of facilitative leadership.

Suffice it to say, there are a number of views on leadership and how leadership is defined. There are other types of leadership defined beyond those addressed in this document. Perhaps most importantly, the definitions appear to focus on the traits of leaders and the process or processes for interaction between the leader and the people with whom he or she is responsible to lead. These behaviors and interactions have an impact both on the leaders and on the perceptions of those for whom they are responsible to lead.

Leadership Role of the Building Principal

Leadership and approaches to leadership can be applied to the building principalship. The role of the building principal as a leader continues to evolve over the years and even decades. Bartky (1956) noted that the principal has three leadership roles (a) a leader of children, (b) a community and parent leader, and (c) a leader of teachers.

The predominant role enacted by principals, from the 1920s until the 1960s, was one of administrative manager (Hallinger, 1992). This leadership style focuses on the functions, tasks, or behaviors of the leader and assumes that if these functions are carried out competently the work of the others in the organization will be facilitated (Leithwood & Duke, 1999). While this approach is effective in managing the day-to-day activities of a building, there is little emphasis on building instruction or reform within the building. Managerial leadership is sometimes referred to as "crowd control" and "instructional order" (Cuban, 1988).

Cuban (1988) also described the political leadership role of the building principal. The principal is in a position of being forced to respond to pressures from a growing number of sources: parents, students, teachers, superintendents, and local and national policy makers.

How the principal responds to these groups – exercising moral judgment, for example – may well shape the manner in which education proceeds, or doesn't, in a democratic society.

During the 1960s and 1970s, a new role emphasis emerged for American principals as they became increasingly responsible for managing federally-sponsored and funded programs designed to serve special student populations (Hallinger, 1992). During these decades, principals assumed a new set of change implementation functions that ranged from monitoring compliance with federal regulations, to helping with staff development, to providing direct classroom support for teachers. While this was reform, in a sense, the changes generally came about due to externally devised solutions to social or educational problems. Cuban (1988) also found that many principals demonstrated more concern for meeting criteria for compliance than for program outcomes.

In 1979, Ron Edmunds published a seminal article in which he stated that strong administrative leadership was a characteristic of instructionally effective schools. By the mid-80s, this document led to a new form of educational leadership referred to as instructional leadership.

> The instructional leader was viewed as the primary source of knowledge for the development of the school's educational program. The principal was expected to be knowledgeable about curriculum and instruction and able to intervene directly with teachers in making instructional improvements. High expectations for students and teachers, close supervision of classroom instruction, coordination of the school's curriculum, and close monitoring of student progress became synonymous with the role definition of an instructional leader. (Hallinger, 1992)

Rallis and Highsmith (1986) suggested that school management and instructional leadership are two different tasks. They question whether one person can perform both tasks simultaneously and well. Instructional leadership has been broadly interpreted to encompass

those actions that a principal takes, or delegates to others, to promote growth in student learning (De Bevoise, 1984). Highlights of this work on the analysis of the research on the principal as instructional leader include the following:

1. Principals cannot exercise instructional leadership in a vacuum. They need support from teachers, students, parents, and the community.
2. Common leadership functions that must be fulfilled in all schools including communicating the purpose of the school, monitoring performance, rewarding good work, and providing staff development. Whether or not these functions must be carried out by the principal depends upon the make-up of the teaching staff and the organization of the school district.
3. While previous studies have generally concentrated on only one facet of instructional leadership – such as personal traits, leadership styles, management behaviors, or organizational contexts – current studies tend to address the interrelationships between these factors.
4. The personal characteristics of the principal cannot be ignored when studying what constitutes effective instructional leadership. (p. 18)

In a synthesis of studies, Smylie and Hart (1999) noted that there are two consistent patterns when studying the research on teacher collegial relations, collaboration, and professional community as it relates to the principalship. First, principals have substantial influence on the development, nature, and function of teacher social relations, teacher learning, and change. Second, principal leadership is strong and purposeful in schools with high levels of collaboration and strong professional communities.

Blase and Blase (1999a) conducted a study in which they asked teachers what they believed to be the principal behaviors that lead to the most effective instructional leadership. Their findings produced two major themes as being most effective (a) talking with teachers to promote reflection and (b) promoting professional growth. Three sub-themes that ran through the study included teacher choice and discretion in relation to career path, non-threatening interactions that were growth oriented, supportive, positive, and built on mutual

respect and trust; and authentic interest of the part of the principal, reflecting true caring and interest in the teachers' professional growth.

Instructional leadership is an idea that has served many schools well throughout the 1980s and early 1990s. However, in light of current restructuring initiatives designed to take schools into the 21st century, instructional leadership no longer appears to capture the heart of what school administrators will have to become. Rather than instructional leadership, transformational leadership evokes a more appropriate range of practice; it ought to subsume instructional leadership as the dominant leadership, at least during the 1990s (Leithwood, 1992).

At the same time Burns was formulating transformational leadership, he also spoke of transactional leadership (Burns, 1979). Leithwood (1992) contrasted transactional leadership – leadership based on the exchange of services for rewards the leader controls, in part, at least; with transformational leadership – which provides the incentives for people to attempt the improvements in their practices. Sergiovani (1990) argued that transactional leadership is a first stage in transformational leadership and central to getting day-to-day routines carried out.

Transformational school leaders are in more or less continuous pursuit of three fundamental goals (Leithwood, 1992): (a) helping staff members develop and maintain a collaborative, professional school culture, (b) fostering teacher development, and (c) helping teachers solve problems together more effectively. Sergiovani (1990) suggested that student achievement can be improved remarkably by transformational leadership.

Transformational leadership assumes that the central focus of leadership ought to be the commitments and capacities of organizational members (Leithwood, Jantzi, & Steinbach,

1999). Burns (1979) noted that the ultimate test of practical leadership is the realization of intended, real change that meets people's enduring needs. These belief systems begin to signal the shift in the principalship from the focus on the leader to a focus of the leader on the participation of constituents in the process of leading the building. The complex issue here is for the principal to empower the teachers through team leadership without abdicating the principal's authority (Glatthorn, 1996).

Leithwood (1994) used the following dimensions to define transformational leadership behaviors:

1. The leader identifies and articulates a vision (often collaboratively) developing, articulating, and inspiring others with a vision of the future.
2. The leader conveys high-performance expectations – demonstrating the leader's expectations for excellence, quality, and/or high performance on the part of the staff.
3. The leader provides appropriate models – behavior on the part of the leader sets an example for staff to follow and is consistent with the values espoused by the leader.
4. The leader provides intellectual stimulation – challenging staff to reexamine some of the assumptions about their work and to rethink how it can be performed.
5. The leader provides individualized support – behavior indicating respect for individual members of the staff and concern about their personal feelings and needs.

Completing studies against these dimensions, Leithwood (1994) concluded that transformational approaches to school leadership are especially appropriate to the challenges facing schools – especially the challenges of restructuring. He also found that accomplishments of transformational leadership in schools depends on attention to all its facets, that transformational leadership is value added, that the implementation of the practices themselves are contingent on the context and vary widely, that expert thinking lies behind effective leadership practices, and distinctions between management and leadership cannot be made in terms of overt behavior.

Following thinking similar to that of Leithwood, in 1996 Glatthorn described what he calls learning-centered leadership. In a summary of research, he described a number of principal behaviors as vital in executing the role of learning-centered leadership. Those behaviors considered to be connected to this study are:

1. Discerns and articulates a vision of excellence and enables others to share effectively in the visioning process.
2. Is aware of the moral dimensions of schooling, acts ethically, and sensitizes teachers and students to the moral aspects of teaching and learning.
3. Maintains a focus on curriculum and instruction, by informing teachers of current developments, sharing knowledge gained through experience, observing teachers at work, monitoring the implementation of the curriculum, and rewarding effective teaching.
4. Uses routine activities and informal interactions as means of reinforcing the teachers' commitment to learning.
5. Creates a climate of high expectations for all, while providing the support needed to realize such expectations: assists teachers in acquiring new skills, supports teachers in disciplinary matters, builds a school curriculum that is developmentally appropriate.
6. Uses a problem-solving approach in fostering continuing improvement: evaluates formatively and summatively, investigates problems, uses reflection to gain insight into problems, uses structures that enable others to become involved in the problem-solving process.
7. Provides services and structures to entire faculty and to individuals that foster their professional growth.
8. Communicates effectively, giving earned and timely praise to all who merit it.

Not everyone is sold on the concept of transformational leadership. Clabaugh (2001) indicated that the literature on transformational leadership is replete with solemn assurances that a visionary change agent, who is an expert at dealing with complexity and ambiguity, can successfully convince everyone to serve goodness, righteousness, duty, and obligations. He argues that this is impossible to achieve due to the fact that there is such a wide variety in constituents, and we as individuals, want different things from schools based on circumstances. When these various role differences are played simultaneously, the balancing

act becomes impossible and the transformational leaders will never be good enough to be considered effective.

Allix (2000), through the use of a coherentist epistemology, showed that the claims of Burns cannot be satisfactorily sustained as they relate to transformational leadership. He noted that Burns' suppositions are based on emotive preferences rather than on empirical content. He also argued that because the conception is focused on persuasion and influence, that the sheer number of people in a social structure will determine the outcome of value conflicts rather than moving toward the common good.

A study by Day, Harris, and Hadfield (2001) suggested that we are entering an era of post-transformational leadership. It is their view that the failure of existing leadership theory to capture, explain, and represent current leadership practice lies in a reluctance to acknowledge that leadership is a complex, messy, and, at times, wholly non-rational activity that is value-laden and value-driven. Their findings indicate that morality, emotion, and social bonds provide far more powerful stimulants to motivation and commitment than the extrinsic concerns of other styles of leadership.

Most recently, instructional leaders are those who lead cultural change (Fullan, 2002). Williams-Boyd (2002), provided what she considered a definition for educational leadership for the 21st century:

> Leadership, then, is a collaborative process of engaging the community in creating equitable possibilities for children and their families that result in academic achievement. The new notion of leadership dreams of changing the world rather than maintaining it. It celebrates embrace of others, championing differences, and nurtures young minds toward seeking questions and posing solutions. It is a notion of stewardship that seeks a higher moral purpose and a more communal humanity born of shared vision and common purpose. (pp. 5-6)

The manner in which the principal leads and the belief systems he or she brings to the role have an impact on the building and the activities that occur in schools. In addition, the teachers in the building have a significant impact on student learning. The manner in which they respond to leadership can be impacted through the teacher career stage in which they are currently engaged. While understanding leadership and the role of the building principal is important for this study, so is an understanding of the career stages of teachers. This is especially true relative to the understanding of the perceptions of beginning or novice teachers.

Teacher Career Stages

Just as there are differences in principals and the way in which they lead, there are differences in teachers and the manner in which they respond to feedback as well as what they need in a model for supervision (Armato, 1990; Glatthorn 1996; Glickman, 2002a; McGreal, 1983). These differences in need are often associated with differences in teacher career stages (Glatthorn, 1996; Glickman, 2002a; Steffy et al., 2000).

The stages of a teaching career have been widely discussed in the educational literature, and have an impact on how teachers perceive and respond to principal leadership and ultimately, teacher evaluation. However, the number of stages in a teaching career has not been described in consensus. It is generally regarded in the literature that the careers of teachers can be broken into two stages (Ryan, 1979), three stages (Armato, 1990; Burden, 1982; Newman, 1978; Unruh & Turner, 1970), four stages (Gregoric, 1973; McDonald, 1982; Petersen, 1978), five stages (DeMoulin & Guyton, 1987; Jellinek, 1985), six stages (Casey, 1994), seven stages (Huberman, 1989; Vonk, 1989) and eight stages (Fessler &

Christensen, 1992). Steffy (1987) and Steffy et al. (2000) championed a multi-layered approach in which the stage of the teacher varied according to the context of the teaching situation.

These career stages are sometimes specifically defined by the years of experience in which the instructor is engaged (Burden, 1982). More often, they are generally defined around years of experience – year one, years two and three, or years one to three, but more importantly by the descriptors of the stages – initial teaching stage, the period of building and security, and the fully functioning stage (Unruh & Turner, 1970), survival, consolidation, renewal, and maturity (Katz, 1972), or investigation, acculturation, opportunities, respect and recognition, rejuvenation, and retirement (Eberhart, 1990).

Career stages have also been described by some researchers as the skills that the instructor has acquired through the career, rather than the actual number of years of teaching experience (Glatthorn, 1996; Glickman, 2002a; Steffy et al., 2000). These same studies, among others (Fessler & Christensen, 1992; Huberman, 1989) have also moved the conversation toward a cycle rather than a linear process of career stages through which teachers progress from the beginning of a teaching career to the end of the career.

While a study of career stages in general is important, for this work an emphasis was placed on the beginning portion of the teaching career. This focus needed to include information related to the thoughts, beliefs, and behaviors most commonly associated with teachers who are new to the profession.

Early in teaching careers there is a focus on student discipline, development of skills, and the context of skill implementation, to name a few. Armato's (1990) study revealed that experienced instructors are more creative and less curriculum bound, are active on

committees, their interactions with students are more humanistic, and they have a more broad repertoire of instructional skills and activities. There is also a focus on concerns for themselves and then concerns on teaching tasks and impacts they are having on students (Fuller, 1969). An additional struggle is understanding the relationship with the building administrator due to three categories described as the multiple roles of the principal (especially that of judge), different perspectives on what is occurring in the classroom, and dealing with administrative authority (Ryan, 1986).

Based on the work at the Collegial Research Consortium, Burke et al. (1987) perceived that the management style of the principal is the most important factor in creating or preventing teacher movement among stages. If a principal's management style matched a teacher's level of competence and commitment, the teacher is likely to thrive professionally. If not, the result would be detrimental.

Characteristics of teachers are often associated with studies on teacher career stages. Casey (1994) delineated six stages in the teaching career. More importantly perhaps was a finding from his research demonstrating that all respondents, regardless of career stage, showed a high resistance to change. In addition, earlier in their teaching career the teachers expressed the least concerns about the profession and were the second highest scoring group in terms of enthusiasm about the teaching profession. Backus (1989) also found that the teachers with the most positive attitudes were more likely to be teachers of elementary-age students. He also discovered this positive attitude was most common in both the youngest and the oldest, most experienced teachers. Casey (1994) noted that beginning teachers scored the lowest of all stages in the area of involvement in professional development.

In a further focus on the beginning teacher, Peterson (2000) noted that once hired, new teachers require special care. They have predictably characteristic needs for reassurance, professional development, and sociological induction. Professional growth must first be familiarity with the district, second it must address gaps in teacher training, and third it must include classroom discipline. Finally, without sociological development, new teachers soon lose their idealism and return to the most conservative educational practices because their "personal social norms and sentiments profoundly and predictably change to a school teacher's perspective" (p. 106).

While Peterson's (2000) study focused on beginning teachers in their first two years of teaching, Moir (1999) identified five phases of the first year of teaching: anticipation, survival, disillusionment, rejuvenation and reflection. These phases are not as pronounced by the second year of teaching and may have an impact on the results of the survey for years one and two.

This summary of work on teacher career reinforces the idea that there are distinct differences in teacher perceptions and behaviors as they move through a teaching career. The interaction of teachers and evaluators come to an intersection against the backdrop of instructional supervision.

Instructional Supervision

Kemmerer-Fehr (2001) conducted a historical inquiry/historical dissertation research on the role of the educational supervisor in the United States from 1970 to 2000. Her findings indicated that, from 1970 to the mid-1980s, the view of teachers remained fairly constant with supervision being viewed as one of centralized management, and the role of the

educational supervisor was that of a case manger in charge of changing unwanted teacher behavior. From the mid-1980s, supervision in the public schools has been one of increasing decentralized management where teachers and supervisors worked collaboratively in shared leadership communities toward shared goals. The study concluded that the roles of the supervisor – teacher evaluators, curriculum specialists, human relations specialists, change agents, and problem solvers – did NOT change dramatically from 1970 – 2000; however, what changed more significantly was the "what" and "how" supervisors carried out the role within the larger context of the educational setting. More specifically, since 1985 supervision has become more diverse and individualized, is focused more on teacher thinking, democratic, and participatory.

The following information sheds light on the variety of approaches and philosophies that exist related to instructional supervision. Bartky (1956) noted that there are six theories of supervision: authoritarian, representative, co-operative-democratic, invitational, scientific, and creative. He championed the idea that each theory has merits based on the principal's personality, teacher attitudes, and context in which the evaluations are occurring.

Franseth (1961) indicated that supervision is generally seen as leadership that encourages a continuous involvement of all school personnel in a cooperative attempt to achieve the most effective school program.

In 1974 the National Education Association (NEA) provided a position statement on teacher evaluation. According to that document, the three specific dimensions of an evaluation process are:

1. Program – ranging from economic resources and fiscal effort of the local school district to student characteristics and their readiness to learn;

2. Performance – including knowledge of subject matter, teaching-learning strategies, plus such adjunct activities as planning, evaluation, and community relations; and,
3. Learning Outcomes – involved the difficult matter of translating stated learning objectives into observable phenomena such as knowledge, behavior, attitude, skill, etc. Plus attention to the fact that the education of students is increasingly shared by agencies beyond the school. (p. 22)

The report also noted that evaluation should be based on multiple indices and should involve a wide variety of personnel in the process.

Based on his research and experiences, McGreal (1983) established nine commonalities of effective teacher evaluation systems that he viewed as best practices: (a) an attitude of accountability centered on the expectation of instructional improvement; (b) complementary procedures, processes, and instrumentation based on a relationship of trust between the supervisor and the teacher; (c) a separation of administrative and supervisory behavior; (d) goal setting; (e) a narrowed focus on teaching based on a common framework; (f) use of a modified clinical supervision format by narrowing the range of things observed in the observation; (g) use of alternative sources of data; (h) differentiated requirements for tenured and non-tenured teachers; and (i) a complete training program for supervisors and teachers. Conley (1987) arrived at very similar results in his research on effective evaluation systems. His findings placed an additional emphasis on formative and summative evaluation, and stressed the importance of training for all staff members using the district evaluation model.

Glatthorn (1984) described supervision as a process of facilitating the professional growth of a teacher, primarily by giving the teacher feedback about classroom interactions, and helping the teacher make use of that feedback to make teaching more effective. Based on

that definition and his research, he called for differentiated supervision for teachers. The rationale he provided for differentiated supervisions was threefold:

1. The standard supervisory practice of administrators and supervisors is often both inadequate and ineffective.
2. It is neither feasible nor necessary to provide clinical supervision to all teachers.
3. Teachers have different growth needs and learning styles

Glatthorn went on to provide an overview of a differentiated system; providing teachers with four types of supervision:

1. Clinical supervision is an intensive process designed to improve instruction by conferring with a teacher on lesson planning, observing the lesson, analyzing the observational data, and giving the teacher feedback about the observation.
2. Cooperative professional development is a collegial process in which a small group of teachers agree to work together for their own professional growth.
3. Self-directed development enables the individual teacher to work independently on professional growth concerns.
4. Administrative monitoring, as the term implies, is a process by which an administrator monitors the work of the staff, making brief and unannounced visits simply to ensure that the staff are carrying out assignments and responsibilities in a professional manner.

Duffy (1985) noted there is no evidence that instructional supervision is effective and no data to suggest that supervision makes a difference. He goes on to say that although instructional supervision has not been shown to be effective, he passionately believes that it can make a difference in schools. He describes four basic premises must guide school efforts toward the goal of increasing the effectiveness of instructional supervision:

1. Premise A – it must be recognized that each school district and, indeed, each school within a district has unique organizational characteristics which may either constrain or enhance the effectiveness of supervision.
2. Premise B – there must be a change not only on individual supervisors' attitudes, concepts, and skills, but also in the organizational structure of the school.
3. Premise C – during the process of increasing the effectiveness of supervision, a comprehensive organizational perspective must be taken.

4. Premise D – to maintain the effectiveness of supervision, problem-solving capacity must be built into the school organization. (pp. 1-2)

Based on their summary of research, Duke and Stiggins (1986) concluded that the five keys to success in teacher evaluation are: the integration of the teacher, the evaluator, the performance data, the feedback, and the context in which the evaluations occur.

Buttram and Wilson (1987) indicated that throughout the 1980s school districts invested considerable time and effort into evaluating student progress, and relatively little in monitoring the teaching process. They continued that progressive districts are taking a new look at teacher evaluation. The improved instructional practices include: linking evaluation systems to research on effective teaching practices, providing improved training for evaluators, holding administrators more accountable for conducting evaluations, using evaluation-identified teacher deficiencies to focus staff development, and making teachers active partners in the evaluation process. In a review of the improved evaluator training they found that while the programs varied significantly in their focus and intensity, they usually include a review of the system's focus, content, and procedures, and some supervised practice with a videotape or live classroom observations.

Poplin (1992) indicated that instructional leadership concentrated on growth of students and rarely looked to growth of teachers. Today's scholarship tells us that to promote true growth in any individual, we must be conscious of what drives us to become the best we can be.

Fink and Resnick (2001) observed in working with principals and others who serve as instructional leaders that their emphasis is on leadership rather than content expertise. Instructional leaders must know enough about pedagogical practice, content-specific

knowledge, and critical curriculum issues to understand what it is they are observing, but their primary focus is on determining ways they can provide instructional assistance or professional development for their teachers.

Glatthorn (1996) described four general groups of teachers as they relate to instructional supervision:

1. Novices – teachers at the first stage of career development, usually high in motivation and low in expertise, functioning at the basic level.
2. Marginal – teachers at the second and third stages in career development, low in motivation and still struggling to master several of the fundamental skills of teaching.
3. Passive – those who have lost their motivation to teach; they are passive in their attitudes toward school improvement, in their approach to teaching, and in the kind of learning they provide. Includes those at advanced levels of career development, who have mastered the basic skills, but who have lost their motivation to teach and are not interested in moving to a higher level of skill development.
4. Productive – those who are competent and continue to grow; the group includes teachers at the intermediate level of skill development and expert teachers functioning at an advanced level. They are high in both motivation and competence, the core of each productive faculty. (p. 13)

It is clear that there are essential elements to instructional supervision based on a summary of these findings. Those elements, among many, include well-trained evaluators, differentiated approaches to evaluation based on the career stage and experiences of the teacher, an emphasis on teacher professional growth, feedback from a variety of sources, a teaching staff and district committed to the evaluation process, and an acknowledgement that the circumstances of the building and district have an impact on the evaluation model for the district. The framework for instructional supervision has a direct impact on the formal work of teacher evaluation.

Teacher evaluation

There is an interaction between the principal and teacher around the concept of instructional supervision. Clearly, the attitudes and beliefs that each bring to the equation have an impact on the results of evaluation process. In addition to the people, the underlying belief systems and processes for evaluation play a part in the outcome of the evaluation. There are two broad, sometimes seemingly dichotomous, roles of teacher evaluation – accountability for the instructor and professional growth of the instructor (McGreal, 1983; Sergiovani, 1990; Stiggins & Duke, 1988).

Stiggins and Duke (1988) designed a Teacher Evaluation Profile (TEP) survey focused on attributes of the teacher, of the evaluator from the perspective of the teacher, information that was gathered during evaluation, the feedback the teacher received, and the context of the evaluation.

Their research revealed that at least five factors that contribute to effective teacher evaluation. One of those factors is the attributes of the teacher. The attributes of teachers that were most likely to produce positive results from evaluation were instructional competence, personal expectations, openness to suggestions, orientation to change, knowledge of subject matter, and experience. In regard to experience, teachers with a history of useful evaluations were more likely to benefit from future evaluations than teachers for whom the process has been uninspiring and uninformative.

A second factor influencing the effectiveness of teacher evaluations is the attributes of the evaluator. The evaluator traits most likely to affect the quality of the teacher evaluation experiences are: credibility, persuasiveness, patience, trust, track record, and modeling.

A third factor influencing the effectiveness of teacher evaluations is the attributes of the procedures used to gather data on teacher performance. Specifically, these procedures are well-defined performance criteria and standards, and well-defined and varied data collection procedures.

A fourth factor influencing the effectiveness of teacher evaluations is the attributes of the feedback. The most useful feedback is from a credible source(s), describes specific aspects of teaching along with suggestions for improvement, comes with frequent regularity to measure ongoing growth, and is formal as well as informal.

The fifth factor influencing the effectiveness of teacher evaluations is the attributes of the evaluation context. The history of labor relations in the district, time spent on evaluation, and resources available for professional growth all have an impact on teacher evaluation.

In a study of attributes of teacher evaluation systems that promote teacher growth, from the perspective of teachers of Intense English Programs, Rindler (1994) used the TEP Survey instrument designed by Stiggins and Duke (1988). Rindler's findings indicated that the attributes of the evaluator (usefulness of suggestions and persuasiveness of rationale, credibility and level of trust of the evaluator, evaluator's capacity to model suggestions) and attributes of the feedback (focused on standards that are clear and endorsed by the teacher) are most significant, but all attributes play a part in effective teacher evaluation.

Using the same TEP Survey instrument, Lawler (1992) surveyed Iowa teachers following what was then the implementation of the state-mandated evaluator approval training that occurred as a part of I-LEAD. Laeler's results, which emerged from a sampling of veteran teachers across the state, demonstrated that the quality of teacher evaluation in each of the attribute clusters improved following the implementation of the evaluator

training. Of those, feedback appeared to have the greatest impact on evaluation. Evaluators did score low in frequency of informal and formal feedback, on depth of information, quality of ideas, and specificity of information. This work using the TEP speaks to elements of evaluation, but does not specify the exact evaluation processes used in any of the provided studies. That is to say, while the general framework for evaluation remains the same or similar, there are process, form, and expectation differences from one district to the next in Iowa.

Frameworks for instructional supervision

The act of teacher evaluation comes in many forms and has been defined across a number of contexts. McKenna (1973) proposed a teacher evaluation framework based on a program-performance-outcomes paradigm, and his was the precursory work upon which the NEA position statement was founded in 1973-74. He went on to argue that the process of the system should be considered as an end to itself – not everything can be directly observed and measured, but the actions are likely still worth undertaking and completing: therefore, should be promoted. He also indicated that the first implication of performance evaluation should be staff development and should involve a wide variety of people in the process.

The dominant model of instructional supervision, for decades, was that of clinical supervision. The clinical supervision model is generally regarded as containing the steps of (a) establishment of the teacher-supervisor relationship, (b) planning together by teacher and supervisor, (c) supervisor and teacher plan the objectives and steps for observation, (d) observation of instruction, (e) teacher and supervisor analyze the teaching data, (f) supervisor plans the conference, (g) conference occurs, and (f) renewed planning (Cogan, 1973). It is

clear that many of the elements to clinical supervision persist in some forms in the evaluation models found today.

Larry Barber (1985) proposed a Peer-Mediated Self-Appraisal (PMSA) System for teacher evaluation. In that system, he delineated between the learning needs and evaluation system of the beginning and new teachers (two years) and that of experienced teachers. For the beginning/new teachers, standards for performance are established, and the system is designed to be summative, process and product oriented, and a check on the quality of training and performance.

If a beginning/new teacher is successful after the first two years, they go to a one-year probationary status in which the evaluation is focused on process, and the products are used for employment decisions. If the standards are met, the instructor moves to the parallel systems that exist for veteran teachers. If the standards are not met, the teacher is placed on one year of intensive assistance. This portion of the system is focused on the summative nature of the product and processes; however, the primary purpose of this portion of the system is to provide direct training in perceived areas of deficiency. If this is successfully completed, then the move is back to a year of probation. If it is not, the individual is terminated. This model is similar to what has been described by Danielson and McGreal (2001), and is similar to a portion of the Iowa Model that includes Individual Teacher Career Development Plans (ITCDP).

Parallel supervision systems for experienced teachers are also described in Barber's PMSA Framework. The parallel systems are split into one set of activities in which there is a standard formative evaluation, which is driven by the teacher, is peer mediated, is process

oriented, is internally created and controlled by teachers, and is ongoing. The purpose of this portion of the evaluation is to improve individual teacher performance.

The standards professional review is the second set of activities delineated in the parallel system for experienced instructors. This system is focused on products, is designed to be summative, is externally created and controlled by supervisors, and is a spot check that occurs every three years. The purpose of this system is to check on quality teaching and to improve the system. If this review is successful, the teacher shifts back to the standard formative evaluation.

If the review is not successful, a remedial contract is developed that moves the instructor to the intensive assistance program. If the instructor is successfully trained in the area(s) of deficiency, he/she is moved to one-year probationary status. If that is successful, the instructor returns to the standard formative evaluation. If not, the individual is terminated.

Using his four groups of teachers as they relate to supervision, Glatthorn (1996) provides a typology of supervisory services summarized in Table 1. He argues that the foundation for successful teacher supervision lies in providing a supportive environment

Table 1. Supervisory services for teachers based on teacher group type

Type/ Factor	Career Stage	Level		
		Cognitive	Competence	Motivation
Novices	Career entry	Mixed	Working on basic skills	Generally high
Marginal	Stabilization, reassessment	Generally low	Have not mastered basic skills	Generally low
Passive	Reassessment, conservativism	Mixed	Mixed; most have mastered basic skills	Very low
Productive	Experimentation, self-acceptance	High	Intermediate or advanced	High

for all teachers. However, when looking at the various teacher types, the cognitive levels and competence levels are inversely proportional to the level of supervision required – those with high competence/cognitive level(s) require far less direct supervision than those with lower competence and/or cognitive levels.

Glickman (2002a) indicated that the behavior of instructional leader falls on a continuum as it relates to supervision of instruction. This continuum shifts from nondirective, to collaborative, to directive-information, to directive-control. The factors affecting the type of behavior the leader exhibits relates to the amount of responsibility for which the situation calls, from both the teacher and the instructional leader. These responsibility views are inversely proportional to one another. As the instructor shows more autonomy in instruction, the educational leader reduces the amount of directive approaches to supervision. The reverse can be true, as well.

Professional Development Plans (PDPs), a formalized structure for formative evaluation, designed to focus on the individual learning needs of each teacher and written by the teacher, are increasing in use and support (Holland & Adams, 2002). They advocate for the use of PDPs as the evaluation model for experienced teachers. They note that the use of PDPs is consistent with the helping function of supervision and its concentration on the teacher's growth in specific areas and they emphasize the shift of responsibility for evaluation from the supervisor to the instructor.

Their research indicates that through planning, implementing, and evaluating their PDPs, and in discussions with their administrative supervisors about the PDP experiences, teachers are encouraged and supported in professional inquiry involving reflection, applications of instructional innovation, professional collaboration, and research. They go on

to argue that because the PDPs shift the focus and design of evaluation and growth activities for both teachers and their supervisors, a more professional view of teaching and teachers' professional development is supported. These research findings also support the work that is being implemented in Iowa in the form of ITCDPs.

The work of schools and districts is to take the theory behind quality teacher supervision models and practices and produce meaningful evaluation systems for teachers. In many cases, this process for designing a district model has been guided not only by theoretical models and recommended practices, but also by expectations contained in state-mandated teacher evaluation models.

State-mandated teacher evaluation models

Guidance for district teaching models is often provided by states. School districts in every U.S. state are required by law to evaluate teachers. In each state, the written legislative purpose of the evaluations is to ensure that all American youth are taught by competent teachers (Frase & Streshly, 2000). Some of the states with state-mandated teacher evaluation models/programs, with the greatest longevity, have been in states such as Missouri, Tennessee, Texas, Mississippi, Connecticut, and Georgia.

In a comprehensive study of states with state-mandated teacher-evaluation models, Ellett (1987) noted that each model contained the following important elements:

1. "state-of-the-art" observation instruments that measure teaching behaviors that are well documented as importantly related to school outcomes by results of process/product studies and research on effective teaching;
2. comprehensive (three- to five-day) training programs which include one or more of the "proficiency' tests to certify instrument users;
3. standardized assessment procedures to be followed by all data collectors;

4. on-going programs of research and development to support the technical and psychometric characteristics of the observation instruments and assessment procedures; and most important,
5. training in the use of the observation instruments and assessment data for ongoing supervision and professional development of teachers. (p. 304)

For this study, the emphasis is on the importance of evaluator training using assessment data for supervision. More recently, states like Oklahoma, New Hampshire, and Iowa have joined the ranks of states with state-mandated models. The Iowa model differs from the elements recognized by Ellett (1987) in that no state-of-the-art observation instruments are required by the state (although model evaluation systems are made available through the Iowa Department of Education in cooperation with a number of state-recognized learning organizations such as School Administrators of Iowa, the Iowa State Education Association, and Iowa Association of School Boards), no standardized assessment procedures are in place (although there is agreement on definitions of data sources and data points and state samples are available), and no ongoing programs of research and development to support the technical and psychometric characteristics of the observation and assessment instruments.

Georgia was the first state to implement a performance-based teacher certification model for beginning teachers, beginning in the fall of 1980 (Ellett, 1987). In establishing the program, the Georgia Department of Education established three essential requirements that must be met to obtain a professional, renewable teaching certificate (a) an appropriate degree from an approved college or university teacher preparation program, (b) a "passing" score on a criterion-referenced test of knowledge in the certification filed, and, most uniquely, (c) acceptable on-the-job demonstration of "generic" teaching skills.

Using then current research on best practices in instruction, the Teacher Performance Assessment Instruments (TPAI) were designed by the University of Georgia and the structure was 14 generic teaching competencies with two to five performance indicators under each (for a total of forty-five performance indicators). The model is similar to what is observed in Iowa with the Iowa Teaching Standards (there are eight) and Criteria (there are forty-two). The Georgia model went on to identify performance indicators, scored from one to five, with a set of scoreable descriptors that further operationalize each indicator statement. The Iowa Department of Education has provided performance level descriptors, but their use is not required across the state.

The TPAI are administered by a team of trained observers. Each observer is certified as "proficient" in the use of the instruments by completing a comprehensive five-day training program and meeting a proficiency standard. The assessment team includes the building principal, a peer teacher (usually from the building in which the building teacher is initially employed), and a member of a regional assessment center who is not a district employee. Portions of the Georgia model have been used by many states across the country, and many similar elements appear in the Iowa Teaching Standards and Criteria. In particular is the concept of including the expectation of evaluator training around the state-mandated teaching competencies.

McClanahan and Peterson (1987) conducted research on the types of training in which principals had participated to insure that they would be qualified to evaluate teachers in Arizona. Evaluators indicated that they had received more than one type of training, including (a) reading professional literature (77.7%), (b) district-sponsored workshops and seminars (75.4%), (c) university training (five or more years before the survey) (63.8%), (d)

professional organization workshops or seminars (60.8%), (e) university coursework within five years of the survey (39.2%), (f) Principal's Academy (31.5%), (g) workshop with an independent contractor (23.8%), and (h) Arizona State Department of Education activity (17.7%). Perhaps most surprising in the findings, a small percent (2.3 percent) of the evaluators indicated no formal training to evaluate teachers.

McGreal (1983) stressed the importance of the district providing all members of the school with appropriate training and guided practice in the skills and knowledge necessary to implement and effectively maintain the system. Conley (1987), doing similar work, described the substantive skills needed by evaluators in to conduct quality evaluations: data collection, methods of observation, data analysis, conferencing, goal-setting, report-writing, and teacher remediation techniques. Brown (1997) found that principals were more optimistic about the process than were teachers, and there was a need for additional training for principals in the West Virginia model that was the basis for his study.

Not everyone is convinced that mandates in general, or mandates related to teacher evaluation, specifically, will achieve the results that are intended. In an essay on mandates, Kelly (1999) noted that current models of supervision rely on teachers being evaluated by one or more individuals and places value on compliance, not excellence. Drawing on the work of Deming, Kelly argued that if one wants to improve the product (student achievement), what one must constantly do is ask – not tell – the workers how to do it. The extension to teacher evaluation indicates that improvement will only come when all participants in the educational process undergo extensive self-assessments designed to improve each individual. Absent this change, the only excellence that will come about in public education is amongst those individuals who practice self-assessment on their own.

Frase and Streshly (2000) indicated that one of the myths in public education is that "teacher evaluation ensures high-quality instruction" (p. 112). The myth, based on their research, is threefold: (a) the results of evaluations are accurate, (b) legislation requirements for teacher evaluation ensure the public that only good teachers are in classrooms – the evaluations lead to improved instruction, and (c) school principals are adequately trained to conduct high-quality teacher evaluations and offer legitimate suggestions for improvement. They provide potential solutions suggesting that school boards state publicly and widely that top-notch instruction is required, boards create policies that say that principals must take time to work with teachers in classrooms and evaluate them to bring about improved teaching, budget money for training for teachers to improve instruction each year and for administrators to seek and take high-quality training in the area of teacher evaluation, boards direct superintendents to develop work plans for principals which allow for plenty of time in classrooms – eliminating developing reports, reducing meetings, etc, and having board direct superintendents to develop a series of criteria for assessing each of these actions for quarterly updates to the board.

The state-mandated teacher evaluation model for Iowa does not provide the level of specificity of expectation as is observed in examples such as the TPAI model. The IEATP training manual provides some evidence about the expectations for teacher evaluation in the state of Iowa. In addition to, at minimum, a meet/does not meet approach to carrying out the state teaching standards and criteria, the overall purpose of effective teacher evaluation is defined:

> Obtain *valid (meaningful)* and *reliable (consistent)* information about teacher performance in order to provide *clear, credible feedback* and make *defensible decisions* about the teacher's performance. (Training Module 1, IEATP)

The goals of an effective teacher evaluation system, according to the IEATP Training Manual are (a) accountability in educator performance, (b) improvement of professional development and school improvement systems, (c) improvement of both individual and collective teaching practices in the school, (d) to sharpen the awareness of the process of instruction within teacher's classroom practice, (e) to accelerate remedial help by principals and teacher leaders, and (f) to permit situation-specific rather than standardized assistance. This module goes on to say that evaluation is a process, not an event. This notion of teacher evaluation as an ongoing conversation about professional growth is widely held (Danielson, 1996; Danielson & McGreal, 2000; Glickman, 2002b).

It is clear that there is a wide variation of thought related to the role of state-mandated teacher evaluation systems and the approaches to delivering those systems. While the models for delivery vary widely, all of them place an emphasis on the quality of the evaluator and a focus on the individual needs of the teacher being evaluated as well as the district in which the evaluations are occurring. One of the realities that is also likely true of all of the evaluation models is that they are susceptible to a number of types of bias.

Gender bias in teacher evaluation

Any time that evaluation is occurring, there is opportunity for bias to enter into the equation. There are indications in the literature about types of teacher evaluation biases from age (Amos, 1988), to race (Brown, 2005; Ward & Sistrunk, 1988) and subject area (Ward & Sistrunk, 1988), to work values such as honesty (Winter et al., 1997), to gender (Cioci, 1991; Gougeon & Hutton, 1993; Rinehart & Young, 1996), and to the ability to write a quality anticipatory set (Bourisaw, 1989). In addition, Peterson (1998) found that evaluators in the

middle of their career gave more strict evaluations than those new to administration or near the end of their administrative careers. This study focused on a potential bias based on teacher and/or evaluator gender.

The potential for gender bias can go in both directions. That is to say, evaluators may have a bias based on, among other things, their own gender or the gender of the teacher (IEATP, Participant Manual, Fall, 2002; Peterson, 1998). Peterson (1988) also indicated that female evaluators gave significantly lower ratings than male evaluators, as did evaluators of different racial groups.

The reverse can also be true in that the perceptions of the teacher may be biased toward the evaluator, or the results of evaluation, based on their own gender or the gender of the evaluator (Cunningham, 2004; Gougeon & Hutton, 1993; Rinehart & Young, 1996). Peterson (1998) found in her study that there was no teacher-evaluator gender interaction, while Cioci (1991) found strong support for a teacher-evaluator gender interaction.

Many of the studies indicate that there is a preference toward favoring women evaluators, regardless of the gender of the teacher. The preferences for female evaluators are observed by teachers as the ability to display more instructional leadership behaviors (Cunningham, 2004), for factors of professionalism and instruction (Rinehart & Young, 1996), and as more positive communicators (Gougeon & Hutton, 1993). Gougeon and Hutton (1993) also found that male teachers did not see female principals as using threats or sanctions as often as did male principals. On a bit of a supportive and contrary note, Cioci (1991) observed that female teachers felt more empowered in school's with female teachers whiles males felt less empowered. At the same time, male teachers viewed female principals as ineffective, while females viewed the same principals as above average. These findings

indicate that the gender of the teacher, gender of the evaluator, and gender interaction between a teacher and her/his evaluator have the potential to impact the teacher evaluation process. A final thought related to training adults, in general, is an understanding of adult learning theory. The next section provides background on the unique learning needs of adults.

Adult Learning Theory

Adults have a different approach to learning than is observed in learning approaches for juveniles, which is often referred to as andragogy (Knowles, 1967). While there are differences in beliefs about adult learning across researchers, there are some points of broad agreement (Knowles, 1973; Leypoldt, 1967; Speck, 1996). Those areas of agreement related to how adults learn include: (a) adults will commit to learning when the objectives are realistic and important to them; (b) adults will resist learning they perceive to be an attack on their competence; (c) adults need to see a connection between the training and their every day learning and work activities; (d) adults need to participate in small group activities in order to construct meaning and receive feedback on their learning; and (e) coaching and follow-up supports are needed to make sure that learning has transferred into daily practice.

These principles of adult learning theory apply to training and learning for principals and other evaluators. Fullan (2002) believed that principals learning while at work – in context – is the only type of learning that can make a difference. Conferences and workshops can further development, but do not create collective knowledge and commitments. Amendt (2005) found that Iowa evaluators felt as though additional training would have benefit in the areas of teacher effectiveness and teacher evaluation around the Iowa Teaching Standards

&Criteria. An understanding of adult learning theory has an impact on understanding the Iowa Evaluator Approval Training Program.

Iowa Evaluator Approval Training Program (IEATP)

Iowa has had state-mandated teacher evaluation expectations for decades. As has been noted, there have been at least three major revisions to this area of legislation in the last 30 years. In response to the SF 205, Manatt (1976) proposed a process for a performance evaluation cycle that would provide for procedural due process and sound supervisory practice:

1. Self-appraisal for familiarization and preparation for the postconference.
2. Preobservation conferences to discuss instructional objectives, methods and the learners.
3. Classroom observations – two or three periods per cycle.
4. Postobservation conferences to discuss critical classroom incidents, progress and to exchange questions.
5. Agreement on a plan of action.
6. Time to improve, help to improve and mutual (appraiser-appraisee) monitoring of change.
7. Report of the summary evaluation to appraisee and to superiors.

Following this change, there was legislation in 1986 that required evaluator approval training in the form of I-LEAD. Most recently, a "switch" has occurred in the philosophy such that the mandates for teacher evaluation have become more specific. In response to Iowa Legislation (Iowa Code 284.10), the Iowa Evaluator Approval Training Program (IEATP) was developed. This training program was designed through a partnership of School Administrators of Iowa, Iowa's Area Education system, the University of Northern Iowa, and the Southeast Region Vision for Education program. While the training program was in the process of being designed, a call was put out for those who would be willing to be trained in

order to become regional trainers. 75 regional trainers and 15 higher education representatives spent 12 days becoming trained in the IEATP.

The IEATP was taught in tandem with a four day course referred to as the Data-Driven Leadership (DDL) Program. This training was centered on the QIC-Decide Model for decision-making (DDL Participant Manual, 2002a). In the DDL framework, the first step is for the persons who are involved in the decision making process to determine what they need data for and how the data will be used when collected – this is often done most effectively by writing a questions or addressing the decision (Q). The next step is for these same stakeholders to identify the specific information that will be needed to address the question or decision and need to keep the consequences for the decision in mind (I). The higher the stakes of the decision, then more data that is likely needed to make the best decision. The next step is to collect and summarize the information that is considered to be the action step, as the data will need to be organized and presented in a meaningful way (C). The final step is to use the data summarizations to make a decision about the original question that began the process (Decide) (DDL Participants Manual, 2002a).

Because this training was partially funded by the Bill and Melinda Gates Foundation, it could not be a required training for evaluators in Iowa. The Gates Foundation does not allow payment for required training as a part of the work of the Foundation. As such, some evaluators were able to opt-out of the training by passing a competency test. This study will focus on the IEATP as the DDL Model is regarded as more of a problem-solving process that can be applied across a variety of contexts, rather than necessarily being a part of the evaluation process, and because the DDL training was not technically required training for all Iowa evaluators.

The IEATP Participant's Manual (Fall, 2002b) provides the intent of legislation as it relates to the IEATP:

> Certified evaluators under this program will…
> • Understand theory behind best practices of teacher evaluation.
> • Demonstrate ability to provide data-based leadership.
> • Be able to identify quality instruction in the classroom.
> • Validate effective teaching that supports the Iowa Teaching Standards and Criteria.
> • Provide coaching in a professional growth environment. (p. 7)

The program is divided into three learning components, or modules, and the training occurs over 6 full days. An important portion of the training requires skill validations, rather than simply "setting and getting" the information from the trainer provided by the state of Iowa. The IEATP Participant's Manual indicates the following:

> Participants will successfully complete the following activities for validation of skills. Each of these activities will be collected, reviewed and validated by the trainers:
> • Site-based Activity in each of the three modules.
> • Check for Understanding in each of the three modules.
> • Site-based coaching-feedback activity. (p. 1, "Great Expectations Powerpoint")

There are teachings from the IEATP that will be exhibited at the school building level through changes in evaluator practices. One change in evaluator practices will be an emphasis by administrators on state-mandated teaching standards and criteria. The state has established eight standards and forty-two supporting criteria intended to define teacher behaviors and attitudes that will lead to quality teaching and improved student achievement. The licensure transition activity from the initial license to the career license is intended to ensure that the beginning teacher has demonstrated some level of competence in each of the Iowa Standards and Criteria. It is worth noting that while there are state established standards and criteria, districts are able to define the levels of performance contained in each of the criteria. In other words, districts set the bar for accepted teacher behaviors for each criterion.

A second change in evaluator practices will be an emphasis on teaching behaviors and teaching strategies in the classroom. In the past, principals would have been trained to focus on the behaviors in the classroom such as the number or times boys or girls were called upon, where the teacher stood during the lesson, or the number of affirmations provided to the students. The new model places an emphasis on evaluator focus on teaching behaviors and strategies that demonstrate quality teaching and lead to increased student achievement. This increases the expectations for evaluator training as it implies that evaluators know and can recognize effective teaching. It also implies that by using the knowledge, they are able to provide feedback and reflective conversation that will yield improved teaching in the classroom.

A third change in evaluator behaviors will be an emphasis on data collection from a variety of data sources (the "people") and data points (the "artifacts). No longer is the evaluator the only source for feedback in evaluation frameworks. The teacher is expected to use student feedback and work products, parent feedback, and other data sources. In addition, documentation for evaluation will not only be lesson plans and guides, but additional artifacts that support effective teaching strategies and the implementation of the teaching standards and criteria. Since this study was completed, the state requires that data sources include students, parents, teachers, and other evaluators. In addition, the state provides additional guidance on meaningful artifact collection from a variety of data sources and points. Many schools have interpreted this portion of the state expectation to be that beginning teachers complete a professional teaching portfolio.

A fourth change in evaluator behaviors will center on coaching in a professional development environment. This will be formalized as a part of the post-observation

conference. This change in the delivery of the post-observation conference is predicated on a collegial relationship between the evaluator and the teacher. To move through the coaching questions at the post-observation conference in a meaningful way, there needs to be some level of trust established in the evaluator-teacher relationship. An emphasis in the conference will be on using a variety of levels of questioning referred to as ORID questioning techniques. These questions are described as objective, reflective, interpretive, and decisional (Nelson, 2001).

Nelson (2001) provided explanations about the different levels of questions. Objective questions are about facts and external reality, or impressions. Reflective questions call forth immediate personal reaction to the data, an internal response, emotions or feelings, hidden images and associations with facts. Interpretive questions draw out meaning, values, significance, and implications. Decisional questions elicit resolution, bring the conversation to a close, and enable individuals or the group to make a decision about the future.

Some changes that teachers would notice in the move to the state-mandated teacher evaluation system would be an expectation of some type of data collection devise. This may be generically regarded as a professional teaching portfolio; however, a portfolio is not required in the law. The collection of artifacts would generally be constructed around the Iowa Teaching Standards and Criteria and would be intended to demonstrate teacher implementation of the expected criteria. It would also exhibit information from a number of data points and a variety of data sources including artifacts such as lesson plans, parent notes, student test scores, and evaluator feedback documents.

In addition, there may be a change in the manner in which the teacher and the evaluator interact with one another. With an emphasis on ORID questioning for improved

instruction, rather than on specific classroom behaviors, the post-observation conference may have a different feel, as may the conversation related to pre-observation conference materials such as lesson plans and lesson planning questions. Rather than be directive, it will be intended to be a model based on coaching and reflection. There may also be a sense that there is a shift to a more open and collegial relationship rather than that of a traditional evaluator and evaluatee as a part of this change in the approach to dialogue.

Summary

This chapter has provided a foundation for an understanding of leadership in general, leadership in the principalship and the expectations for evaluators related to teacher evaluation models. In addition, there is background on teacher career cycles, important elements of any teacher evaluation model such as context, content, and participants and state-mandated teacher evaluations. The final talking points of the chapter were on adult learning theory, the IEATP, and perceived changes that may result in teacher evaluation as implementation of the evaluator training is implemented across the state of Iowa. The next chapter will place an emphasis on the research contained in the study and the manner in which the study was conducted.

CHAPTER 3. METHODOLOGY

Introduction

This chapter describes the research methods, research questions, description of the statistical analysis, research design, instrumentation, human subjects release, and data collection, processing, and analysis. This study featured descriptive statistics and correlational methods, including t-tests, pairwise comparisons, multivariate tests, within-subjects tests, and multiple analysis of variance [MANOVA]. The correlational methods were used to test for differences in novice teacher perceptions of their first year of teaching, 2001-02, in which the Iowa Teaching Standards and Criteria were not implemented and their second year of teaching, 2002-03, the first year of the Iowa Teaching Standards and Criteria in Iowa, in tandem with the IEATP. They were also used to study potential differences in perception of the attributes of teacher evaluation based on the gender of the teacher and/or the gender of the evaluator.

Research Questions

The following research questions will guide this study:

1. What do novice teachers report concerning perceptions of attributes of self as a teacher, perceptions of his/her evaluator, perceptions of the attributes of the evaluation processes, perceptions of the attributes of the feedback, and perceptions of the context in which the evaluations occurred?
2. Do factors such as gender of the teacher, gender of the evaluator, or gender interactions between the teacher and the evaluator impact perceptions about teacher evaluation?

3. Are there differences in perceiver data from the 2001-02 school year to the 2002-03 school year (before and after the IEATP and the ITS&C) around attributes of self as a teacher, perceptions of his/her evaluator, perceptions of the attributes of the evaluation processes, perceptions of the attributes of the feedback, and perceptions of the context in which the evaluations occurred?

Population of the Study

The population of this study consisted of all of the Iowa novice teachers who were in their first two years of teaching in the 2001-02 and 2002-03 school years. These teachers taught in an Iowa public school and were evaluated by the same evaluator in both of those school years. The total number of teachers in this study was approximately 350 potential participants, and the names and schools in which they teach were obtained from the Iowa Department of Education through the Basic Education Data Survey (BEDS) database.

This group of teachers was selected as they were the only teachers in the state who would have been evaluated in their first year of teaching by an evaluator not yet trained in the IEATP expectations and in their second year of teaching by the same evaluator who had received the training. This may have an impact on one or more of the five attributes of teacher evaluation (Stiggins & Duke, 1988). They are also the first group who would be held to a licensure transition activity, from initial to career, based on a state-established list of expected teaching standards and criteria. The teachers from the study are also the first group in the state who will be expected to produce artifacts to support their work from a variety of data sources and data points. This group is also likely the first to experience district changes in the products and paperwork associated with district evaluation systems.

Statistical Procedures

A letter was sent to all novice teachers who were in their first two years of teaching, in an Iowa public school, in 2001-02 and 2002-03 (see Appendix A). This letter provided the possible respondents with information about the survey, their individual participation code, and a weblink from which to access the electronic survey (see Appendix B).

The first part of the survey included questions relating to the personal teaching attributes of the novice teachers as well as their perceptions of the evaluator as they relate to traits such as: (a) a credible source of meaningful feedback, (b) trustworthy, and (c) knowledgeable of the technical aspects of teaching. There were also questions relating to the evaluation process and products. Those questions fell into the categories of procedures used to address the dimensions of teaching to be evaluated, attributes of the feedback received, resources available for professional development, and district values and policies as they relate to evaluation. Questions also related to the amount of formal and informal feedback the novice teacher received from the evaluator. Finally, four open-ended questions were asked about topics which included: (a) additional feedback to those who may have responsibility for potential changes to the teacher quality legislation, (b) a focus on the ITS&C, (c) specific personal experiences of interest, and (d) overall experience(s) with the evaluation process.

The second part of the survey sought demographic information from the respondents. That demographic information included items such as gender of both the novice teacher and the evaluator, ethnicity of both the novice teacher and evaluator, and the grade level(s) taught by the novice teacher. The ethnicity and grade levels taught were a part of the data that was lost in a move of the electronic survey from one host server to another host.

Survey Instrument

Development

The framework for the instrument is the Teacher Evaluation Profile (TEP) Questionnaire, which was originally designed and used by Stiggins and Duke (1988) for work they were doing in the area of research on teacher evaluation. Since that time, modified forms of the instrument have been used in other studies on teacher evaluation (Lawler, 1992; Rindler, 1994).

For each of the non-demographic statements in the survey, a 5-point Likert-type range from strongly agree to strongly disagree has been provided for review by the participants. Respondents will also need to provide reactions for the statements from the perspectives of both the 2001-02 school year and the 2002-03 school year.

Validation

The instrument was validated by a committee of Iowa State professors. This group was selected as they approved my proposal and was the group that would be ultimately responsible to hold me accountable to the results of the survey as a part of the process of dissertation defense.

Stiggins was contacted by phone in December of 2003 for consideration as to the validity of the instrument. Stiggins indicated he had not used the instrument in some time and that follow-up work to the original research was likely in order. He also referred this research back to the data analysis that occurred in the original book from which the TEP was drawn.

The original validation completed by Stiggins and Duke (1988) included three phases of data analysis. The first phase of the analysis was an investigation of information related to

the dependability of the original structure of the items built into the questionnaire. Through Pearson product moment correlations, among others, it was observed that there is cohesiveness among the items, but they do not exactly match the five sections of the questionnaire. With the exception of context, it would appear there is evidence that each of the categories has an important place in the structure of the item responses.

The second phase of their original validation work focused on the relationships between attributes of the evaluation and the outcome of the event as observed by the teacher. The purpose was to explore the relative importance of various attributes in determining the outcomes of evaluation. A number of correlations were computed in working on this phase of the analysis. It was discovered that with respect to attribute-outcome relationships, there is a great deal of evidence to suggest that the attributes covered in the questionnaire are related to the teachers' ratings of the overall quality and impact of an evaluation event.

The third phase of the analysis addressed issues related to the sensitivity of the questionnaire to differences in evaluations across the districts being observed in the original study. Correlation matrices were developed and analyzed and a multivariate analysis of variance was carried out with the school districts as the independent variable. The results demonstrated the instrument was sensitive enough to detect differences in district evaluation. Overall, they concluded from the evidence that (a) the instrument provides high-quality information about the teachers' perceptions of their evaluation experience and (b) there is a strong relationship between specific attributes of an evaluation event and the outcomes of that event as perceived by the teachers (p. 115).

The instrument, modified for this study, was also pilot tested in March of 2004 by a small number of novice teachers from the Waukee Community School District in Waukee,

Iowa. The reasons for choosing this population for this portion of the validation were to receive clarity on wording in the instrument as well as the removal of any potential barriers to completion of the survey. The teachers who completed the original draft of the survey provided feedback on clarification of wording for some of the questions and indicated that the survey was easy to navigate and to complete.

Human Subjects Approval

The letter inviting novice teachers to participate in the study was mailed after the Iowa State University Human Subjects Committee approved the letter of invitation and the survey instrument (see Appendix C). Voluntary consent was obtained as respondents completed the online survey or returned the completed paper version of the survey in the mail. In both cases they were reminded of their rights before beginning the survey and were provided the opportunity to discontinue the completion of the survey at any time in the survey completion process.

Data Collection and Analysis

In April of 2004 the online survey website was established in cooperation with staff members from Iowa State University. In May of 2004, the Human Subjects Committee approved letter of invitation was sent via the United States Postal Service to 344 potential participants. This letter contained an explanation of the study, a unique participation code for each respondent, and the web address for access to the survey. The participation code was provided as a tool designed to limit the set of responses from each participant to one completion and to connect the data from each participant to his/her unique responses. The participation code was also intended for use in disaggregating data based on school size and

teacher ethnicity among other characteristics. In an electronic data dropping error, this connection between the data and the participation code was lost which resulted in the loss of some of the data manipulation opportunities.

In February of 2005 e-mails were sent to non-respondents asking each of them to complete the online survey. Again, attached to each e-mail document was the unique participation code as well as a link to the web address for the survey. This process for making e-mail contacts occurred through the first half of the month of February. On March 24, 2005, all non-completers were sent a second e-mail reminder asking each to complete the online survey. On Friday, April 29, Monday, May 2, or Tuesday, May 3, 2005, depending on completion of the mail stuffing, all non-completers were sent a paper copy of the original invitation to complete the survey, a paper copy of the survey (See Appendix D), and a self-addressed stamped envelope in which to return the completed paper survey. On May 2, 2005, an e-mail reminder was sent to all non-completers in order to make them aware that a paper copy of the survey would be coming to each of them in the US Mail. Phone calls were made to the schools of all non-completers on May 23 and 24, 2005. The opportunity to complete either the paper or online survey was brought to a close on July 1, 2005.

Personnel in the Iowa State University Research Institute for Studies in Education (RISE), using the Statistical Package for the Social Sciences (SPSS) software, assisted in loading and analyzing the data. Descriptive statistics and quantitative methodology were used to analyze, report, and interpret the data.

Summary

This chapter focused on the process used to design and implement the research methodology for this study. Some important elements of the methodology include choosing and validating the instrument, collecting the data and the approach taken in analyzing the data. The next chapter of this study will emphasize the results of the data collection and analysis through the use of data tables and written descriptors. It will also include support comments harvested from open-ended questions contained at the end of the survey instrument used for the study.

CHAPTER 4. RESULTS

Introduction

The purpose of this study was to investigate the quality and effectiveness of the newly mandated teacher evaluation process in Iowa based on the potential changes in teacher perceptions of the attributes of teachers, attributes of their supervisor, attributes of the procedures of evaluation, and attributes of the feedback. The study utilized descriptive statistics and quantitative methodology to provide a thorough understanding of this research topic. Qualitative narrative data were also collected as a part of the study and those comments supported and enriched the quantitative findings.

General Characteristics of the Sample

The original survey instrument was developed using the Teacher Evaluation Profile (TEP) Questionnaire, written by Stiggins and Duke (1988). This survey was modified to direct the questioning to the emerging needs of novice teachers through the process of teacher evaluation. The survey instrument design included Likert-scale items, four open-ended responses, and demographic information as it relates to gender of the teachers and evaluators. The questionnaire attempted to elicit responses related to the attitudes and opinions of the novice teachers.

The Iowa Department of Education provided a listing of all of the 2002-03 public school second-year teachers as well as a listing of all of the 2003-04 public school third-year teachers. Teacher licensure folder numbers were used to identify those teachers who were in both their second year of public school teaching, in Iowa, in 2002-03 and third year of teaching, in Iowa, in 2003-04. A sort and comparison of the lists of teachers, using the

teacher licensure folders to verify that it was the same person in 2002-03 and 2003-04, regardless of the name, completed this combined list. A total of 344 teachers met these original criteria. These teachers were mailed a letter inviting them to participate in the study by visiting an online website to complete an electronic survey.

An e-mail reminder was sent to each non-completer. There were 64 potential respondents who could not be contacted via e-mail due to no e-mail address or an incorrect e-mail address. The Iowa Department of Education was contacted, and it was learned that 20 of these 64 potential respondents were no longer teaching in a public school in Iowa. As a result, they were removed from the candidate pool of possible respondents. The school and address of each of the other 44 possible respondents were provided. These potential respondents were often located in a school other than that in which they taught during the first two years of teaching, sometimes in the same district and sometimes in a new district, while others had a change in last name. All of these possible respondents were also contacted via e-mail.

Approximately three months later, a paper copy of the survey was then sent to each non-responder followed by two e-mail reminders. Two of the paper surveys came back as "not at this address – no forwarding address". In making a call to each school, the person had left and no forwarding information was available. As a result, these two possible respondents were removed from the candidate pool.

Phone calls were also made to the schools of those who had not completed the survey. In the course of that reminder process, it was discovered that five possible respondents were no longer working in the indicated district, and there was no follow-up information from their current school. It was also learned that one of the possible respondents was killed in a

car accident, one had completed her first two years of teaching in a private school, one was working under a grant, and as such, was not on a teaching contract, so she was not eligible for formal evaluation, and one had two evaluators in her first two years of teaching. All of these potential respondents were removed from the candidate pool.

The first two questions on the survey indicate the need for the same evaluator for the first two years of teaching as well as the successful transition of the first two years of teaching as defined by the transition from a provisional to a standard teaching license. There were 99 total respondents who did not meet either or both of these two criteria, as indicated by a "no" to either or both of the first two questions in the survey. As a result, these respondents were removed from the survey population.

From the original list of 344 possible respondents, a total of 31 were removed for the reasons noted above. Of the remaining possible 313 respondents, 202 responded for a 64.5% response rate. Of these 202 respondents, 99 were ineligible due to the additional criteria listed in the survey. A total of 103 participants responded from the remaining potential pool of 214 respondents for a 48.1% completer response rate to the survey.

Because of the number of possible responders who were removed from the study for the reasons described or for marking "no" to one or more of the opening questions, the respondents appear to be a survivor population. As a result, it is possible that this group has a more positive outlook on the profession, given the fact that they have continued to participate in the occupation. Knowing that about the population, it is more likely that the results are conservative and would be more pronounced given a higher percentage of teachers who met all of the criteria.

Demographics of the Respondents and their Evaluators

In addition to data collected about perceptions of novice teachers related to evaluation, demographic data were collected relative to the gender of the teachers as well as the gender of the evaluators. Table 2 shows the gender breakdown of the respondents as well as the gender breakdowns of the respondent's evaluator. This table represents all returned surveys in which both the gender of the respondents and the gender of the respondent's evaluator were indicated.

Table 2. Gender of the respondents and the respondent's evaluator

	Female		Male	
Population	N	Percent	N	Percent
Respondent	67	72.0	26	28.0
Respondent's evaluator	31	33.3	62	66.7

Analysis of returned surveys revealed demographic distributions closely resembling the total populations represented in the survey. For example, the breakdown of female respondents at 72.0% is similar to the percentage of female teachers in the Iowa teaching population in 2002-03 at 72.0% (Iowa Department of Education, 2005) and the national teaching population at 74.4% (U.S. Dept. of Ed., 2001). The comparison is similar as it relates to the gender of evaluators. In the completed surveys, 33.3% of the respondent's evaluators were female. The percentage of female administrators in the Iowa population in 2002-03 was 34.1% (Iowa Department of Education, 2005) and national population was 34.5% (U.S. Dept. of Ed., 2001).

Statistical Analysis of t he Data

The SPSS statistical software package was used to calculate both descriptive and correlational statistics. The use of descriptive statistics is intended to provide a comprehensive view of the results for each of the Likert-type questions on the study survey. A mean for each Likert-type question was calculated as an arithmetic average for all of the completed respondents for each of the sets of compared questions from 2001-02 to 2002-03. In cases where one or more pieces of data were missing for a respondent relative to the subset of observed traits, all of the data for the traits described by that respondent were eliminated from the average for that subset. Standard deviations provide the reader with a sense for how far each of the values reported differ from the mean and from one another. That is to say, the smaller the standard deviation for a particular survey question, the less variability there is in the set of responses observed for the question (Hinkle, Wiersma, & Jurs, 1998)

Correlational methods including t-tests, pairwise comparisons, multivariate tests, within-subjects tests, and multiple analysis of variance [MANOVA] were used to test for relationships among variables. Using a two-tailed test, a significance level of less than .05 would be a cause for further investigation of the relationship between the variables by the observer. However, in many comparisons of variables that, at face value, appeared to show significance, the Bonferroni-t adjustment for multiple comparisons further tightened the scrutiny level for the definition of significance among the variables. As would be expected, this adjustment led to a decrease in the number of statistically significant comparisons.

The traditional repeated measures ANOVA assumes sphericity. This condition is satisfied if the correlation is identical between responses for each pair of categories of the

within-subjects factor (Agresti & Finlay, 1997). Mauchly's Test for Sphericity was often calculated as a part of the comparisons contained in this study. In all cases, the test for sphericity was violated, which indicates that more than one factor may be impacting the responses from the novice teacher.

When the sphericity condition is violated, the true distribution of the F statistic with the degrees of freedom can be approximated using a number of responses to sphericity such as Greenhouse-Geisser, Huynh-Feldt, and Lower-bound (Kirk, R.E., 1995). These responses to a lack of sphericity were calculated for the survey results and in all cases, they were also violated. Some potential factors that were not accounted/ controlled for in the survey design, but may be impacting the results include factors such as the following: layers in school (elementary, middle, high school), gender percentage differences in various grade levels, styles of the leadership, or longevity of the leaders. There may be other unidentified factors involved that this study did not address.

Attributes of self as a teacher

Research question one focused on traits novice teachers reported concerning perceived attributes of self as a teacher, perceptions or their evaluator, perceptions of the evaluation processes, perceptions of the attributes of the feedback, and perceptions of the context in which the evaluations occurred.

Respondents rated nine attributes of self as a teacher. Table 3 provides a breakdown of responses for all of the traits of attributes of self as a teacher for 2001-02. Using a Likert-type scale, responses were 1 for strongly disagree, 2 for disagree, 3 for neither agree nor disagree, 4 for agree and 5 for strongly agree. The responses from 98 respondents indicated

Table 3. Traits for attributes of self as a teacher, in 2001-02 (N=98)

Trait	Mean	S.D.
In terms of professional expectations, I demand a great deal from myself. (T1)	4.50	.542
I consider myself to be open to constructive criticism. (T5)	4.36	.482
As it relates to my classroom, I consider myself relatively open to change. (T3)	4.33	.570
I have a great deal of knowledge about the subject matter that I am responsible to teach. (T7)	4.16	.796
I am quick to implement suggestions I have received from my evaluator. (T8)	4.13	.698
I consider myself to be a teacher who frequently engages in instructional experimentation in the classroom. (T4)	3.89	.823
I embraced the district teaching standards and criteria as appropriate for my classroom. (T9)	3.65	.839
In terms of my professional orientation, I consider myself to be an instructional "risk-taker". (T2)	3.62	.891
I have a great deal of knowledge about the technical aspects of teaching. (T6)	3.45	.839

that for the year 2001-02, the trait of *demanding a great deal from myself* (M = 4.5, SD = .542) was the mostly highly agreed with trait in the survey. This most agreed upon trait was followed closely by the traits of *open to constructive criticism* (M = 4.36, S.D. = .482) and *relatively open to change* (M = 4.33, S.D. = .570). The trait of a *great deal of knowledge about the technical aspects of teaching* (M = 3.45, S.D. = .839) was the least agreed upon trait as it related to attributes of self as a teacher.

Table 4 provides a breakdown of responses for all of the traits of attributes of self as a teacher for 2002-03. The 98 respondents indicated that the teacher trait of *demanding a great deal from myself* (M = 4.61, SD = .490) was also the most highly agreed upon trait. The trait of *a great deal of knowledge about the subject matter I teach* (M = 4.42, S.D. = .535) was the next most agreed upon trait in the survey, followed closely by the traits of *open to constructive criticism* (M = 4.38, S.D. = .546) and *relatively open to change* (M = 4.37,

Table 4. Traits for attributes of self as a teacher, in 2002-03 (N=100)

Trait	Mean	S.D.
In terms of professional expectations, I demand a great deal from myself. (T1)	4.61	.490
I have a great deal of knowledge about the subject matter that I am responsible to teach. (T7)	4.42	.535
I consider myself to be open to constructive criticism. (T5)	4.38	.546
As it relates to my classroom, I consider myself relatively open to change. (T3)	4.37	.544
I am quick to implement suggestions I have received from my evaluator. (T8)	4.21	.656
I consider myself to be a teacher who frequently engages in instructional experimentation in the classroom. (T4)	4.07	.700
In terms of my professional orientation, I consider myself to be an instructional "risk-taker". (T2)	3.81	.837
I embraced the district teaching standards and criteria as appropriate for my classroom. (T9)	3.78	.733
I have a great deal of knowledge about the technical aspects of teaching. (T6)	3.70	.798

S.D. = .544). The least agreed upon trait was, again, the trait of a *great deal of knowledge about the technical aspects of teaching* (M = 3.7, S.D. = .798).

From 2001-02 to 2002-03, in general terms related to descriptive statistics, there was an overall increase in the means for all of the traits in the attributes of self as a teacher. This indicates that the perceptions of novice teachers were more in agreement with the traits from the first year of the study to the second year of the study. There was also some change in the order of the means from highest to lowest in the transition from 2001-02 to 2002-03. In addition, the standard deviations, overall, tended to be smaller numbers from year one to year two. This indicates that the survey numbers used to calculate the means for each trait showed less variation in 2002-03 than in 2001-02 and as such, more agreement.

In addition to comparing the overall means of the traits, the difference between the means provides an indication of those traits that represent significant agreement or disagreement to other traits in the same cluster. A numerical guideline can be generated to

indicate whether a trait or traits is/are considered significantly different from other traits in the same cluster of traits. This guideline is generated by taking the number of comparisons in a cluster of traits (traits minus one, or n-1 as a Bonferroni correction) and multiplying that number by the level of significance for statistical scrutiny in the study. The product gives an approximation of the chance of being wrong about the level of relative significance contained within the cluster of traits. If this product is then multiplied by the number of traits in the cluster and rounded to the nearest whole number, it can be generalized that a trait or traits the exceed(s) this number in a positive or negative direction is different, in general, from the traits in the cluster that do not meet this numerical threshold.

For example, in the cluster of traits known as attributes of self as a teacher, there are eight comparisons. Multiplying the comparisons (eight) by the significance used in this study (.05) yields what could be described as an error approximation of .4 for the cluster of traits. When that product is multiplied by the number of traits in the cluster (9), the product is 3.6, which is rounded up to 4. That is to say, if a trait in the cluster is significant in the same direction, either positive or negative, from four or more traits in the cluster, it can be described as a more frequently (positive) or less frequently (negative) agreed upon trait than the typical trait in the cluster of traits. Table 5 provides an indication of the relative significance, either positive (+) or negative (-), of difference between each of the personal teaching attributes of the novice teachers.

The * symbol used in Table 5 indicates a significant difference in means between two traits for the 2001-02 school year. The Δ symbol used in Table 5 indicates a significant difference in means between two traits for the 2002-03 school year. The center of the table from top left to bottom right is blank as that is where each trait is compared to itself. Table 5

Table 5. Comparison of the attributes of self as a teacher, in 2001-02 and 2002-03

	T1	T2	T3	T4	T5	T6	T7	T8	T9
T1		* (+) Δ (+)	Δ (+)	* (+) Δ (+)	Δ (+)	* (+) Δ (+)	* (+)	* (+) Δ (+)	* (+) Δ (+)
T2	* (-) Δ (-)		* (-) Δ (-)	* (-)	* (-) Δ (-)		* (-) Δ (-)	* (-) Δ (-)	
T3	Δ (-)	* (+) Δ (+)		* (+) Δ (+)		* (+) Δ (+)			* (+) Δ (+)
T4	* (-) Δ (-)	* (+)	* (-) Δ (-)		* (-) Δ (-)	* (+) Δ (+)	Δ (-)		
T5	Δ (-)	* (+) Δ (+)		* (+) Δ (+)		* (+) Δ (+)			* (+) Δ (+)
T6	* (-) Δ (-)		* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)		* (-) Δ (-)	* (-) Δ (-)	
T7	* (-)	* (+) Δ (+)		Δ (+)		* (+) Δ (+)			* (+) Δ (+)
T8	* (-) Δ (-)	* (+) Δ (+)				* (+) Δ (+)			* (+) Δ (+)
T9	* (-) Δ (-)		* (-) Δ (-)		* (-) Δ (-)		* (-) Δ (-)	*(-) Δ (-)	

$p < .05$ (2-tailed with Bonferroni adjustment)
* = 2001-02
Δ = 2002-03

contains both symbols in order to provide a pictorial representation of a comparison of the significance of the means of the traits. Both the top and bottom of the chart, above and below the blank diagonal, is completed in order to be able to count the number of significant incidences of each trait compared to all others.

Overall, in 2001-02, the trait of *demand a great deal from self* (T1) shows a significant positive difference from six of the other eight traits. The traits of *relatively open to change* (T3) and *open to constructive criticism* (T5) show significant positive differences from four of the other eight traits. Conversely, the traits of *instructional risk-taker* (T2) and *knowledge about the technical aspects of teaching* (T6) show significant negative differences from six of the other eight traits. The trait of *embracing the district standards as appropriate*

to my classroom (T9) shows significant negative differences from five of the eight other traits.

In 2002-03, the trait of *demand a great deal from self* (T1) shows a significant positive difference from seven of the other eight traits. The trait of *great deal of knowledge about the subject matter* (T7) shows a significant positive difference from four of the other eight traits. The traits of *relatively open to change* (T3) and *open to constructive criticism* (T5) show significant positive differences from four of the other eight traits, and one significant, negative difference from the other eight traits. The trait of *knowledge about the technical aspects of teaching* (T6) shows a significant, negative difference from six of the other eight traits. *Embracing the district teaching standards as appropriate to the classroom* (T9) shows a significant, negative difference from five of the other eight traits. The trait of *frequently engaging in instructional experimentation* (T4) shows a significant negative difference from 4 of the other traits, and 1 significant, positive difference from one of the other traits.

For the 2001-02 school year, three of the nine traits of attributes of self as a teacher showed four or more significant, positive differences in mean when compared to the other traits. The traits include the following: *demanding a great deal from self* (T1), *relatively open to change* (T3) and *open to constructive criticism* (T5). For the 2002-03 school year, two of the nine traits of attributes of self as a teacher showed four or more significant, positive differences in mean when compared to the other traits. Those traits are *demand a great deal from self* (T1) and *possessing a great deal of knowledge about the subject matter* (T6). The traits of *relatively open to change* (T3) and *open to constructive criticism* (T5) show

significant positive differences from four of the other eight traits, and one significant, negative difference from the other eight traits.

The trait of *demanding a great deal from self* (T1) goes from being significantly different from six of the other eight traits in year one to being significantly different from seven of the other eight traits in year two. The traits of *relatively open to change* (T3) and *open to constructive criticism* (T5) are observed in both years as meeting the earlier defined threshold of being significantly different from four of the other traits. However, for both, a trait of significant, negative difference appears for both as the trait of *demanding a great deal from self* (T1) becomes even more different from either of the two traits in terms of a comparison of the means. The trait of *possessing a great deal of knowledge about the subject matter* (T6) emerges as meeting the threshold for difference as a trait in the second year of the study.

In 2001-02 three of the nine traits of attributes of self as a teacher showed more than four significant, negative differences in means when compared to the other traits. Those traits include the following: *instructional risk-takers* (T2), *knowledgeable about the technical aspects of teaching* (T6) and *embracing the district standards as appropriate to their classrooms* (T9). In 2002-03, three of the nine personal traits showed four or more significant, negative differences in mean when compared to the other traits, as well. Those traits include the trait of *knowledge about the technical aspects of teaching* (T6), *embracing the district teaching standards as appropriate to the classroom* (T9), and *frequent engagement in instructional experimentation in the classroom* (T4). It should also be noted that the trait of *frequent engagement in instructional experimentation in the classroom* (T4)

shows a significant negative difference from 4 of the other traits, and 1 significant, positive difference from one of the other traits.

The traits of *knowledge about the technical aspects of teaching* (T6) and *embracing the district teaching standards as appropriate to the classroom* (T9) appear in both years of the study. By year two of the study the trait of *instructional risk-takers* (T2) no longer meets the defined threshold as different from the other traits. However, the trait of *frequent engagement in instructional experimentation in the classroom* (T4) emerges as a trait that is different from that of the rest of the cluster of traits by 2002-03.

Overall, as it relates to attributes of self as a teacher, in both years of the study there was continued positive agreement with the trait *demanding a great deal from self* (T1), and the difference from other traits became even more pronounced from year one of the study to year two. Novice teachers see themselves as possessing this trait. In this study there was also statistically significant agreement with the traits *relatively open to change* (T3) and *great deal of knowledge about the subject matter* (T7), although not to the extent as was observed with *demanding a great deal from self* (T1). Both did become less like other traits from year one of the study to year two of the study and as a result, more of a trait of a novice teacher. It can be said that novice teachers perceive themselves to possess these traits when compared to other traits in the attributes of self as a teacher. In year one of the study the trait *open to constructive criticism* (T5) appeared to be a trait of novice teachers but did not meet the threshold definition by year two of the study.

The traits of *instructional risk-takers* (T2), *knowledge about the technical aspects of teaching* (T6), and *embracing the district teaching standards as appropriate to the classroom* (T9) are not traits novice teachers perceive to possess when compared the other traits in the

attributes of self as a teacher. This is true in both years of the study. By the second year of the study, the trait of *frequent engagement in instructional experimentation in the classroom* (T4) becomes a trait that is not like the others in the grouping and as such, it is not a trait of novice teachers when compared to others in the trait cluster of attributed of self as a teacher.

An indication about the high expectations that novice teachers have for themselves is indicated in a number of comments related to expectations they have for all teachers. A sampling of seemingly related comments from the open-ended survey questions include:

> "I think the evaluation process is good, to keep teachers on track and teaching. But this needs to happen for all teachers, not just new teachers. New teachers are up on most recent changes – unlike those who have been teaching 20+ years and may not be accepting of change."

> "I am a go-getter, but I really fretted over meeting all the criterion and collecting all the artifacts. It was extremely overwhelming to a new teacher."

In summary, the attributes of self as a teacher for the years 2001-02 and 2002-03 are defined in Table 6. In both 2001-02 and 2002-03, novice teachers in this study regarded themselves as possessing the traits of *demand a great deal from self, relatively open to*

Table 6. Traits and non-traits of novice teachers, in 2001-02 and 2002-03

Trait	2001-02	2002-03
Demand a great deal from self (T1)	+ (4.50)	+ (4.60)
Relatively open to change (T3)	+ (4.32)	+ (4.35)
Open to constructive criticism (T5)	+ (4.35)	+ (4.37)
Great deal of knowledge about the subject matter (T7)		+ (4.41)
Instructional risk-taker (T2)	- (3.59)	
Knowledge about the technical aspects of teaching (T6)	- (3.43)	- (3.69)
Embracing the district standards as appropriate to my classroom (T9)	- (3.64)	- (3.78)
Frequently engaging in instructional experimentation (T4)		- (4.07)

change, and *open to constructive criticism.* By 2002-03, the trait of *great deal of knowledge about the subject matter* emerges as a perceived trait of the group.

In 2001-02 the trait of *instructional risk-taker* is not observed as a trait of the novice teachers but the same perception is not observed for the trait by the second year of the study. In both 2001-02 and 2002-03 the traits of *knowledge about the technical aspects of teaching* and *embracing the district standards as appropriate to my classroom* are perceived to not be traits of the group as they relate to self as teachers. By 2002-03 the trait of *frequently engaging in instructional experimentation* emerges as not being a trait of the group when compared to all other traits in the perceptions of self cluster of traits.

This information provides a glimpse into the perceptions of self as it relates to novice teachers; both the traits which with then tend to agree and those with which they do not agree. The next section of data organization is focused on novice teacher perceptions of traits of evaluators.

Perceptions of the evaluators

Respondents rated eleven attributes of their evaluator using the same Likert-type scale as was defined in the attributes of self as a teacher. Table 7 provides a breakdown of responses for traits of evaluators. The responses from 100 respondents indicate that for the year 2001-02, the trait of *the evaluator demonstrated flexibility through the evaluation process* ($\underline{M}$ = 4.11, SD = .875) was the most highly agreed with trait in this portion of the survey. This most agreed upon trait was followed by the traits of *my evaluator is trustworthy* ($\underline{M}$ = 4.02, SD = 1.054) and *my evaluator's interpersonal manner is non-threatening* ($\underline{M}$ = 4.01, SD = .980).

Table 7. Traits for attributes of evaluators, in 2001-02 (N=100)

Trait	Mean	S.D.
My evaluator demonstrated flexibility through the evaluation process. (T5)	4.11	0.875
My evaluator is trustworthy. (T3)	4.02	1.054
My evaluator's interpersonal manner is non-threatening. (T4)	4.01	0.980
My evaluator is knowledgeable about the technical aspects of teaching. (T6)	3.97	0.915
My evaluator is a credible source of feedback. (T1)	3.68	1.188
The suggestions my evaluator provides are useful to my professional development. (T10)	3.66	0.945
My evaluator is helping in promoting my ongoing professional growth. (T2)	3.60	1.054
My evaluator is able to provide a persuasive rationale for suggestions for improvement. (T11)	3.57	1.037
My evaluator is familiar with my particular classroom. (T8)	3.57	1.139
My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s). (T7)	3.36	1.106
My evaluator is familiar with classrooms in general within my school. (T9)	2.95	1.306

The least agreed upon trait was *my evaluator is familiar with classrooms in general within my school* (M = 2.95, SD = 1.306). The trait of *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s)* was next (M = 3.36, SD = 1.106) followed by a tie for the next two least agreed upon traits of *my evaluator is able to provide a persuasive rationale for suggestions for improvement* (M = 3.57, SD = 1.037) and *my evaluator is familiar with my particular classroom* (M = 3.57, SD = 1.139).

The same eleven traits of evaluators were also rated by the novice teachers for the 2002-03 school year. Table 8 provides a listing of the order of perceived traits of evaluators for 2002-03. The trait of *my evaluator demonstrated flexibility through the evaluation* was again the most agreed upon trait (M = 4.13, SD = .881) in this section of the survey. The next three most agreed upon traits include the following: *my evaluator's interpersonal manner is non-threatening* (M = 4.01, SD = 1.010), *my evaluator is trustworthy* (M = 3.98, SD =

Table 8. Traits for attributes of evaluators, in 2002-03 (N=100)

Trait	Mean	S.D.
My evaluator demonstrated flexibility through the evaluation process. (T5)	4.13	0.881
My evaluator's interpersonal manner is non-threatening. (T4)	4.01	1.010
My evaluator is trustworthy. (T3)	3.98	1.045
My evaluator is knowledgeable about the technical aspects of teaching. (T6)	3.94	0.929
The suggestions my evaluator provides are useful to my professional development. (T10)	3.72	0.871
My evaluator is a credible source of feedback. (T1)	3.69	1.143
My evaluator is helping in promoting my ongoing professional growth. (T2)	3.65	1.036
My evaluator is able to provide a persuasive rationale for suggestions for improvement. (T11)	3.64	0.977
My evaluator is familiar with my particular classroom. (T8)	3.52	1.133
My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s). (T7)	3.35	1.104
My evaluator is familiar with classrooms in general within my school. (T9)	2.89	1.307

1.045), and *my evaluator is knowledgeable about the technical aspects of teaching* (M = 3.94, SD = .929).

The least agreed upon trait was again *my evaluator is familiar with classrooms in general within my school* (M = 2.89, SD = 1.307). The next least agreed upon trait was *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s)* (M = 3.35, SD = 1.104).

In regard to descriptive statistics for the study relating to attributes of evaluators, the mean scores from 2001-02 to 2002-03 tended to increase very little, and in many cases, were the same number or even decreased from year one to year two. In addition, the overall mean scores for the perceptions of evaluators were generally lower than the overall mean scores that were observed for the attributes of self as a teacher. That is to say, in general, the novice teachers were more in agreement with the traits of teachers than the traits of evaluators.

The standard deviations tended to be the same from year one to year two for the perceptions of evaluators, and in some cases, there was increased variability in the scores. This increase in standard deviations indicates that there was less agreement in perceptions of the novice teachers related to the traits from 2001-02 to 2002-03. In addition, this cluster of traits contains a score that demonstrates that the respondents disagree with the contents of the trait *my evaluator is familiar with classrooms in general within my school* (T9).

A comparison of the significance of the means, as was completed with the traits of teacher perceptions of self, is also tabulated for the traits of evaluators. Table 9 provides an

Table 9. Comparison of the attributes of evaluators, in 2001-02 and 2002-03

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
T1			* (-) Δ (-)		* (-) Δ (-)		Δ (+)		* (+) Δ (+)		
T2			* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-)			* (+) Δ (+)		
T3	* (+) Δ (+)	* (+) Δ (+)					* (+) Δ (+)	* (+) Δ (+)	* (+) Δ (+)	* (+)	* (+) Δ (+)
T4		* (+) Δ (+)					* (+) Δ (+)	* (+) Δ (+)	* (+) Δ (+)	* (+)	* (+) Δ (+)
T5	* (+) Δ (+)	* (+) Δ (+)					* (+) Δ (+)	* (+) Δ (+)	* (+) Δ (+)	* (+) Δ (+)	* (+) Δ (+)
T6		* (+)					* (+) Δ (+)	* (+) Δ (+)	* (+) Δ (+)	* (+)	* (+) Δ (+)
T7	Δ (-)		* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)			* (+) Δ (+)	* (-) Δ (-)	Δ (-)
T8			* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)			* (+) Δ (+)		
T9	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)		* (-) Δ (-)	* (-) Δ (-)
T10			* (-)	* (-)	* (-) Δ (-)	* (-)	* (+) Δ (+)		* (+) Δ (+)		
T11			* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	Δ (+)		* (+) Δ (+)		

$p < .05$ (2-tailed with Bonferroni adjustment)
* = 2001-02
Δ = 2002-03

indication of the relative significance, either positive (+) or negative (-), of difference between each of the personal attributes of the evaluators as perceived by novice teachers. In this portion of the survey, this relative significance of means is calculated by multiplying the number of comparisons, ten (11-1) by the significance level used in the study (.05) which yields an error approximation of .5. This number is then multiplied by the number of traits (11) and rounded up to the next whole number of six.

The attributes of their evaluator for the years 2001-02 and 2002-03 are summarized in Table 10. The results for 2001-02 show the traits of *my evaluator demonstrated flexibility through the evaluation process* (T5) and *my evaluator is trustworthy* (T3) as being significantly positive in difference from seven of the other ten traits. The trait of *my evaluator's interpersonal manner is non-threatening* (T4) and the trait of *my evaluator is knowledgeable about the technical aspects of teaching* (T6) are significantly different, in a positive direction, from six of the other ten traits. The trait that demonstrates a significant negative difference from 10 of the other 10 traits is *my evaluator is familiar with classrooms in general within my school* (T9).

Table 10. Traits and non-traits of evaluators, in 2001-02 and 2002-03

Trait	2001-02	2002-03
My evaluator demonstrated flexibility through the evaluation process (T5)	+ (4.11)	+ (4.11)
My evaluator is trustworthy (T3)	+ (4.01)	+ (3.99)
My evaluator's interpersonal manner is non-threatening (T4)	+ (4.02)	
My evaluator is knowledgeable about the technical aspects of teaching (T6)	+ (3.94)	
My evaluator is familiar with classrooms in general within my school (T9)	- (2.95)	- (2.87)
My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice (T7)		- (3.32)

In 2002-03, the trait of *my evaluator demonstrated flexibility through the evaluation process* (T5) showed a significant positive difference from seven of the other ten attributes of evaluators. The trait of *my evaluator is trustworthy* (T3) showed similar results for six of the ten attributes of evaluators. The attribute of *my evaluator is familiar with classrooms in general within my school* (T9) showed a significant negative difference from all ten of the other attributes of evaluators. The attribute of *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s)* (T7) was significantly different from seven of the ten attributes and exhibited a positive difference from one of the traits.

For the 2001-02 school year there were four traits the showed a significant, positive difference from six or more of the other ten traits. Those traits include the following: *my evaluator demonstrated flexibility through the evaluation process* (T5), *my evaluator is trustworthy* (T3), *my evaluator's interpersonal manner is non-threatening* (T4), and *my evaluator is knowledgeable about the technical aspects of teaching* (T6). By 2002-03, there were two traits that emerged as significantly different in the positive direction from six or more of the other ten traits. Those two traits were *my evaluator demonstrated flexibility through the evaluation process* (T5) and *my evaluator is trustworthy* (T3).

The trait in year one of the study that meets the threshold for six or more of the ten traits in the negative direction is *my evaluator is familiar with classrooms in general within my school* (T9). In year two the traits that meet this same threshold are *my evaluator is familiar with classrooms in general within my school* (T9), and *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s)* (T7).

From 2001-02 to 2002-03 the traits that no longer meet the numerical threshold, on the positive side, statistically significant from six or more of the other traits, are *my*

evaluator's interpersonal manner is non-threatening (T4) and *my evaluator is knowledgeable about the technical aspects of teaching* (T6). In the other direction, the trait of *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s)* (T7) emerges by the second year of the study.

Overall, in both year one of the study and year two of the study there were two traits to which novice teachers demonstrated agreement as traits of the principal, when compared to all other traits in the group. Those two traits are, *My evaluator demonstrated flexibility through the evaluation process* (T5), and *My evaluator is trustworthy* (T3). The traits of *My evaluator's interpersonal manner is non-threatening* (T4), and *My evaluator is knowledgeable about the technical aspects of teaching* (T6), were agreed upon in year one of the study but were not agreed upon by year two of the study.

In both years of the study it was agreed upon, more so than any other trait in the principal trait cluster, that the trait *my evaluator is familiar with classrooms in general within my school* (T9) is not a trait of principals. In addition, the trait of *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s)* (T7) was perceived as not a trait of principals by the second year of the study. No other traits in the principal grouping met the definition for significant difference from other traits in the novice teacher perceptions of the attributes of evaluators trait cluster.

In both years of the study novice teachers perceive the evaluators as possessing the traits of *my evaluator demonstrated flexibility through the evaluation process* and *my evaluator is trustworthy*. In 2001-02 the traits of *my evaluator's inter-personal manner is nonthreatening* and *my evaluator is knowledgeable about the technical aspects of teaching* are observed as a perceived trait of evaluators, but they are not observed as traits by 2002-03.

The trait of *my evaluator is familiar, in general, with classrooms in my school* is not perceived as a trait of evaluators in either year of the study. The trait of *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice* emerges as not being a trait of the group by year two, based on perceptions of novice teachers.

Frequent comments from the open-ended questions in the survey that might further describe perceptions of evaluators include the following:

"Evaluators should have to spend more time observing."

"I would have liked to have been evaluated, just to know how I am doing as a first year teacher. My Principal [sic] did not come into my classroom at all and made stuff up to cover himself."

"New teachers need time for specific, positive, regular, informal evaluation with administration."

"Evaluators [sic] should be in classrooms more than twice a year."

"My evaluator, who was also the principal of my school, did not give me any constructive [sic] criticism. I would like suggestions for improvement."

"My evaluations are basically a replay of my lesson. Really enjoy my principal. He is just never in the classroom. He has not seen one lesson this year."

It is clear that traits related to evaluators emerge in the data, and differ a bit from year one of the study to year two. The next section of data interpretation will center on the evaluation processes. These processes center on activities such as direct observation in the classroom and reviews of evaluation artifacts.

Perceptions of the evaluation processes

In the survey, respondents were originally expected to rate six attributes of the evaluation processes. However, when the survey was posted to the web address, one of the questions, *a review of student achievement data was part of my evaluation(s)*, was

accidentally omitted. It was retained in the paper copy of the survey. As a result, approximately one third of the surveys contain feedback associated with this attribute. Because of this discrepancy, data from this question was not included in the study.

Table 11 provides a breakdown of responses for traits of evaluation processes for 2001-02. Respondents rated five attributes of the evaluation processes, again, using the Likert-type scale that was defined in the attributes of self as a teacher. The responses from 101 respondents indicate that for the year 2001-02, the trait of *direct observation of my classroom performance was used extensively in my evaluation(s)* (M = 3.64, SD = 1.361) was the most highly agreed with trait in the survey. The least agreed upon trait was *the district teaching standards and criteria were differentiated to meet my unique learning needs* (M = 2.92, SD = 1.093).

Table 11. Traits of evaluation processes, in 2001-02 (N=101)

Trait	Mean	S.D.
Direct observation of my classroom performance was used extensively in my evaluation(s). (T4)	3.64	1.361
The district teaching standards and criteria were communicated to me in detail. (T1)	3.43	1.228
The content of the district teaching standards and criteria were clear to me. (T2)	3.33	1.193
A review of classroom or school records (lesson plans, etc.) was used extensively in my evaluation(s). (T5)	3.27	1.207
The district teaching standards and criteria were differentiated to meet my unique learning needs. (T3)	2.92	1.093

Table 12 provides a breakdown of responses for traits of evaluation processes for 2002-03. The responses from 99 respondents indicate that for the year 2002-03, the trait of *the district teaching standards and criteria were communicated to me in detail* (M = 4.00, SD = .915) was the most highly agreed with trait in the survey. Again, the least agreed upon

Table 12. Traits of evaluation processes, in 2002-03 (N=99)

Trait	Mean	S.D.
The district teaching standards and criteria were communicated to me in detail. (T1)	4.00	0.915
The content of the district teaching standards and criteria were clear to me. (T2)	3.87	0.911
Direct observation of my classroom performance was used extensively in my evaluation(s). (T4)	3.73	1.391
A review of classroom or school records (lesson plans, etc.) was used extensively in my evaluation(s). (T5)	3.37	1.200
The district teaching standards and criteria were differentiated to meet my unique learning needs. (T3)	3.13	1.027

trait was *the district teaching standards and criteria were differentiated to meet my unique learning needs* (M = 3.13, SD = 1.027).

The means for all of the traits increased, that is to say there was generally more agreement with the traits, from year one to year two of the study. There was not an even increase in the means of the traits with one trait, *direct observation of my classroom performance was used extensively in my evaluation(s)* (T4) increasing .09, while the trait *the district teaching standards and criteria were communicated to me in detail* (T1) increased as much as .57. It is also observed that that the standard deviations tended to decrease from year one to year two as the scores showed less variability in 2002-03 when compared to 2001-02.

A comparison of the significance of the means, as completed with the traits of teachers perceptions of self, was also tabulated for the traits of evaluation processes. The number of comparisons, four, is multiplied by the significance level used in the study (.05) for an error approximation of .2. This product is then multiplied by the number of traits, five, for a numerical threshold of difference in traits of one. Because the number of traits in the cluster is small, a decision was made by the researcher to use a threshold difference of more

than one as the threshold for a trait to be considered different from the other traits in the group.

Table 13 provides an indication of the relative significance, either positive (+) or negative (-), of difference between each of the evaluation processes as perceived by novice teachers. In 2001-02 one of the traits, *direct observation of my classroom performance was used extensively in my evaluation(s)* (T4), showed more than one mean difference of significance when compared to the other traits of the evaluation processes. The directions of the significant comparisons were both positive. The trait of *the district teaching standards and criteria were differentiated to meet my unique learning needs* (T3) showed negative significance from more than one of the other four traits.

In 2002-03 three of the traits showed more than one mean difference of significance, in the positive direction, when compared to the other traits of the evaluation process. Those three traits were *the district teaching standards and criteria were communicated to me in*

Table 13. Comparison of the evaluation processes, in 2001-02 and 2002-03

	T1	T2	T3	T4	T5
T1			* (+) Δ (+)		Δ (+)
T2			* (+) Δ (+)		Δ (+)
T3	* (-) Δ (-)	* (-) Δ (-)		* (-) Δ (-)	
T4			* (+) Δ (+)		* (+) Δ (+)
T5	Δ (-)	Δ (-)		* (-) Δ (-)	

$p < .05$ (2-tailed with Bonferroni adjustment)
* = 2001-02
Δ = 2002-03

detail (T1), *the content of the district teaching standards and criteria were clear to me* (T2), and *direct observation of my classroom performance was used extensively in my evaluation(s)* (T4). For the same year of the study, the traits of *a review of classroom or school records (lesson plans, etc.) was used extensively in my evaluation(s)* (T5) and *the district teaching standards and criteria were differentiated to meet my unique learning needs* (T3) showed more than one mean difference in the negative direction.

The attributes of the evaluation processes for the years 2001-02 and 2002-03 are summarized in Table 14. The trait of *direct observation of my classroom performance was used extensively in my evaluation(s)* (T4) was one of the most agreed upon traits in this portion of the survey for both years of the study. Novice teachers agree that in both years of the study, direct observation in the classroom was a foundation of the evaluation process. By the second year of the study the most agreed upon traits added the two traits of *the district teaching standards and criteria were communicated to me in detail* (T1), and *the content of the district teaching standards and criteria were clear to me* (T2).

Table 14. Traits and non-traits of the evaluation process, in 2001-02 and 2002-03

Trait	2001-02	2002-03
Direct observation of my classroom performance was used extensively in my evaluation(s). (T4)	+ (3.63)	+ (3.75)
The district teaching standards and criteria were communicated to me in detail. (T1)		+ (3.97)
The content of the district teaching standards and criteria were clear to me. (T2)		+ (3.87)
The district teaching standards and criteria were differentiated to meet my unique learning needs (T3)	- (2.91)	- (3.12)
A review of classroom or school records (lesson plans, etc.) was used extensively in my evaluations. (T5)		- (3.38)

Conversely, the trait of *the district teaching standards and criteria were differentiated to meet my unique learning needs* (T3) was the least agreed upon trait in year one and persisted into year two. The trait of *a review of classroom or school records (lesson plans, etc.) was used extensively in my evaluation(s)* (T5) was one of the least agreed upon traits by year two. That is to say, there was more disagreement with this trait from the first year of the study to the second year of the study.

Direct observation of the classroom is a trait of the group, while differentiation of the process to meet teacher needs is not a trait of the processes. A comment from one of the novice teachers appears to hit this sentiment related to differentiation of the teaching standards and criteria:

> "Legislators need to be aware that there is not a one-fits-all evaluation tool that can be used effectively for every teacher [sic]. There [sic] must be some flexibility to evaluation programs that will meet the qualifications of the teachers."

Traits related to the evaluation processes emerge as those traits that are more agreed upon, such as direct observation, is used in the process, and those traits such as differentiation occurs, that are not. The next portion of the survey centered on the attributes of the feedback including dimensions of feedback such as specificity, and a focus on district teaching standards.

Perceptions of the attributes of the feedback

Respondents rated five attributes related to feedback received during the evaluation process, using the same Likert-type scale as that used in the attributes of self as a teacher survey. Table 15 provides a breakdown of responses for traits of attributes of the feedback for 2001-02. The responses from 103 respondents indicated that the trait of *the nature of the*

Table 15. Traits of attributes of the feedback, in 2001-02 (N=103)

Trait	Mean	S.D.
The nature of the information provided by my evaluator was descriptive rather than judgmental. (T4)	3.95	0.964
I received a great deal of feedback from my evaluator that was directly applicable to my classroom. (T1)	3.51	1.259
The ideas and suggestions contained in the feedback from my evaluator were of high quality. (T2)	3.46	1.195
The information provided by my evaluator was very specific. (T3)	3.45	1.297
The feedback from my evaluator was focused on district teaching standards and criteria. (T5)	3.38	1.147

information provided by my evaluator was descriptive rather than judgmental (M = 3.95, SD = .964) was the most highly agreed with trait in the survey. The least agreed upon trait was *the feedback from my evaluator was focused on the district teaching standards and criteria* (M = 3.38, SD = 1.147).

Table 16 provides a breakdown of responses for traits of attributes of the feedback for 2002-03. The responses from 103 respondents indicate that *the nature of the information provided by my evaluator was descriptive rather than judgmental* (M = 3.98, SD = .918) was again the most highly agreed with trait in the survey. The least agreed upon trait was *the information provided by my evaluator was very specific* (M = 3.53, SD = 1.227).

Table 16. Traits of attributes of the feedback, in 2002-03 (N=103)

Trait	Mean	S.D.
The nature of the information provided by my evaluator was descriptive rather than judgmental. (T4)	3.98	0.918
The feedback from my evaluator was focused on district teaching standards and criteria. (T5)	3.65	1.045
I received a great deal of feedback from my evaluator that was directly applicable to my classroom. (T1)	3.59	1.175
The ideas and suggestions contained in the feedback from my evaluator were of high quality. (T2)	3.55	1.118
The information provided by my evaluator was very specific. (T3)	3.53	1.227

The mean scores for all of the traits in this cluster of traits increased from 2001-02 to 2002-03. It is worth noting that for the first time in the study none of the mean scores in a cluster of traits exceeded agree (4 on the Likert-type scale) in either year of the study. It is also observed that the standard deviations tended to decrease from year one to year two as an indication that the scores showed less variability.

A comparison of the significance of the means, as was completed with the traits of teachers' perceptions of self, was also tabulated for the traits of attributes of the feedback. The number of comparisons, four, is multiplied by the significance level used in the study (.05) for an error approximation of 0.2. This product is then multiplied by the number of traits, five, for a numerical threshold of difference in traits of one. Because the number of traits in the cluster is small, a decision was made by the researcher to use a threshold difference of more than one as the threshold for a trait to be considered different from the other traits in the group. Table 17 provides an indication of the relative significance, either positive (+) or negative (-), of difference between each of the attributes of the feedback received as perceived by novice teachers.

Table 17. Comparison of the attributes of the feedback, in 2001-02 and 2002-03

	T1	T2	T3	T4	T5
T1				* (-) Δ (-)	
T2				* (-) Δ (-)	
T3				* (-) Δ (-)	
T4	* (+) Δ (+)	* (+) Δ (+)	* (+) Δ (+)		* (+) Δ (+)
T5				* (-) Δ (-)	

$p < .05$ (2-tailed with Bonferroni adjustment)
* = 2001-02
Δ = 2002-03

The attributes of the feedback for the years 2001-02 and 2002-03 are defined in Table 18. The trait of *the nature of the information provided by my evaluator was descriptive rather than judgmental* (T4) showed a positive significance from all four of the other traits in both 2001-02 and 2002-03. There was generally agreement with this statement, especially when comparing the mean score for the traits by year two of the study.

Table 18. Traits and non-traits of the attributes of the feedback, in 2001-02 and 2002-03

Trait	2001-02	2002-03
The nature of the information provided by my evaluator was descriptive rather than judgmental. (T4)	+ (3.95)	+ (3.98)

This trait represents the only trait that was agreed upon by novice teachers for this dimension of evaluation. Some comments from novice teachers related to evaluator feedback include:

"Hold the administrators accountable for evaluating the teachers. I wanted him to take this as seriously as I did. He never looked at my portfolio, was in my room for a total of 10 min. and had nothing to say other than I already know you are a 'good teacher'. I was looking to get some guidance so I could be a better teacher."

"My evaluator was a former P.E. teacher he (in my opinion) had little experience in the classroom. He knew little about my subject matter. Because of these things I felt my feedback wasn't useful or applicable to my course."

"I had an awful experience with my evaluator she did not do any formal observation of me and did not provide me with any feedback. However, I had a wonderful mentor the second year that made up for it."

There is not broad agreement with the process for delivery of feedback to novice teachers, nor with the content and effectiveness of the feedback. This appears to be an area for further discussion as it relates to the conclusions contained in this document. The next segment of the survey was focused on the context in which evaluations occurred, which asks

questions related to district professional growth opportunities and expectations as well as district policy statements related to evaluation.

Context in which the evaluations occurred

Respondents rated five attributes of the context in which the evaluations occurred, using the same Likert-type scale used previously when measuring the attributes of self as a teacher. Table 19 shows a complete breakdown of responses for traits of the context in which the evaluations occurred for 2001-02. The responses from 101 respondents indicated that the trait of *the evaluation(s) is/are intended to place a high emphasis on teacher professional growth* (M = 3.69, SD = .845) was the most highly agreed with trait in this section of the survey. The least agreed upon trait was *an adequate amount of time was allotted during the school day for professional development* (M = 2.51, SD = 1.171).

Table 20 provides a breakdown of responses for traits of the context in which the evaluations occurred for 2002-03. The responses from 103 respondents indicated that the trait of *the evaluation(s) is/are intended to place a high emphasis on teacher professional growth*

Table 19. Traits of the context in which the evaluations occurred, in 2001-02 (N=101)

Trait	Mean	S.D.
The evaluation(s) is/are intended to place a high emphasis on teacher professional growth. (T5)	3.69	.845
The evaluation(s) is/are intended to place a high emphasis on teacher accountability. (T4)	3.59	.918
The district has clear policy statements regarding the purpose(s) of evaluation. (T3)	3.48	1.016
My district makes many training programs/models of best practices available to teachers. (T2)	3.20	.990
An adequate amount of time was allotted during the teaching day for professional development. (T1)	2.51	1.171

Table 20. Traits of the context in which the evaluations occurred, in 2002-03 (N=103)

Trait	Mean	S.D.
The evaluation(s) is/are intended to place a high emphasis on teacher professional growth. (T5)	3.82	.813
The evaluation(s) is/are intended to place a high emphasis on teacher accountability. (T4)	3.68	.899
The district has clear policy statements regarding the purpose(s) of evaluation. (T3)	3.56	.957
My district makes many training programs/models of best practices available to teachers. (T2)	3.20	1.023
An adequate amount of time was allotted during the teaching day for professional development. (T1)	2.58	1.159

($\underline{M}$ = 3.82, SD = .813) was, again, the most highly agreed upon trait in this section of the survey. The least agreed upon trait was again *an adequate amount of time was allotted during the school day for professional development* ($\underline{M}$ = 2.58, SD = 1.159).

The mean scores for all of the traits in this cluster increased except for the trait of *my district makes many training programs/models of best practices available to teachers* (T2), which remained the same from 2001-02 to 2002-03. It is also worth noting that this is the first section of the survey in which the order of the means did not change from year one of the study to year two of the study. This set of traits also contains the lowest mean, 2.58, of any of the clusters of traits in the study. The standard deviations tended to show less variability by the second year in higher scoring means but demonstrate increased variability in the less agreed upon two traits with the lowest scores. This cluster of traits also contains a score that indicates the respondents disagree with the trait described as *an adequate amount of time was allotted during the teaching day for professional development* (T1).

A comparison of the significance of the means, as was completed with the traits of teachers' perceptions of self, was also tabulated for the traits of attributes of the context in

which the evaluations occurred. The number of comparisons, four, is multiplied by the significance level used in the study (.05) for an error approximation of .2. This product is then multiplied by the number of traits, five, for a numerical threshold of difference in traits of one. Because the number of traits in the cluster is small, a decision was made by the researcher to use a threshold difference of more than one as the threshold for a trait to be considered different from the other traits in the group. Table 21 provides an indication of the relative significance, either positive (+) or negative (-), of difference between each of the attributes of the context in which the evaluations occurred.

In 2001-02, the traits of *the evaluation(s) is/are intended to place a high emphasis on teacher professional growth* (T5), and *the evaluation(s) is/are intended to place a high emphasis on teacher accountability* (T4) demonstrated more than one positive difference in significance of means when compared to the other traits. The trait of *my district makes many*

Table 21. Comparison of the context in which the evaluations occurred, in 2001-02 and 2002-03

	T1	T2	T3	T4	T5
T1		* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)	* (-) Δ (-)
T2	* (+) Δ (+)		Δ (-)	* (-) Δ (-)	* (-) Δ (-)
T3	* (+) Δ (+)	Δ (+)			
T4	* (+) Δ (+)	* (+) Δ (+)			
T5	* (+) Δ (+)	* (+) Δ (+)			

$p < .05$ (2-tailed with Bonferroni adjustment)
* = 2001-02
Δ = 2002-03

training programs/models of best practices available to teachers (T2) showed more than one negative difference from the other traits. It also showed one positive trait as the trait of *the evaluation(s) is/are intended to place a high emphasis on teacher professional growth* (T5) possesses a negative significance from all of traits in the group.

In 2002-03, three of the traits *the evaluation(s) is/are intended to place a high emphasis on teacher professional growth* (T5), *the evaluation(s) is/are intended to place a high emphasis on teacher accountability* (T4), and *the district has clear policy statements regarding the purpose(s) of evaluation* (T3) demonstrated more than one positive difference in significance of means when compared to the other traits. The trait of *my district makes many training programs/models of best practices available to teachers* (T2) showed more than one negative difference from the other traits. It also showed one positive trait as the trait of *the evaluation(s) is/are intended to place a high emphasis on teacher professional growth* (T5) possesses a negative significance from all of traits in the group.

The trait of *the district has clear policy statements regarding the purpose(s) of evaluation* (T3) appeared as different from more than one trait, in the positive direction, by the second year of the study. The trait of *an adequate amount of time was allotted during the teaching day for professional development* (T5) was significantly different from all other traits, in the negative direction, in both 2001-02 and 2002-03.

Overall, the traits *the evaluation(s) is/are intended to place a high emphasis on teacher professional growth* (T5), and *the evaluation(s) is/are intended to place a high emphasis on teacher accountability* (T4) were more agreed upon than the other traits in the group. That is to say, they are agreed upon as expectations for the context of evaluations. The

trait *the district has clear policy statements regarding the purpose(s) of evaluation* (T3) joins the agreed upon group by the second year of the study.

The attributes of the context in which the evaluations occurred for the years 2001-02 and 2002-03 are summarized in Table 22. The trait *an adequate amount of time was allotted during the teaching day for professional development* (T1) was the least agreed upon trait in both 2001-02 and 2002-03 and was disagreed with in the descriptive statistics. The trait *my district makes many training programs/models of best practices available to teachers* (T2) was different from a number of traits in year one of the study and even more so by year two of the study. Neither of these two traits would be seen as traits of the context in which evaluations occurred based on the perceptions of novice teachers.

Overall, there was not broad agreement around any of the traits of the context in which evaluations occurred. While teacher accountability appears to be an emerging portion of teacher evaluations, conversations about best practice and adequate time to implement

Table 22. Traits and non-traits of the context in which evaluations occurred, in 2001-02 and 2002-03

Trait	2001-02	2002-03
The evaluation(s) is/are intended to place a high emphasis on teacher professional growth. (T5)	+ (3.70)	+ (3.81)
The evaluation(s) is/are intended to place a high emphasis on teacher accountability. (T4)	+ (3.60)	+ (3.68)
The district has clear policy statements regarding the purpose(s) of evaluation. (T3)		+ (3.56)
My district makes many training programs/models of best practices available to teachers. (T2)	- (3.19)	- (3.22)
An adequate amount of time was allotted during the teaching day for professional development. (T1)	- (2.52)	- (2.58)

those practices remain a concern based on the perceptions of novice teachers. This marks the shift from looking at traits of evaluation to a focus on the role that gender plays, if any, in the teacher evaluation process.

Gender influence on the respondents' perceptions

Research question two asked if variables such as gender of the teacher and gender of the evaluator influenced their perceptions about evaluation. A variety of comparisons related to gender were completed as a part of the study. Comparisons of mean scores by gender of the novice teacher, gender of the evaluator, gender of both the novice teacher and the evaluator, and a combined model of the means of the gender of the novice teacher, and the evaluator by the average means of the scores related to the gender of both groups.

Some general assertions can be made relative to the descriptive statistics contained in the study. Overall, female novice teachers tended to agree less with the surveyed traits than did the male novice teachers. This was evidenced by lower overall means scores for the majority of traits in the survey for the perceptions of females compared to the males. In addition, the overall mean scores for the level of agreement for both genders increased from the 2001-02 school year to the 2002-03 school year. This was true for all of the types of comparisons that were completed for the survey.

Teacher

A comparison of the mean scores of the perceptions of the novice teachers, by gender of the teacher, produced one significant difference around a trait for the 2001-02 school year. The trait *I have a great deal of knowledge about the subject matter that I am responsible to teach* showed a significant difference in means of females ($\underline{M}$ = 4.031) to males ($\underline{M}$ = 4.500).

That is to say, male novice teachers were more in agreement with this statement than were the female novice teachers. The same comparison of means was completed for the 2002-03 school year. By the second year, there were no traits in which there was a significant difference, by gender of the teacher, in the mean scores of the perceptions of novice teachers.

Evaluator

A comparison of the mean scores of the perceptions of the novice teachers, by gender of the evaluator, produced significant differences around six traits for the 2001-02 school year. Table 23 provides the traits that differed by the gender of the evaluator in the first year of the study, as well as a comparison of the means by gender of the evaluator.

The trait, *the district teaching standards and criteria were differentiated to meet my unique learning needs,* was less agreed upon relative to male evaluators than it was relative to female evaluators. The trait *as it relates to my classroom, I consider myself relatively open to change*, while generally agreed upon relative to both genders of evaluators, was less

Table 23. Traits with a difference by gender of the evaluator, in 2001-02

	Evaluator	
Trait	Female	Male
The district teaching standards and criteria were differentiated to meet my unique learning needs.	3.517	2.810
As it relates to my classroom, I consider myself relatively open to change.	4.657	4.168
My evaluator is helping in promoting my ongoing professional growth.	3.960	3.379
My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s).	4.003	3.074
My evaluator is able to provide a persuasive rationale for suggestions for improvement.	4.020	3.305
Please rate your overall experience related to your district's formal teacher evaluation process.	2.607	3.262

agreed upon relative to male evaluators than female evaluators. *My evaluator is helping in promoting my ongoing professional growth*, as a trait, is more agreed upon relative to female evaluators than male evaluators. The most pronounced difference in means between male and female evaluators was the trait of *my evaluator has the capacity to demonstrate or model needed improvement in my classroom practices*. The trait was more agreed upon relative to female evaluators than male evaluators by almost an entire step in the Likert-type scale. The trait *my evaluator is able to provide a persuasive rationale for suggestions for improvement* was more strongly agreed upon for female evaluators than for male evaluators. The final comparison of significance by gender of the evaluator is a rating of the overall experience of formal observation with a score of 1 being excellent and 5 being very poor. It is described in the survey as *please rate your overall experience related to your district's formal teacher evaluation process*. In this case, the overall mean score for female evaluators is closer to excellent than that of the overall mean score for male evaluators.

A comparison of the mean scores of the perceptions of the novice teachers, by gender of the evaluator, produced significant differences around seven traits for the 2002-03 school year. Of the seven traits, five were also observed as significant in difference between the gender of evaluators in 2001-02. Table 24 provides the traits that differed by the gender of the evaluator in the second year of the study, as well as a comparison of the means by gender of the evaluator.

As shown in Table 24, the trait, *the district teaching standards and criteria were differentiated to meet my unique learning needs,* was observed in both 2001-02 and 2002-03, and, in both cases, was rated higher relative to female evaluators, when compared to male evaluators. The trait, *my district makes many training programs/models of best practices*

Table 24. Traits with a difference by gender of the evaluator, in 2002-03

Trait	Evaluator	
	Female	Male
The district teaching standards and criteria were differentiated to meet my unique learning needs.*	3.617	2.992
My district makes many training programs/models of best practices available to teachers.	3.573	3.023
As it relates to my classroom, I consider myself relatively open to change.*	4.717	4.185
My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s).*	3.983	3.049
My evaluator is familiar with classrooms in general within my school.	3.390	2.627
My evaluator is able to provide a persuasive rationale for suggestions for improvement.*	4.020	3.411
Please rate your overall experience related to your district's formal teacher evaluation process.*	2.540	3.418

*Denotes a trait seen in both 2001-02 and 2002-03.

available to teachers emerges as a trait of significant difference by year two of the study, with more agreement relative to female evaluators, although the significance level is not as strong (.047) as other traits in the survey. The trait *as it relates to my classroom, I consider myself relatively open to change* appears as significant in both years of the survey and is increasingly agreed upon relative to female evaluators rather than male evaluators. The trait *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s)* is observed as significant in difference in years one and two of the survey; however, the scores for both have become less agreed upon for both genders by 2002-03. *My evaluator is familiar with classrooms in general within my school* emerges as a trait of significant difference in 2002-03, with the trait more agreed upon in females compared to males. The mean scores for the trait *my evaluator is able to provide a persuasive rationale for suggestions for improvement* remain similar from year one to year two of the survey, with more agreement relative to female evaluators than male evaluators. In

the description of *please rate your overall experience related to your district's teacher evaluation process*, the scores are even less alike and are moving closer to excellent, with the female evaluators, and farther away from excellent than the male evaluators when comparing 2001-02 to 2002-03.

The trait of *my evaluator is helping in promoting my ongoing professional growth* is observed as significant in difference by gender of the evaluator in year one of the study only. The traits of *my district makes many training programs/models of best practices available to teachers,* and *my evaluator is familiar with classrooms in general within my school* emerge as traits seen as significantly different in year two of the study. As has been noted, there are five traits that are significant in difference of means for both 2001-02 and 2002-03. It is also worth noting that for all of these traits, the more agreed means are connected to the female evaluators, rather than the male evaluators.

Table 25 provides a summary of the traits observed as being statistically significant in difference based on the gender of the evaluator. Over all, in both years of the study, the traits *the district teaching standards and criteria were differentiated to meet my unique learning needs, as it relates to my classroom, I consider myself relatively open to change, my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s),* and *my evaluator is able to provide a persuasive rationale for suggestions for improvement* were statistically significant in agreement toward female evaluators. The same was true for the question that asked novice teachers to *rate your overall experience related to your district's teacher evaluation process*, as that score was closer to "excellent" for female evaluators than for male evaluators.

Table 25. Traits with a difference in response due to evaluator gender, in 2001-02 and 2002-03

Trait	2001-02	2002-03
My evaluator is promoting my ongoing professional growth.	Female (3.960) Male (3.379)	
The district teaching standards and criteria were differentiated to meet my unique learning needs.	Female (3.517) Male (2.810)	Female (3.617) Male (2.992)
As it relates to my classroom, I consider myself relatively open to change.	Female (4.657) Male (4.168)	Female (4.717) Male (4.185)
My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s).	Female (4.003) Male (3.074)	Female (3.983) Male (3.049)
My evaluator is able to provide a persuasive rationale for suggestions for improvement.	Female (4.020) Male (3.305)	Female (4.020) Male (3.411)
Please rate your overall experience related to your district's teacher evaluation process. (1=excellent)	Male (3.262) Female (2.607)	Male (3.418) Female (2.540)
My district makes many training programs/models of best practices available to teachers.		Female (3.573) Male (3.023)
My evaluator is familiar with classrooms in general within my school.		Female (3.390) Male (2.627)

As shown in Table 25, there is a gender difference based on the gender of the evaluator across a number of traits in the study. In every case, there was more agreement to the traits relative to female evaluators when compared to male evaluators. The only exception was a "higher" score for males in rating the overall experience of evaluation, although a lower score is closer to "excellent" in the evaluation process. A comparison was also made involving the gender of the teacher and the gender of the evaluator, rather than looking at them overall.

Novice teacher and the evaluator

A comparison of the mean scores by both gender of the novice teacher and gender of the evaluator provide two significant results for 2001-02. Table 26 provides a comparison of

Table 26. Traits with a difference by gender of the novice teacher and gender of the evaluator, in 2001-02

	Teacher (Evaluator)			
Trait	Female (Female)	Female (Male)	Male (Female)	Male (Male)
The district has clear policy statements regarding the purpose(s) of evaluation.	3.680	3.357	2.667	3.650
As it relates to my classroom, I consider myself relatively open to change.	4.480	4.286	4.833	4.050

the mean scores, by trait, at a significant level, based on the gender of both the novice teacher and the evaluator for year one of the study.

As shown in Table 26, the trait, *the district has clear policy statements regarding the purpose(s) of evaluation,* demonstrated the lowest mean score in the perception of male teachers who had a female evaluator. Relative to that trait, there was an indication that there was disagreement with this trait when compared to the other mean scores for the group. All of the groups indicated agreement with the trait *as it relates to my classroom, I consider myself relatively open to change*; however, the group of male teachers with male evaluators provided the lowest mean score for this trait.

Table 27 shows a comparison of the mean scores, by trait, at a significant level, based on the gender of both the novice teacher and the evaluator for year two of the study. A comparison of the mean scores by both gender of the novice teacher and the evaluator provided two significant results in 2002-03, as well. One of the traits remained the same as in year one of the study, while one trait was no longer significant in difference and a different trait emerged as being significant in difference.

Table 27. Traits with a difference by gender of the novice teacher and gender of the evaluator, in 2002-03

Trait	Teacher (Evaluator)			
	Female (Female)	Female (Male)	Male (Female)	Male (Male)
The district has clear policy statements regarding the purpose(s) of evaluation.	3.720	3.409	3.000	3.789
Please rate your overall experience related to your district's formal evaluation process.	3.080	3.205	2.000	3.632

The gender interactions for both years of the study are summarized in Table 28. The trait *the district has clear policy statements regarding the purpose(s) of evaluation* followed a similar pattern in mean scores as was observed in 2001-02. The combination of a male teacher with a female evaluator produced the lowest average mean score. In the description *please rate your overall experience related to your district's formal evaluation process*, male teachers gave the marks closest to an excellent process when combined with a female teacher, and farthest from excellent when paired with a male evaluator.

Table 28. Traits with a difference in response due to teacher gender and evaluator gender, in 2001-02 and 2002-03

Trait	Teacher (Evaluator)
The district has clear policy statements regarding the purpose(s) of evaluation.	Male (Female) lowest mean score 2001-02 (2.667) 2002-03 (3.000)
As it relates to my classroom, I consider myself relatively open to change.	Male (Male) lowest mean score 2001-02 (4.050)
Please rate your overall experience related to your district's formal evaluation process (1=excellent)	Male (Female) lowest (BEST) score 2002-03 (2.000)

In both years of the study, the interaction of a male teacher and a female evaluator produced the lowest mean score, in a statistically significant difference, related to the trait *the district has clear policy statements regarding the purpose(s) of evaluation.* In year one, the interaction of a male teacher and a male evaluator produced the lowest mean score with regard to the trait of *as it relates to my classroom, I consider myself relatively open to change.* In year two the interaction of a male teacher and a female evaluator produced the lowest mean score relative to the trait of *please rate your overall experience related to your district's formal evaluation process.* By way of reminder, rather than agreement/disagreement, this indicates excellence in experience when combining a male teacher and female evaluator. A combined model of comparison of means for gender differences was also completed as a part of the data analysis for the study. The results of those comparisons are contained in the next portion of this chapter.

Comparisons based on gender differences

The previous comparisons of means indicate significant differences that appear in a portion of the combined model. A comparison of the mean scores in the combined model reveals a number of traits that emerge as impacted by the interaction of the gender of the novice teachers and the evaluators. In 2001-02, there were five traits that demonstrated a significant difference in means within the combined model related to gender. The differences in means may be best demonstrated in a table format. The following five tables are intended to provide a numeric representation of the differences in means contained in the combined models of gender for the first year of the study.

As shown in Table 29, in regard to the trait, *The content of the district teaching standards and criteria was clear to me,* the mean score that is most different is the combination of a female teacher and a male evaluator. The mean is the lowest score in the table and is an indication that the trait is the least agreed upon by this combination of novice teachers and evaluators. It is also worth noting that the highest level of agreement within this trait is the combination of male teachers and female evaluators.

Table 29. Combined interaction for the trait, *The content of the district teaching standards and criteria was clear to me*, in 2001-02

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	3.64	3.83	3.68
Male	2.90	3.60	3.13
Combined Mean	3.18	3.65	

As shown in Table 30, the trait, *The district teaching standards and criteria were differentiated to meet my unique learning needs*, contains a mean score in a combination of a male teacher and a female evaluator that is the most discrepant from all of the other scores in the table. It is the most agreed upon in terms of the previously described combination of novice teacher and evaluator. The pairing that contains the least agreement with the trait is again the combination of female teachers and male evaluators.

Table 30. Combined interaction for the trait, *The district teaching standards and criteria were differentiated to meet my unique learning needs,* in 2001-02

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	3.20	3.83	3.32
Male	2.62	3.00	2.74
Combined Mean	2.84	3.19	

As shown in Table 31, the trait, *As it relates to my classroom, I consider myself relatively open to change,* is similar to the previous trait in that the combination of male teachers and female evaluators contains the mean score that is the most different from all of the other combinations. It is also the case that it is the most agreed upon combination of novice teacher and evaluator, by gender, in a trait that contains a significant amount of overall agreement. The combination of teacher and evaluator, which contains the least agreement, is the combination of male teachers and male evaluators.

Table 31. Combined interaction for the trait, *As it relates to my classroom, I consider myself relatively open to change,* in 2001-02

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	4.48	4.83	4.55
Male	4.29	4.05	4.21
Combined Mean	4.36	4.23	

As shown in Table 32, the trait, *My evaluator has the capacity to demonstrate or model needed improvements in my classroom,* is a bit different from other traits observed to this point in that the overall difference in means, when comparing the male evaluators with female evaluators, contains much more agreement related to female evaluators than male evaluators. This agreement is seen in both the female and male teachers, with the male teachers providing a higher average mean than the female teachers.

As shown in Table 33, the trait, *My evaluator is able to provide a persuasive rationale for suggestions for improvement,* is similar to the previous trait in that the overall mean scores for female evaluators are much higher than that for the male evaluators.

Table 32. Combined interaction for the trait, *My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s),* in 2001-02

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	3.84	4.17	3.90
Male	3.05	3.10	3.06
Combined Mean	3.34	3.35	

Table 33. Combined interaction for the trait, *My evaluator is able to provide a persuasive rationale for suggestions for improvement,* in 2001-02

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	4.04	4.00	4.03
Male	3.31	3.30	3.31
Combined Mean	3.58	3.46	

In the 2002-03 school year there are four traits that are observed as having significant differences within the combined model for the gender of the teacher and the gender of the evaluator. The observation that emerges by year two is related to the comment on the overall experience of the teachers related to the district's formal evaluation process. That is described as *please rate your overall experience related to your district's formal teacher evaluation process.* The two traits that were observed as containing a significant difference in 2001-02, but not in 2002-03 are *the content of the district teaching standards and criteria was clear to me,* and *the district teaching standards and criteria were differentiated to meet my unique learning needs.* The following four tables provide representations of the numeric comparisons contained in the combined model.

As shown in Table 34, the trait, *As it relates to my classroom, I consider myself relatively open to change*, contains overall agreement, as evidenced by a mean score over 4.0 in all gender combinations, by the novice teachers, regardless of the gender of the teacher or

Table 34. Combined interaction for the trait, *As it relates to my classroom, I consider myself relatively open to change,* in 2002-03

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	4.60	4.83	4.65
Male	4.32	4.05	4.24
Combined Mean	4.42	4.24	

the evaluator. That being said, there is a significant difference in the mean scores; with the combination of male teachers and male evaluators showing the least agreement based on the means. The combination of the gender of the novice teacher and gender of the evaluator with the most agreement is that of male teachers and female evaluators. As is seen in other traits, there is generally more agreement with this trait relative to female evaluators.

As shown in Table 35, the trait, *my evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s),* is similar to what was observed in the first year of the study. However, by the second year of the study, the difference in the overall means, as it relates to the gender of the evaluator, is not as pronounced as was observed in 2001-02. Only one of the combinations of the gender of the novice teacher and the evaluator is in the agreement range, with a mean score above 4.0; male teachers and female evaluators.

Table 35. Combined interaction for the trait, *My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s),* in 2002-03

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	3.80	4.17	3.87
Male	3.05	3.05	3.05
Combined Mean	3.32	3.32	

As is the case for many of the traits in the combined gender model, there is greater general agreement related to female evaluators as compared to the mean scores for male evaluators for the trait, *My evaluator is able to provide a persuasive rationale for suggestions for improvement.* This trait indicates the highest overall mean score average for the combination of female teachers and female evaluators and the lowest mean score for male teachers with male evaluators (Table 36).

Table 36. Combined interaction for the trait, *My evaluator is able to provide a persuasive rationale for suggestions for improvement,* in 2002-03

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	4.04	4.00	4.03
Male	3.45	3.37	3.43
Combined Mean	3.67	3.52	

This combined interaction related to gender as shown in Table 37 is the reverse indication of what was seen in the previous traits. To this point, a higher mean score is an indication of greater overall agreement with the trait. The set of data for the trait, *Please rate your overall experience related to your district's formal teacher evaluation process,* defines an "excellent" overall experience as a 1.0 and a "very poor" overall experience as 5.0. The discrepancy in mean scores between the combination of female evaluators and male teachers

Table 37. Combined interaction for the trait, *Please rate your overall experience related to your district's formal teacher evaluation process,* in 2002-03

	Teacher		
Evaluator	Female	Male	Combined Mean
Female	3.08	2.00	2.87
Male	3.20	3.63	3.33
Combined Mean	3.16	3.24	

($\underline{M}$ = 2.00) compared to the combination of male evaluators and male teachers ($\underline{M}$ = 3.63) is significant.

Table 38 provides a comparison of the observed traits in 2001-02 and 2002-03, as well as the gender interactions that exhibited the statistically significant differences in scores.

Table 38. Combined interactions for traits based on gender, in 2001-02 and 2002-03

Trait	Year	Gender Interaction Teacher (Evaluator) (mean score)
The content of the district teaching standards and criteria was clear to me.	2001-02	Female (Male) (2.90) Male (Female) (3.83)
The district teaching standards and criteria were differentiated to meet my unique learning needs.	2001-02	Female (Male) (2.62) Male (Female) (3.83)
As it relates to my classroom, I consider myself relatively open to change.	2001-02	Male (Male) (4.05) Male (Female) (4.83)
	2002-03	Male (Male) (4.05) Male (Female) (4.83)
My evaluator has the capacity to demonstrate or model needed improvements in my classroom.	2001-02	Female (Male) (3.05) Male (Female) (4.17)
	2002-03	Female (Male) (3.05) Male (Male) (3.05) Male (Female) (4.17)
My evaluator is able to provide a persuasive rationale for suggestions for improvement.	2001-02	Male (Male) (3.30) Female (Male) (3.31) Male (Female) (4.00) Female (Female) (4.04)
	2002-03	Male (Male) (3.37) Female (Female) (4.04)
Please rate your overall experience related to your district's formal teacher evaluation process (1=excellent)	2002-03	Male (Male) (3.63) Male (Female) (2.00)

In both 2001-02 and 2002-03,the three traits of, A*s it relates to my classroom, I consider myself relatively open to change, My evaluator has the capacity to demonstrate or model needed improvements in my classroom,* and *My evaluator is able to provide a persuasive rationale for suggestions for improvement,* showed significant differences in gender interactions. Generally speaking, the gender interactions with male evaluators and

male teachers demonstrate the lowest agreement. This is always the lowest agreement gender interaction when only the gender of the evaluator (male) is taken into consideration. In general the highest agreement level when considering traits occurs between male teachers and female evaluators.

Two of the traits, *The content of the district teaching standards and criteria was clear to me* and *The district teaching standards and criteria were differentiated to meet my unique learning needs,* demonstrated gender interactions that were gone by the second year of the study. In both traits, the pattern of male evaluator containing the least agreement continues, but the interaction occurs with the female teachers. The greatest agreement persists between male teachers and female evaluators. By the second year of the study, the trait, *Please rate your overall experience related to your district's formal teacher evaluation process (1=excellent),* emerges as containing a gender interaction difference. In this case, the average number is lowest in the male teacher and female evaluator combination, but it indicates the best overall experience in the teacher evaluation process. The interactions that create the least satisfying evaluation experience are with male evaluators, with the greatest dissatisfaction occurring in the interaction between male teachers and male evaluators. A summary of the findings comparing gender differences is found in Table 39.

Evidence exists for differences in the combined model for gender interactions between evaluators and teachers. The remaining portion of this study focused on differences in teacher perceptions from year one of the study to your two of the study. By way of reminder, this represents the year before the implementation of the IEATP and the year after the implementation of the evaluator training.

Table 39. Summary of findings when comparing gender differences, in 2001-02 and 2002-03

Trait	Gender Differences by year Teacher (Evaluator) (mean score)
The content of the district teaching standards and criteria was clear to me.	2001-02 Female (Male) (2.90) Male (Female) (3.83)
The district teaching standards and criteria were differentiated to meet my unique learning needs.	2001-02 Female (Male) (2.62) Male (Female) (3.83)
As it relates to my classroom, I consider myself relatively open to change.	2001-02 Male (Male) (4.05) Male (Female) (4.83) 2002-03 Male (Male) (4.05) Male (Female) (4.83)
My evaluator has the capacity to demonstrate or model needed improvements in my classroom.	2001-02 Female (Male) (3.05) Male (Female) (4.17) 2002-03 Female (Male) (3.05) Male (Male) (3.05) Male (Female) (4.17)
My evaluator is able to provide a persuasive rationale for suggestions for improvement.	2001-02 Male (Male) (3.30) Female (Male) (3.31) Male (Female) (4.00) Female (Female) (4.04) 2002-03 Male (Male) (3.37) Female (Female) (4.04)
Please rate your overall experience related to your district's formal teacher evaluation process. (1=excellent)	2002-03 Male (Male) (3.63) Male (Female) (2.00)

Changes in evaluator and self perception from 2001 to 2003

Research question three queried on differences that may exist in perceiver data from the 2001-02 school year to the 2002-03 school year (before and after the implementation of the ITS&C and the IEATP) around attributes of self as a teacher, perceptions of the evaluator, perceptions of the attributes of the evaluation processes, perceptions of the attributes of the feedback, and perceptions of the context in which the evaluations occurred.

A paired t-test was constructed for each of the pairs of the Likert-type questions in the survey. The pairs were defined as the comparison of responses for the same question from the 2001-02 and 2002-03 school years. Of the 39 comparisons, 13 showed significance at the .05 level or less. The Bonferroni test (.050 level of significance divided by the 39 comparisons) was applied to the t-test and a modified significance level of .00128 was defined for the comparisons. Of the 13 comparisons that met the original level of significance, only six were significant upon application of the Bonferroni calculation for significance.

The research question centers on the five clusters of traits that have received a significant amount of attention in the document. Of those five clusters, four contained at least one trait that was significant before the Bonferroni correction. The cluster of traits measured concerning the perceptions of the evaluator indicated no significance relationship to one another. Table 40 indicates the 13 traits that are significant from 2001-02 to 2002-03, as well as the trait cluster from which each trait emerges.

All the scores for mean difference shown in Table 40 are negative. This indicates that, for all of the question pairings that demonstrated some level of significance, a higher mean score was observed in year two of the study when compared to year one. It also denotes that for each of these mean scores there was some level of increased agreement in perception on the part of the novice teachers from 2001-02 to 2002-03.

The seven traits that are found to be significant before the Bonferroni adjustment would be better described as interesting rather than significant. While there is some

Table 40. Comparison of traits from 2001-02 to 2002-03

Trait	Trait Cluster	Mean Difference
The district teaching standards and criteria were communicated to me in detail.	Perceptions of the evaluation processes.	-.549*
The content of the district teaching standards and criteria was clear to me.	Perceptions of the evaluation processes.	-.574*
The district teaching standards and criteria were differentiated to meet my unique learning needs.	Perceptions of the evaluation processes.	-.210*
A review of classroom or school records (lesson plan, etc.) was used extensively in my evaluation(s).	Perceptions of the evaluation processes.	-.126
The feedback from my evaluator was focused on the district teaching standards and criteria.	Attributes of the feedback.	-.272
The district has clear policy statements regarding the purpose(s) of evaluation.	Attributes of the evaluation context.	-.087
The evaluation(s) is/are intended to place a high emphasis on teacher professional growth.	Attributes of the evaluation context.	-.118
In terms of my professional expectations, I demand a great deal from myself.	Attributes of self as a teacher.	-.099
In terms of my professional orientation, I consider myself an instructional "risk-taker".	Attributes of self as a teacher.	-.214*
I consider myself to be a teacher who frequently engages in instructional experimentation in the classroom.	Attributes of self as a teacher.	-.147
I have a great deal of knowledge about the technical aspects of teaching.	Attributes of self as a teacher.	-.267*
I have a great deal of knowledge about the subject matter I am responsible to teach.	Attributes of self as a teacher.	-.255*
I embraced the district teaching standards and criteria as appropriate for my classroom.	Attributes of self as a teacher.	-.136

$p < .05$ (2-tailed)
* significant with Bonferroni adjustment

indication that they could provide a suggestion for changes in perceptions by the novice teachers, they do not make the cut for true significance in difference from year one to year two of the study.

The six traits that meet the Bonferroni adjustment requirement are considered significant for the purposes of the study. It is also worth noting that the significant changes in mean scores from 2001-02 to 2002-03 are only located in the two trait clusters known as the perceptions of the evaluation processes and the attributes of self as a teacher.

The three statistically significant traits from the trait cluster of perceptions of the evaluation processes are: *The district teaching standards and criteria were communicated to me in detail, The content of the district teaching standards and criteria was clear to me,* and *The district teaching standards and criteria were differentiated to meet my unique learning needs.* While all of these traits increased in agreement from year one to year two, the trait, *The district teaching standards and criteria were differentiated to meet my unique learning needs,* went from disagree to neither agree nor disagree in the Likert-type scale from year one to year two. The other two traits were closer to agree in the Likert-type scale, especially by year two of the study.

The three statistically significant traits from the attributes of self as a teacher cluster include: *In terms of my professional orientation, I consider myself an instructional "risk-taker", I have a great deal of knowledge about the technical aspects of teaching,* and *I have a great deal of knowledge about the subject matter I am responsible to teach.* As was true with the other three traits, these traits increased in agreement from the first year of the study to the second. However, only the trait, *I have a great deal of knowledge about the subject matter I am responsible to teach,* demonstrated agreement in the Likert-type descriptive statistics. That was not the case for either of the other two traits, as both of them demonstrated neither agree nor disagree in the Likert-type scale.

Another area worth discussion, but in which no significant change occurred, was in the area of the number of formal and informal observations. The scores for the number of formal observations (observations that were pre-announced and followed by a conference with the evaluator) was also on a Likert-type scale with a 6 representing more than 4 formal observations and a 1 representing 0 formal observations. In 2001-02 the scores represented

between 2 and 3 formal observations (M=3.68). In 2002-03 the scores represented a similar number of visits (M=3.63). This may be partly driven by legal requirements for visits contained in the master contracts of many Iowa school districts.

The scores for the number of informal observations (refers to an unannounced drop-in visit that was accompanied by some type of written and/or oral feedback) follows a similar pattern to that observed for formal evaluations. The number of informal observations was also on a Likert-type scale related to frequency of visits with a 5 representing daily and a 1 being none. In 2001-02 the scores represented visits approximately once per month by evaluators (M=3.23). In 2002-03 the scores represented a similar frequency in visits (M=3.28). In both cases there was no significant change in observations, neither formal nor informal, from year one to year two of the study. The last set of information in this chapter centers on general comments from the open-ended questions.

Comments from Open-Ended Questions

Some topics emerged from teacher comments in the open-ended questions that were not originally addressed as a part of this study. While the evidence supporting them is not considered qualitative, some examples of representative teacher comments would lead to some general conclusions related to the topics. They may also be observed as topics worthy of further study. There were a total of two hundred and nineteen unique responses provided across the four open-ended questions. Some comments included more than one topic area for discussion, so the total for each of the areas may not match the original number of two hundred and nineteen.

There were 79 different comments made related to the Iowa Teaching Standards and Criteria. Of those comments, 57 had a more positive feel, 21 had a more negative feel and one felt more neutral. As this language implies, there is a sense in the comments that there is a positive feeling or a negative feeling related to the ITS&C. Some examples of each type of comment are included for review.

Positive comments related to the ITC&S include items such as:

"The Iowa Teaching Standards has made me stop and analyze my teaching. I have internalized what I do on a daily basis. REFLECTION!"

"I feel I am more clear of expectations and it is easier to set goals."

"It has brought a greater amount of awareness of the many aspects of teaching. Sometimes, we forget all of the little details that we need to focus on to be the best teachers we can be."

"It gave my evaluator and me specific items to discuss and evaluate."

"I feel like I am definitely monitoring my students as well as the way I am teaching to see how they are progressing."

Negative comments related to the ITC&S include items such as:

"It really hasn't [helped]. I feel like it has been more busy work than anything. The practices and strategies that we use are ones we do everyday. It seemed like a lot of extra paperwork and took all my extra time."

"None, it created more work for me. I was already doing my job."

"It has done little. It nearly pushed me out of the profession because how was I as a new teacher supposed to do all of these things when veteran teachers continually told me they did not do it all."

"I do not feel these standards have changed the quality of my teaching but just added to the realization of they want more and more from me with less time to do it in. Examples – grades, IEPS (graphs), staffings, teaching (Holy Moly I do that), extra programs after school, lesson plans, what more?"

"I knew nothing of the Teaching Standards the first year. They were all thrown on me the second."

The neutral comment related to the ITS&C, interesting in its own right, was:

"I was sort of an exception and was not required to fulfill the 8 standards/criteria."

Portfolios were widely discussed in the novice teacher comments. Of the comments, there were 21 total related to portfolios, generally centered on those who found the process had value (4comments), did not have value (12 comments), or simply made comment on the process (5 comments). A representative set of comments related to portfolios include:

"I think the evaluation process is very necessary and beneficial. As far as the professional portfolio is concerned, it seems as though it's just more work to add to our busy days… I understand the idea behind it, I just wish there was a different way."

"Allow new teachers one year with the new process before they are evaluated with a portfolio process."

"Completing my portfolio helped me to focus on areas within the eight standards in which I needed to improve."

There were 15 comments included related to mentoring. Of those, 8 were positive about mentors or the mentoring process, three were neither positive nor negative and four were negative towards mentoring. A few teacher comments related to mentoring include:

"My mentor turned this seemingly impossible task into a possible one."

"I am grateful for the opportunity to work with my mentor. she [sic] shared so much support and knowledge with me that I couldn't have made it without her."

"I really enjoyed working with my mentor and discussing many different educational topics with her. She was very helpful in my professional development."

Pay the first/second year teachers the extra money like the mentors. My mentor "made over two thousand dollars, and I was the one that had to do the majority of the work."

"I think this whole mentoring, two year program was a joke. It made a ton of extra work for a beginning teacher, who already has a hard time trying to keep up."

The topic of differentiation of expectations appeared as a less overt theme in the comments from novice teachers. A total of ten comments appeared and generally centered

around a desire to have all teachers held to these expectations, or a sense that the content area of the person was so unique that the process did not apply to his/her work. A couple of comments related to differentiation include:

> "Minimal impact because the standards were very broad and did not always apply to my area of teaching: early childhood special education."

> Evaluate teachers that have been in the field for several years, in my first three years of teaching I have seen teachers that are stagnant [sic] in their approach to teaching. They continue with the old strategies and give little effort to their teaching... I think older teachers have a lot of experience that they could bring to implementing the standards into their current techniques."

In novice teacher comments there was an indication of a desire to be evaluated using a variety of sources and to receive feedback from other sources. Those other potential evaluators and sources of feedback as defined in the open-ended questions are (a) people specially trained by districts to evaluate teachers, (b) teams, (c) several evaluators who are not associated with the district where the teacher teaches, (d) other teachers, (e) department reviews, (f) mentors, and (g) "someone else". The comments also indicated to not include student achievement data in the process (2 comments).

This study was designed to investigate the perceptions of novice teachers as they relate to teacher evaluation. The group of novice teachers chosen for the study were in the unique situation of having completed their first year of teaching in 2001-02, the year before the implementation of the Iowa Teaching Standards and Criteria and their second year of teaching was the first year of implementation for the ITS&C. The perceptions were gathered into five clusters of traits: attributes of self as a teacher, perceptions of their evaluator, perceptions of the evaluation processes, attributes of the feedback, and the context in which the evaluations occurred. In addition, data was collected based on the gender of the teacher

and the gender of the evaluator. A comparison was completed on potential changes in novice teacher perceptions related to the questions from 2001-02 to 2002-03. Finally, data was presented on comments from the open-ended questions.

Summary

This chapter provided a foundation for organization of the data as well an indication of the trends that are found relative to each of the research questions that were addressed through the survey. In addition, the chapter has broken down the research questions in an organized manner, such as trait clusters and/or gender of the teacher or evaluator. The next chapter will include an analysis of the trends that emerged from this set of data.

CHAPTER 5. CONCLUSIONS AND RECOMMENDATIONS

Summary

State legislated expectations for teacher evaluation around the then newly-constructed Iowa Teaching Standards and Criteria were enacted as a part of the Teacher Quality Program and Evaluator Approval Training. According to the legislation, Evaluator Approval Training was expected to: (a) improve the skills of school district evaluators in making employment decisions; (b) make recommendations for licensure; and (c) move teachers through a career path as established under this chapter. The Iowa Evaluator Approval Training Program Participant's Manual (Fall, 2002b) further refines the expectations for evaluators by providing the key development points for the training as: (a) consistent expectations statewide for quality instruction; (b) increased inter-rater reliability of evaluators for quality instruction; (c) identification and validation of the Iowa Teaching Standards and Criteria; and (d) a link to ongoing professional development based on definition of quality instruction.

The Evaluator training program was divided into three modules with knowledge and skill expectations for evaluators, upon completion, in each of the areas. Module one centered on participant knowledge in the intent and purpose of the Iowa Teaching Quality legislation, the Iowa Teaching Standards and Criteria, and best practices in teacher evaluation. Demonstration of ability for this module is centered on identifying teaching examples that support the Iowa Teaching standards and criteria.

Module two of the training focused on knowledge of Personnel Evaluation standards as they apply to data collection, best practices in data collection for teacher evaluation, and development and collection of multiple data. Participants would prove this knowledge

through collection and management of appropriate data that demonstrate support for the Iowa Teaching Standards and Criteria.

Module three knowledge spotlights the Personnel Evaluation Standards as they apply to feedback (bias, confidentiality, etc.) and best practices of feedback for the purpose of teacher evaluation. Trained evaluators would demonstrate the knowledge and skill in alignment of feedback to the Iowa Teaching Standards and Criteria. Participants will demonstrate the ability to identify best practices in teacher evaluation and provide constructive feedback to teachers relating to the Iowa Teaching Standards and Criteria.

The Evaluator Training would conclude with a site-based activity protocol. This protocol activity defined in the training manual, involves the: (a) identification of a teacher willing to allow the participant to practice the skills learned from the training; (b) identification of a standard and criteria to be addressed by the teacher and the evaluator through the protocol; (c) teacher and evaluator will design a question based on the training; (d) teacher and evaluator will agree on expected data to support the criteria; (e) the teacher and evaluator will work to compile the data; (f) teacher and evaluator will discuss the data and place it in a meaningful format during an informal conference; (g) teacher and evaluator will decide how well the question was answered using the conferencing guidelines from the training; (h) the evaluator completed the paperwork for the protocol; and (i) trainer and the evaluator will reflect on the process. In order to be evaluator approved, this protocol was to be completed in an acceptable manner as defined by the state expectations.

The novice teachers chosen for this study were in the unique position of having been evaluated in their first year of teaching by an evaluator who was not trained in the newly-designed Evaluator Approval Model. These same teachers were then evaluated in their

second year of teaching by the same evaluator, who had successfully completed the Iowa Evaluator Approval Training. This evaluator training completion is assumed in that no evaluator was allowed to evaluate in Iowa in 2002-03 unless she or he had successfully completed the Data-Driven Leadership and Iowa Evaluator Approval Training Program.

The study participants completed a seventy-eight question survey that was based an original teacher evaluation survey by Stiggins and Duke referred to as the Teacher Evaluation Profile (TEP) Questionnaire. In addition to the 78 Likert-type responses, teachers were invited to answer four open-ended questions at the end of the survey. While these responses were not intended to be a formal, qualitative portion of the research, they did provide additional insight into some of the results that were observed in the qualitative data.

For each of the non-demographic statements in the survey, a 5-point Likert-type scale, from strongly agree, 5 on the scale, to strongly disagree, 1 on the scale, was provided for review by the participants. The questions were paired such that respondents needed to provide reactions for the statements from the perspectives of both the 2001-02 school year (before the IEATP) and the 2002-03 school year (after the IEATP). The responses to the questions were designed to better understand novice teachers based on their perceptions, based on agreement/disagreement, of the five dimensions or traits, that are widely regarded as the most influential in teacher evaluation (Danielson, 1996; McGreal, 1983; Stiggins & Duke, 1988). The study was also designed to study demographic data in order to demonstrate if there is a gender effect in teacher evaluation from the perspective of the novice teacher. In addition, it was designed to identify differences, if any, in novice teacher perceptions of the dimensions of teacher evaluation from before (2001-02) and after (2002-03) the implementation of the knowledge and skills of evaluators based on Iowa Evaluator Approval

Training Program. The study addressed three research questions, for which the findings are discussed in this chapter.

Conclusions

Several conclusions are made based on the findings of the study.

Traits

Novice teachers

Research Question 1 focused on novice teachers reports concerning perceptions of attributes of self as a teacher, perceptions or their evaluator, perceptions of the evaluation processes, perceptions of the attributes of the feedback, and perceptions of the context in which the evaluations occurred. As it relates to novice teachers perception of self, the respondents from this study had high expectations for themselves, considered themselves to be open to change and constructive criticism, and believed that they had a strong knowledge of the content matter they teach. These are similar to the traits that emerged in the study of Iowa teachers completed by Lawler (1992) using a modified TEP. This optimism and enthusiasm, especially by the second year of teaching, was similar to findings observed by Lawler (1992) and Casey (1994). Knowledge of the subject matter, especially for those trained in secondary degrees with content-specific degrees, would be an expected comfort area for the group.

In the first year of teaching, the novice teachers did not perceive themselves as instructional risk-takers, when compared to the other traits, but that perception was not observed by the second year of the study. In the second year of the study, the teachers did not see themselves as possessing the trait of frequently engaging in instructional experimentation

when compared to the other traits. At first glance, this trait's emergence appears to be in contrast to the traits of open to change and open to constructive criticism. However, the trait was agreed upon by novice teachers, just not at the level as is observed in the most agreed upon traits. In both years of the study, the teachers did not perceive themselves to be as knowledgeable in the technical aspects of teaching, which would be anticipated based on the work of Armato (1990) and Peterson (2000) among others, in that veteran teachers are more creative and have a more broad repertoire of instructional skills and activities. This knowledge in the technical aspects of teaching is contrary to what was observed by Lawler (1992). It is worth noting that his study also included veteran teachers who would be more likely to have a sense for possessing a stronger instructional skill set. A logical outcome of teachers gaining experience in the classroom would be additional confidence in their ability to provide the students with more varied, high-quality instruction.

The novice teachers also did not perceive themselves as embracing the district standards in the classroom, when compared to other traits in the perceptions of self set of traits related to teacher evaluation. Given the training modules of the Iowa Evaluator Approval Training, and the expectations around implementing the Iowa Teaching Standards and Criteria, this point seems especially germane in that there would be an expectation in the second year that this would emerge as a trait of Iowa Teachers. This would assume that district standards are being communicated to the teachers. It is likely that even by the second year, in most districts, there was not an overt connection between district standards and the ITS&C.

None of the observed traits of novice teachers found in this study were particularly different from what would have been expected based on other research as well as on

interactions with beginning teachers. People new to the profession generally bring an expectation of hard work, and openness to change and constructive criticism to the workplace.

The trait not associated with novice teachers that gives a moment of pause is that, by year two, the novice teachers did not perceive themselves as frequently engaging in instructional experimentation. This may be connected to a later finding related to a lack of meaningful professional development provided by the district. Peterson (2000) noted that beginning teachers have specific professional development needs. There are indications from the survey that these unique needs are not being met. If not, perhaps a lack of willingness to engage in instructional experimentation flows from not having access to meaningful professional development models. It may also flow from the finding that evaluations are frequently unable to provide meaningful feedback or a persuasive rationale for change. Combining those novice teacher perceptions on not receiving meaningful professional development opportunities or meaningful feedback from evaluators, it is understandable how this trait emerges as has been observed. These results provide a sense about the traits of novice teachers based on their perceptions of self. The next section centers on the results of perceptions of the traits of evaluators.

Evaluators

Regarding their perceptions of evaluators, the respondents perceived their evaluators as flexible and trustworthy through the evaluation process. These are both characteristics that need to be in place in order for teacher evaluation to be successful, and are similar to what was observed by Lawler (1992). On the other hand, the trustworthy trait finding from this

study was in opposition to findings by Bulach and Peterson (1999) who found a lack of trust between teachers and principals. The finding is, however, in alignment with Bulach and Peterson's findings which indicated that teachers are willing to talk to principals about good things and open to comments and reactions from principals. The results indicated that evaluators in Iowa are working to form meaningful relationships with novice teachers, an important first step in the evaluator-novice teacher relationship (Peterson, 2000).

More troubling are the other results from the study in the dimension of teacher evaluation. In the first year of teaching the evaluators were observed as possessing a manner that is non-threatening. This trait was not observed by the second year of the study and appeared to run contrary to the notion of evaluators being observed as trustworthy. In looking more closely at the results for the non-threatening trait, nonthreatening was still an agreed upon trait by novice teachers, just not at the defined level for difference in the study. The trait of knowledgeable about the technical aspects of teaching was similar to nonthreatening in that it was generally agreed with in 2001-02 and in 2002-03. It was also a case that the other traits were more or less agreed upon and caused this trait to no longer meet the defined threshold as different from other traits.

Conversely, evaluators were seen as not being familiar with classrooms in their school(s). In addition, by year two of the study evaluators were seen as unable to demonstrate or model needed improvements in classroom practice. It was observed by Lawler (1992) that the lowest rating for evaluators were about modeling desired teaching behaviors and being able to provide a persuasive rationale to help teachers change. These findings appeared to be in line with observations by Tishler (1987) who found that 80% of first year teachers considered their evaluators to be good at what they do while only 32% of experienced

teachers felt the same way about their evaluators. In the current study, the first year teachers considered principals as knowledgeable about teaching, however, by year two, the principals were no longer perceived as knowledgeable about the technical aspects of teaching, or unable to demonstrate or model needed improvements in the classroom.

It is also worth noting that Armendt (2005) completed an Iowa study on principal perceptions of the implementation of the IEATP. Some of Arment's findings appear to demonstrate a disconnect between the perceptions of novice teachers toward their evaluators and the evaluators' perceptions of self. For example, 52% of the evaluators in Arment's study felt that no additional evaluator training was needed beyond the IEATP, and 48% felt that no additional training on identifying teacher effectiveness was necessary beyond IEATP. Evaluators felt that teachers needed additional training in teacher effectiveness (67%) and evaluation as it relates to the ITS&C (63%).

Given that goals of the IEATP include identifying best practices in teacher evaluation and providing constructive feedback on the Iowa Teaching Standards and Criteria, which include planning and instruction, it is troubling that these trait statements emerge as not being traits of Iowa evaluators. It is also troubling, but not surprising, to learn that novice teachers are generally more in agreement with traits of self than with the traits of evaluators. The same could also be said for evaluators when their perceptions of teachers and evaluators are compared. There appears to be some truth to the thought that, in the end, it is all about the individual. The findings in this study related to novice teacher perceptions have implications for future participants in the IEATP in that future participants will need to demonstrate competence in the eyes of the novice teacher and they will need to establish a meaningful coaching relationship early in the evaluation process.

Evaluation process

Perceptions

While there are fewer traits for comparison in this grouping of traits, it is clear that that the trait related to differentiating the process to meet the unique needs of individual teachers is different from all of the other traits in both 2001-02 and 2002-03. This information provides an indication that novice teachers perceive that differentiation of the district teaching standards and criteria is not occurring, especially as it relates to other traits of the evaluation process. The results appear to be in sharp contrast to a call for the need to provide, at minimum, different approaches to assessment for beginning teachers (Barber, 1985; Danielson, 2001; Glatthorn, 1997; Glickman, 2002a; McGreal, 1983; Pajak, 2003).

The same expectation for differentiation appears true based on the expectation in module three for the on-site protocol. In that protocol there is practice between the evaluator and the teacher in identifying the data that will demonstrate the criteria to be observed, presenting it in a meaningful way, and deciding if the data support the chosen criteria. It is somewhat predictable that the lack of differentiation is perceived to be occurring. When considering the work of Cuban (1988) which focused on compliance rather than outcomes, it is reasonable that in many schools evaluators met the expectation of the training first. That is to say, in first learning a new task it is likely there was a focus by evaluators on implementing what had been learned rather than not only implementing, but also modifying for individual teacher needs. In addition, adult learning theory would remind educators that coaching and follow-up support are needed to make sure that learning has transferred to daily practice (Knowles, 1973; Leypoldt, 1967; Speck, 1996). It is likely evaluators did not receive

that level of feedback for implementation from their supervisors or IEATP coaches, and the ability to differentiate teacher evaluation would be impacted as a result.

In addition to the need for differentiated evaluation procedures, there is an expectation that evaluation should include feedback from multiple sources (Conley, 1987; Danielson, 2001; McGreal, 1983). In the current study, by 2002-03, a review of classroom records did not emerge as a trait of the evaluation processes. This would indicate that classroom or school records are not necessarily as an extensive portion of the evaluation process. This perceived lack of use of multiple data sources and data points can likely be extended to other sources of feedback and data beyond the example of lesson plans. The same expectation for multiple data sources from the literature is evidenced in Iowa Code which requires feedback to teachers come from a number of sources including students, parents, teachers and other evaluators. It is also exhibited in the expectation of the on-site protocol experience for evaluators, with feedback from the IEATP trainers. By not emerging as a trait in year two, it creates a question related to the implementation of the training as well as expectations by districts for evidence of meeting the ITS&C. Certainly, the licensure activity, from initial license to standard license, would be impacted by a lack of use of these types of artifacts.

The traits relating to communicating the district teaching standards and criteria as well as the contents of the standards and criteria being clear, both reached the descriptor of agree on the Likert-type range by the second year of the study and were also agreed upon traits of the evaluation processes. This would likely be an anticipated outcome from the evaluator approval training. One of the training pieces in module one was a focus on the Iowa Teaching Standards, which was also to be applied to the evaluation process at the

district level. An indication that this is an agreed upon trait indicates that there is improvement in communicating these expectations to teachers from 2001-02 to 2002-03. It is also likely that there is a clarification of what the standards and criteria actually are with the advent of the ITS& in place by 2002-03. This also further supports the notion that compliance is occurring. One of the seemingly easiest points of implementation would be to communicate the new ITS&C, and this appears to be occurring based on the statistically significant increase in change from year one to year two of the study. In addition, that increase is in the direction of further agreement with the observed trait.

Attributes of evaluation feedback

Only one of the traits of the attributes of the feedback was significantly different from the other traits; however, it was significant in all of the traits in both 2001-02 and 2002-03. The agreed upon trait for this portion of the study related to the idea that the information provided by evaluators was descriptive rather than judgmental. This is a trait worth noting, in that part of the emphasis of the Iowa Evaluator Approval Training focuses on the need for the use of evidence and data in the evaluation process rather than personal judgments. It is interesting that, by appearing in both of the years, there may be a question as to how much of a change, if any, this represented in the evaluation habits of many of the evaluators in the state as a result of Evaluator Approval Training.

One of the most important elements of instructional leadership is the role of the evaluator (De Bevoise, 1984; Frase & Streshly, 2000; Leithwood, 1992). Within that role, providing feedback to promote reflection is one of the best indicators of instructional leadership (Blase & Blase, 1999a). It is also repeatedly viewed as one of the essential

elements of effective teacher evaluation (Duke & Stiggins, 1988; Glatthorn, 1996; Glickman, 2002a; McGreal, 1983) and was a point of emphasis in the IEATP, so much so that the training manual spent an entire module on ORID questioning (Nelson, 2001), and this was an activity that was completed back in schools with IEATP trainer feedback. Given all of that information, it would have been anticipated that there would be a significant increase in agreement to the statements related to feedback.

The observed means for none of the other traits in this dimension of teacher evaluation reach the agree level in the Likert-type rangeThis indicates that the novice teachers did not agree with any of the other traits of feedback in the group which include (a) feedback novice teachers received from their evaluator was focused on district teaching standards and criteria, (b) feedback received from their evaluator was directly applicable to the classroom, (c) feedback received was of high quality, and (d) feedback was very specific.

No significant changes in perceptions of evaluator feedback beyond feedback being descriptive rather than judgmental were observed in this study. This lack of observed change in the types and quality of feedback for teachers may be a casualty of a lack of repeated coaching feedback for evaluators in building skills for quality implementation at the building level.

This appears to be an area for further discussion as it relates to the conclusions contained in this document.

Context in which the evaluations occurred

In 2001-02, the novice teachers considered the context in which the evaluations occurred as being intended to both place a high emphasis on teacher professional growth and

teacher accountability. By 2002-03 the mean scores reinforce those traits as different from the other traits in the cluster and that the district has clear policy statements regarding the purpose(s) of evaluation to the group. The findings may indicate that the evaluation models are trying both to hold teachers accountable, and lead to professional growth. This is one of the great challenges of teacher evaluation frameworks, and it is generally regarded as being achievable through a differentiated model based on teacher career levels (Danielson, 1996; Danielson & McGreal, 2000; McGreal, 1983) or teacher skill level acquisition (Glatthorn, 1997; Glickman, 2002a). The state of Iowa appears to have a similar intention in combining a model based on accountability and professional growth. This is evidenced by expectations in the IEATP, backed by legislation, that evaluators are to both transition licenses of new-to-the-profession teachers, a high accountability activity, and design three-performance review models for career teachers, an activity which emphasizes teacher professional growth.

In 2001-02 and 2002-03, the perceptions of novice teachers indicate that an inadequate amount of time was allotted during the teaching day for professional development. In 2002-03 a second trait, centered on districts making models of best practices available to teachers, emerged as a least agreed upon trait. This is in line with what was observed by Peterson (2000) that one of the specific needs of beginning teachers is additional professional development. Specifically, the professional growth must address first familiarity with the district, second gaps in teacher training, and third classroom discipline. It is also troubling in that one of the two most important traits that designate a principal as an educational leader based on teacher perceptions is promoting professional growth (Blase & Blase, 1999a).

It seems intuitive that if novice teachers perceived that they were not being provided appropriate time for staff development that the models provided to them would be considered to be somewhat limited in scope. If Peterson's assertions are correct and the district is not differentiating the professional development, but rather providing the same to all, beginning teachers may not be perceiving the activities as appropriate because they are being provided training beyond their experience/competence, rather than meeting the more basic needs of familiarity with the district or classroom discipline. It may also be a situation of increased awareness on the part of the novice teachers, in that as they become more aware of potential training programs/models, they are less agreeable to this trait amongst the collection of traits used for evaluation. This is also important as it relates to IEATP, as the Training Manual notes that a key development point for the training is to link teacher evaluation to ongoing professional development based on the definition of quality instruction.

An additional finding related to the context of teacher evaluation was the emergence of clear policy statements regarding the purposes of evaluation. In alignment with compliance in communicating district criteria for evaluation, a likely extension of the conversations would be communication around the purposes of evaluation.

Clearly, some positives appear to have emerged related to novice teacher perceptions around the dimensions of teacher evaluation. It is also clear that there appear to be gaps in delivery of teacher evaluation based on the IEATP model and the implementation of the system in districts across the state. This will be addressed further in the conclusions and suggestions for next steps contained later in this chapter. The next section takes a look at potential differences in perceptions of evaluation and evaluators based on teacher and evaluator gender.

In transitioning to the gender portion of this study, it is worth noting that many types of bias are acknowledged in the research as well as in the IEATP participant manual. The research that has been discussed regarding bias includes topics such as race, ability to write a quality anticipatory set, and gender, to name a few. In the IEATP participant manual bias is described as a source of concern as a part of the second module of study. The second research question included in this study specifically deals with the issue of gender of the teacher, gender of the evaluator, and possible interactions that may occur as a result of the gender of the teacher and the gender of the evaluator.

Gender-related conclusions

Teacher gender and evaluation

Of all of the surveyed questions in this study, only one trait emerged as statistically significant in difference of response based on the gender of the teacher.

The only trait in which there is a gender difference is related to knowledge of subject matter and it does not appear by the second year of the study. In general, females scored the answer with less agreement than did males. That may have had an impact on the outcome of this portion of the study. An additional thought may relate to training of the teachers. Males are predominantly represented in teaching secondary schooling (US Dept. of Ed., 2000), which generally requires a degree or minor in the content area. Conversely, females are more commonly represented in elementary teaching (US Dept. of Ed., 2000), which generally does not require a content major. As a result, it may lead to a difference in perception of content knowledge in the first year of teaching that has resolved itself by the second year of teaching when there is additional comfort after having gone through all of the content material in the

previous (first) school year. The data appear to indicate that, in general, there are very limited differences in novice teacher perceptions based on teacher gender.

Evaluator gender and evaluation

Unlike what was seen in terms of teacher gender, there do appear to be differences in teacher perception based on the gender of the evaluator. There is generally a statistically significantly higher score, for female evaluators than male evaluators for a number of different traits. The differences between perceptions of female and male evaluators are many, and actually increase in number from the first year of the study to the second. This study indicates that novice teachers perceive female evaluators as having the capacity to demonstrate or model needed improvements in the classroom and as having a persuasive rationale for suggestions for improvement, as demonstrated by general agreement in the Likert-type averages. While not necessarily demonstrating agreement based on the Likert-type range, female teachers tend to have more agreement than males in the areas of differentiating standards and criteria based on the needs of the teacher; by 2002-03, they make many training programs/models of best practices available to teachers and they are familiar with classrooms in general in the school. In addition, teachers have more agreement with the statement about being open to change in the classroom with a female administrator than with a male administrator. A difference was observed in a focus on professional growth in 2001-02, but that was not significant by the second year of the study.

Perhaps the most telling piece of information of all is the indication of overall experience with the district evaluation process. Using a scale from 1, defined as excellent, to 5, defined as very poor, the scores for male evaluators were higher than those for female

evaluators. More specifically, the scores for female evaluators were closer to excellent, and the scores for male evaluators were closer to very poor. Even more enlightening is the fact that in year two of the study, after IEATP, the scores for female evaluators were even lower, or closer to excellent, and the scores for male evaluators were even higher, or closer to very poor.

These findings appear to fall in line with the work of many researchers (Cunningham, 2004; Gougeon & Hutton 1993; Rinehart & Young, 1996) that there is a gender preference as it relates to evaluators. This study shows that it tilts toward female evaluators, and that, too, is in line with other researcher's findings. It is in direct opposition to the findings by Cioci (1991) that male evaluators rate female evaluators lower than do female teachers; however, that is not the prevailing finding when compared to the results from this study as well as the work of other researchers.

A particularly important finding is the idea from both Rinehart & Young (1996) and Cunningham (2004) that female evaluators are generally regarded, more so than males, as instructional leaders. An additional finding, likely the most important of all as it relates to evaluator gender, is the finding from Gougeon & Hutton (1993) that teachers, regardless of gender, view female principal's as more positive communicators. One of the main points of emphasis in the IEATP is an emphasis on communication with teachers – for the criteria for evaluation, for choosing data, for making instructional decisions, etc. It is clear from this study that there is a difference around many topics related to the gender of the evaluator, and they favor the work done by female evaluators.

The next set of data examination centers on the interaction of gender of the evaluator and gender of the teacher. While there aren't nearly as many sources of difference as

observed when looking strictly at evaluator gender, it is clear that differences exist in the interaction of the gender of both teacher and evaluator.

Influence of gender based on gender of the novice teacher and the evaluator

Based on the results of this study, there does appear to be a gender interaction based on both the gender of the evaluator and the gender of the teacher. The number of statistically significant differences, based on this interaction, is not as pronounced as was observed in a focus on the gender of the evaluator alone, but an interaction appears nonetheless.

The results of the gender interactions in this teacher-evaluator combination are mixed when compared to what was observed in a strict focus based only on the gender of the evaluator. In this portion of the study the lowest-scoring interaction in both years was that of male teachers with female evaluators around the trait of the district having clear policy statements. For the first year of the study the lowest score was an interaction between male teacher and male evaluators on being relatively open to change, but this was not significant in year two. In the second year of the study an interaction between male teachers and female evaluators, described as an excellent overall experience in the teacher evaluation process, demonstrated the best overall experience in teacher evaluation found in the study.

Comparison of gender differences

A potentially more important set of findings relate to a combined model that compares the mean scores of the gender of evaluators and gender of teachers. A set of statistically significant findings exists as it relates to the gender interaction found in a comparison of means for teachers and evaluators based on gender.

In the majority of these gender interactions the lowest score (least agreement with the trait) is between male teachers and male evaluators. Conversely, the highest level of agreement with the traits is observed in the interactions between male teachers and female evaluators.

The interaction between female teachers and male evaluators and female teachers and female evaluators, in some cases, is similar in mean score to other gender interactions. Generally, however, the average scores for those two types of gender interactions fall somewhere between those observed in the interactions with male teachers and either gender of evaluators. As an additional dynamic in gender interaction, it was observed female teachers, overall, tend to score all evaluators lower than do male teachers. This finding related to the manner in which female teachers tend to score evaluators, may at least partially explain the means in the middle of the scores for male evaluators. It does not explain how male evaluators did the least well in terms of agreement with the male teachers and the female teachers do the most well with male teachers. Clearly there is some type of gender interaction in perceptions of teacher evaluation. This is opposite of the results observed by Cioci (1991) indicating that male teachers tended to rate female teachers lower than did female teachers. Perhaps times have changed such that the tradition has passed of male evaluators being observed as effective and female evaluators as less effective.

The most telling gender interaction may be the interaction of male teachers and male evaluators and male teachers and female evaluators relative to the perception of an overall rating of the teacher evaluation experience. The mean score difference is quite pronounced between the male teacher and male evaluator interaction and the male teacher and female evaluator interaction. The implication is that the experience of male teachers working with

female evaluators is significantly better (closer to "excellent") than the experience of male teachers working with male evaluators.

One cautionary note in these findings is observed when looking at cell populations in the various types of gender interactions. One of the least common interactions is that of a male teacher with a female evaluator. As a result, the population for this interaction tends to be smaller and that has an impact on means scores that are produced for comparison to the mean scores used for comparison in more heavily populated cell comparisons. An example of a more common interaction is that of female teachers with male evaluators. This does not explain away the gender interaction, and there is still likely power in these findings. It is worth noting that other important interactions may not have been included in this study. This would be underscored by the lack of sphericity that has been observed as well as the possibility of other interactions such as race (Brown, 2005), career stage of the evaluator (Peterson, 1988), or subject area being observed (Ward & Sistrunk, 1988). More specific thoughts related to gender interaction findings will be addressed in the final section of the study. It appears that there is some alignment between changes in evaluator and self-perception and the gender interaction. That is to say, the gender of the evaluator, in combination with the gender of the teacher, impacts the perceived outcome of teacher evaluations.

Summary

Based on information provided by novice teacher perceptions, there is a gender influence in teacher evaluation. This difference is most pronounced as it relates to the interaction of gender of the evaluator and gender of the teacher. More specifically, female evaluators are better regarded for the feedback they provide, communicating standards and criteria, and modeling needed improvements, among others. Also, the most frequent positive (more agreement) interactions are in the gender combination of male teachers and female evaluators. The least agreement with traits is observed most often in the interaction between male teachers and male evaluators. A related concern is the statistically significant difference in male teacher perceptions of the overall evaluation process – female evaluators score much closer to excellent and by the second year of teaching were closer to excellent, and the gap in scores between male and female evaluators increased.

Female evaluators likely are better communicators and better relationship-builders than are their male counterparts (Gougeon & Hutton, 1993). They also are likely to be perceived as better teachers due to their ability to display more instructional leadership behaviors (Cunningham, 2004). These perceptions are also supported in these research findings. The strength in communication likely leads to the improved ability to present the standards in a clear manner and to provide a persuasive rationale for suggestions for improvement. The ability to build meaningful relationships with staff members will likely lead to the openness to change and perception that evaluation is differentiated based on teacher needs. A background in teaching at the elementary level, which is emphasizes pedagogy and classroom instruction, supports the perceptions of the ability to model

improvements in the classroom and reinforces the position in providing suggestions for improvement. This is based on the idea that many of the female evaluators in this study would be serving in administrative positions at the elementary level. The combination of these traits leads to a significant difference in teacher experiences of the evaluation process. Specifically, female evaluators provide a teacher evaluation experience that novice teachers, especially male novice teachers, consider to be closer to excellent. The final analysis of this document centers on changes from before and after implementation of the IEATP training.

Changes Before and After IEATP (2001 to 2003)

Of the 40 items that were observed in the survey, 13 were significant in a change from 2001-02 to 2002-03, and in all of those cases the change was an increase in agreement from the first year of the study to the second. Of those 13 traits, 7 of them were observed only before the Bonferroni correction. What may be most interesting about these findings is that there are significant changes in each of the trait cluster areas except attributes of the evaluator. With the training and expected implementation of IEATP, one would presume that the strongest changes from the first year of the study to the second year would be in the trait cluster related to evaluators. However, no traits changed significantly over the course of this study.

More important than these results are the findings that were significant following the Bonferroni correction. It can be said that for the remaining 6 traits there was a significant change from 2001-02 to 2002-03. Each of those 6 traits comes from either the trait cluster of perceptions of the evaluation processes, with 3 traits, or attributes of self as a teacher, also with 3 traits. The changes in traits about the evaluation processes would be expected based

on the IEATP. Novice teachers perceived that the district teaching standards and criteria were communicated better and were clearer, or more understandable. There also was movement in the area of differentiating the process to meet the unique learning needs of individual teachers. This was demonstrated in a statistically significant increase in agreement with this trait.

Teacher evaluation in Iowa did improve as a result of the IEATP. There were 13 areas that showed initial significant change in the study, and all of them were negative in results. That is to say, there was more agreement with the statements in 2002-03 after implementation of the IEATP tenets, than was observed in 2001-02, before the training program was implemented. Of those 13 traits, 6 were observed as significant even after the Bonferroni correction, and, again, they were more agreed upon by the second year of the study. The areas in which there was more agreement by year two are the areas of evaluation processes and teacher attributes. These are two areas where an impact expected would be expected due to changes in the teacher evaluation process in Iowa.

Conversely, one of the interesting conclusions from this research is the lack of significant results in the area of feedback to teachers. This is especially troubling in the context of the expectations placed on administrators, especially related to the importance of meaningful feedback, as a part of the ITS&C and the modules of the IEATP; specifically the significant work around ORID questions and postobservation conferences. Admittedly there was statistically significant improvement in feedback as it relates to a focus from the evaluator on the district teaching standards and criteria, but the improvement was not significant following the Bonferroni correction. Also, the trait relating to information from the evaluator as descriptive was the only trait that reached the level of "agree" on the Likert-

type range by 2002-03. The phrases relating to feedback in which there is not statistically significant improvement, nor Likert-type agreement by 2002-03, are related to receiving a great deal of feedback that is directly applicable to the my classroom, ideas and suggestions from evaluator were of high quality, information provided was very specific, and to a lesser degree, the feedback was focused on district teaching standards and criteria. These phrases represent the type of training upon which the IEATP was centered related to feedback. By not being observed as changing in statistical significance, or agreement, this would appear to be a weakness in the implementation of the IEATP.

Another related thought is the lack of appearance of any level of significant change from 2001-02 to 2002-03, before or after the Bonferroni adjustment, in the trait cluster of perceptions of the evaluator. While there is agreement with 4 of the 11 evaluator traits by the second year of the study, these perceptions of evaluator behavior would be expected to change through the evaluator training. Additionally, with a focus in the modules on evaluator knowledge and skills related to identifying best practices in teacher evaluation, conversations in data collection and interpretation and the ITS&C, as well as extending it to professional development, one would have anticipated significant changes in novice teacher perceptions by 2002-03 around evaluator traits such as knowledge about the technical aspects of teaching, suggestions are useful for professional improvement, and is a credible source of meaningful feedback. None of that was observed in this study.

As a follow-up to the notion of compliance to expectations of program outcomes, and adult learning theory that advocates for coaching and follow-up supports to transfer learning, it may be understandable that change was not observed in traits of evaluators or feedback. To go back a step, it is often observed that new learning goes from knowledge, to application, to

implementation. Much of what is observed in the changes centers on the knowledge and compliance aspects of the training, as they are the most easily and immediately implemented back at the school. This has sometimes been referred to and may be explained by the notion of an implementation dip (Fullan, 2002) following real and perceived change. The real change in collective knowledge and commitments will take time and practice (Fullan, 2002). In addition, the observed findings are partially supported in Amendt's (2005) study, as the perceptions of evaluators following the training is that teachers could use additional training in teacher evaluation more so than the evaluators. A final thought related to a dearth of change in evaluators from before the training to after flows out of adult learning theory. One of the criteria for success in training adult learners is that the objectives need to be realistic and important to the learners. The IEATP was mandated by the state rather than pursued by evaluators in the state, which may play a part in changes not being observed in evaluator traits. These findings, as well as others from the study, will have an impact on recommendations for revisions to the first round of IEATP as well as on the IEATP renewal process.

Recommendations for the IEATP Renewal Process

The intent of the legislation as it relates to the IEATP indicates that under the program certified evaluators will (a) understand theory behind best practices for teacher evaluation, (b) demonstrate ability to provide data-based leadership, (c) be able to identify quality instruction in the classroom, (d) validate effective teaching that supports the ITS&C, and (e) provide coaching in a professional growth environment. Suggestions for changes to the IEATP, based on the novice teacher perceptions contained in this study, will be framed around these five broad goals for training.

1. Understand theory behind best practices of teacher evaluation

Evaluators need to be reminded of the five broad clusters of traits related to teacher evaluation (as well as other specifics such as legal expectations and nuances between formative and summative evaluation) and novice teacher perceptions around each of the clusters. This will help evaluators better understand how novice teachers perceive themselves, their evaluators, and the other dimensions of teacher evaluation.

There is already a portion of the training that speaks to bias in evaluation. That conversation is generally focused on the bias of evaluators. It may be worth extending that conversation to include the gender interaction between teachers and evaluators and the bias of female teachers to providing lower overall agreement to the elements of evaluation. Not only are evaluators evaluating and providing feedback; the interaction between the gender of the teacher and the evaluator impacts how the evaluation comments and evaluator behaviors are being interpreted by the novice teachers.

2. Demonstrate ability to provide data-based leadership

There may not be as obvious a change needed in this portion of the IEATP, when compared to the other portions. Novice teachers perceive, following the evaluator training, that district teaching standards and criteria are communicated in detail and that they are clear to novice teachers. The next step of improving pre-observation conferences to make the connections between the district teaching standards and data collection more obvious seems like a natural, logical next step in training evaluators. This also would be supported in the limited change that was observed in formal and informal observations from the first year of the study to the second. It is difficult to collect data without direct observation of teaching, and it has been observed (Brown, 2004) that increases in informal observations and feedback

improve the relationship with evaluators and make it more likely individual teachers will remain in the profession.

3. Be able to identify quality instruction in the classroom

While training evaluators on the theory behind quality instruction, there needs to be an emphasis on training them to identify specific behaviors that are a part of quality instruction. This is currently contained as a knowledge and skill expectation in training Module One. In addition, there needs to real world application to the learning elements to meet the needs of adult learning theory (Knowles, 1973; Leypoldt, 1967; Speck, 1996).

4. Validate effective teaching that supports the ITS&C

Ongoing practice needs to be a part of the training in identifying effective teaching strategies. There needs to be fewer novice teacher comments about room cleanliness and how high the shades are, and more comments related to a scientific approach to teaching. Iowa evaluators need continued training, including those focused on research-based strategies, when implemented with fidelity, that are regarded as leading to increased student achievement. This knowledge not only will assist evaluators in improving instruction, it likely will lead to increased agreement of perceptions by novice teachers that evaluators are knowledgeable about the technical aspects of teaching and able to demonstrate needed improvements in the classroom. It also will support evaluator ability to provide persuasive rationale for changes to teacher practices.

5. Provide coaching in a professional growth environment

IEATP currently focuses coaching conversations on the use of ORID questions. This study indicates that the feedback being provided by evaluators is not focused on district teaching standards and criteria, not specific, or of high quality. Ongoing evaluator training

needs to provide for a coaching model focused on high-quality questions and, at the same time, be able to allow for high quality feedback.

The training also may improve through an increased focus on the connection between professional growth and teacher evaluation. While novice teachers perceive that evaluations are intended to place a high emphasis on professional growth (and accountability), they also indicate neither that adequate time is allotted in the school day for professional growth nor that many training programs/models of best practices are made available by districts.

Some additional thoughts related to the training of evaluators would include an emphasis on skills and differentiation within the training. Just as is observed in skill differences in needs of teachers, whether it is based on years of experience or skill level obtained, a similar approach needs to be taken in training Iowa evaluators. For example, several comments from novice teachers indicated being pleased with the evaluation process as completed by an Iowa evaluator. These comments, when compared to other comments in this study, indicate that Iowa evaluators vary in implementation and could use differentiated, ongoing training:

"My evaluator, [evaluator name], utilized this process as a teaching experience. I received more from my time with him than in any of my classes at UNI."

"Evaluation procedure was very nonthreatening and beneficial to me. I credit my principal with this."

My [sic] evaluator does a fantastic job of describing positives he witnessed in my classroom, made me feel real good about working under him."

It could also be argued, at some level, that there could be differentiated trainings for evaluators based on evaluator gender. Male evaluator training may include additional work in communication techniques related to the content of district teaching standards and criteria as

well as in providing a rationale for changes in classroom practices and behaviors. It also may include additional training in differentiation of evaluation and in knowledge/skills related to demonstrate or model needed improvements in teacher classrooms. This training change is training with an emphasis on males is unlikely to occur given that there are continuums of ability to provide meaningful feedback across both genders of evaluators. It is also unlikely to occur given the impractical nature of singling out training needs based on the gender of evaluators.

Another dimension of this work that would likely prove valuable would be to provide a training program for evaluators of evaluators for the ongoing professional growth of Iowa evaluators. Frequent comments from novice teachers focused on who was holding the evaluators accountable to implementing this state expectation. Just as teacher evaluation often has the double track of accountability and professional growth, there is a perception by novice teachers that something similar needs to be in place for evaluators. It would be valuable to design a research-based model to evaluate and professionally grow Iowa evaluators in the area of implementation of teacher evaluation, among other topics. Adult learning theory supports this idea in the context of the need for ongoing coaching for improved implementation of training. It is also supported in the Fullan work that indicates that real change in evaluator behaviors while in the workplace.

This next stage of training will need to be crafted carefully in order to get buy-in from both evaluators and evaluators of evaluators as to the necessity of such work. Otherwise, we may see an extension of thought similar to what was observed by Armendt (2005) when he discovered that only 41% of evaluators he surveyed felt that student achievement in Iowa would increase as a result of the implementation of the IEATP and 20% of that same group

of evaluators expected no improvement in student achievement in Iowa as a result of the implementation of the training. We likely also would see limited coaching back in districts.

Areas for Further Study

This study was focused on the perception of novice teachers; similar work done by Lawler (1992) used feedback from career teachers. A similar study using feedback from career teachers would be of value, especially with the recent expectations of applying the lessons and teachings of this training to career teachers. Another potential study would connect student achievement to the results of the implementation of the evaluator feedback following the IEATP. A follow-up study using the novice teachers from this study, after four years of implementation of the ITS&C, may provide for additional feedback on the implementation of the training.

This study was limited to a focus on gender. Additional studies on other potential sources of bias in teacher evaluation would have value. Some of the other sources of bias that would be worth studying include the career stage of the teacher and/or evaluator, race of the teacher and/or evaluator, grade level cluster of the teacher and/or evaluator, or the subject area taught by the teacher.

Feedback from evaluators was also used by Lawler (1992) in his study on evaluator training. A similar study using the perceptions of evaluators from the modified TEP survey likely would have merit and would be a companion to the work of Amendt (2005). By comparing the perceptions of evaluators to the perceptions of teachers on the same instrument the differences in perceptions likely would be more immediately obvious than through a comparison of studies that did not use the same survey instrument.

Leadership styles should be used as a foundation for a follow-up survey as well as teacher learning styles. There may be an interaction in that relationship between leadership style and teacher learning styles that impacted this model, but were not accounted for in the modified TEP.

An additional follow-up study would center on the teacher who have left the profession since the implementation of the IEATP. The group of teachers who participated in this study represent those educators who stayed with the teaching profession. A study that includes teachers that left the profession would potentially provide additional feedback on the IEATP.

The topics that emerged from the open-ended comments are likely worth further study. They are the ITS&C, teacher professional portfolios, mentoring, differentiation, and other evaluators and potential sources of feedback. The topics may be studied through the use of surveys, or through follow-up surveys with the novice teachers who were surveyed for this study.

APPENDIX A. PARTICIPANT INVITATION LETTER

May 15, 2004

Dear _____

I am a doctoral candidate at Iowa State University as well as the principal at Waukee Middle School. My dissertation topic centers on novice teachers' perceptions toward their district evaluation procedures, their personal teaching attributes, the frequency of evaluations and the perceptions of their evaluators. The intention is to collect data on the perceptions of teachers who completed their second year of teaching under the Iowa Teaching Standards and Criteria in 2002-03. This study will compare their perceptions of their experiences with the teacher evaluation process during both the 2001-02 and 2002-03 school years.

I am asking for your help in improving the state initiatives in the areas of teacher mentoring and induction, standards and criteria for teaching, and teacher evaluation. **Your responses may have an impact on any policy changes that may be made in improving these areas in upcoming legislative sessions.**

I received your name through the Iowa Department of Education. Using the Fall 2003-04 Basic Educational Data Survey, completed by your school district and sent to the Department of Education, they provided me with a list of all Iowa public school teachers who are in their third year of teaching this school year. **I am requesting that you go to the following website and complete an electronic survey**.

The name of the URL is:
http://www.educ2.iastate.edu/database/brad/Content/brad_asp_survey.asp

Your participation code:

The participation code will be used to keep track of response rates. If at the end of the survey you indicate that you would be willing to be contacted in the future, it would be used for that purpose, as well. Please know that the individual responses to this survey will be kept strictly confidential. Scores will be reported in the final document in an aggregated form. As you probably know, the better the participation rate of the respondents, the more accurate the information and results that will be contained in the survey. Please take the time – approximately 15 minutes – to visit the site and complete the survey. **As an added incentive, anyone who completes the survey will have his/her name placed in a drawing for one of two $50.00 checks, made out directly to each of the two winners**!

Your participation in the survey is completely voluntary and it can be terminated at any point in the process of completing the survey. Please know that if the survey is aborted prematurely, the results will not be compiled by the survey program.

If you have questions, please feel free to contact me at 515.987.5177 or bbuck@waukee.k12.ia.us. If you prefer, you may contact my major professor at ISU, Dr. Donald Hackmann (515-494-4871 or hackmann@iastate.edu). Thank you in advance for completing the survey!

Respectfully,

Brad Buck

APPENDIX B. ELECTRONIC SURVEY

Perceptions of Iowa Novice Teachers on the Effects of the Implementation of a State-Mandated Teacher Evaluation Framework

In recent years, the teaching profession has been marked by rapid change and the emergence of a number of issues and concerns. One of the areas receiving a great deal of attention is teacher evaluation. As you are aware, the state of Iowa has implemented the Iowa Teaching Standards and Criteria and requires novice teachers to participate in a mentoring and induction program prior to being recommended for their standard licenses.

This survey is designed to allow you to describe your experiences with teacher evaluation. Your responses will be combined with those of other teachers from Iowa to provide a picture of the key elements in an effective teacher evaluation experience.

This research is intended to determine if and how evaluation can be improved to serve relevant and useful purposes and to determine the effects of the evaluator approval training for Iowa educators. As a member of the first group of teachers to complete the provisional licensure requirements in Iowa, your input is especially important to this research.

This survey will take approximately 20 minutes to complete. **Your participation is very much appreciated.**

Informed Consent

By logging into the electronic survey you are providing consent to use the data as has been described in this document as well as in the cover letter for the survey. Please remember that the individual survey results will be kept strictly confidential.

Please insert the user code number provided on the cover letter.

Start

Did you **have the same evaluator** (typically the principal, but not always) for **both** the w001-02 and 2002-03 schools years?

Yes No N/A

Dis you complete your first two years of teaching in the 2001-02 and 2002-03 school years **and** transitions from a provisional license to a standard teaching license for the 2003-04 school year?

Yes No N/A

If you selected "No" for either or both of these questions, **then scroll to the bottom of the survey to "submit survey" and thank you for your time in completing the survey.** If you chose "Yes" for both of these questions **most people will answer "Yes" for both questions), please continue with the survey, starting with question number 1 on the next page.**

The following questions will ask for your views regarding your attributes as a teacher, your perceptions of your evaluator, the procedures used during your observations, the feedback you received, and the evaluation process within the context of the school year. You will be asked to make comparisons between your first year of teaching in 2001-02 and your most recently completed year of teaching (2002-03).

<u>Questions 1-3</u> address attributes of the procedures used during your observation(s) in 2001-02 and your observation(s) in 2002-03, in particular, the procedures used to address the dimensions of your teaching/(standards) to be evaluated.

1. The district teaching standards and criteria were communicated to me in detail.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

2. The content of the district teaching standards and criteria was clear to me.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

3. The district teaching standards and criteria were differentiated to meet my unique learning needs.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

<u>Questions 4-6</u> relate to the procedures that were used to address the dimensions of your teaching that were evaluated.

4. Direct observation of my classroom performance was used extensively in my evaluation.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

5. A review of classroom or school records (lessons plans, etc.) was used extensively in my evaluation(s).

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

6. A review of student achievement data was part of my evaluation(s).

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

<u>Questions 7-8</u> relate to the extent of observation in your classroom. Note: in the items, <u>FORMAL</u> refers to observations that were pre-announced and followed by a conference with the evaluator; <u>INFORMAL</u> refers to unannounced drop-in visits that were accompanied by some type of written and/or oral feedback.

7. Number of FORMAL observations.

A. 2001-02

More than 4 3 2 1 0

B. 2002-03

More than 4 3 2 1 0

8. Approximate frequency of INFORMAL observations.

A. 2001-02

Daily Once per week Once per month Less than once per month None

B. 2002-03

Daily Once per week Once per month Less than once per month None

<u>Questions 9-13</u> relate to the attributes of the feedback you received.

9. I received a great deal of feedback from my evaluator that was directly applicable to my classroom.

A. 2001-02

Strongly Disagree Agree

Disagree Strongly Agree

Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree Agree

Disagree Strongly Agree

Neither Agree nor Disagree.

10. The ideas and suggestions contained in the feedback from my evaluator were of high quality.

A. 2001-02

Strongly Disagree Agree

Disagree Strongly Agree

Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree Agree

Disagree Strongly Agree

Neither Agree nor Disagree.

11. The information provided by my evaluator was very specific.

A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

12. The nature of the information provided by my evaluator was descriptive rather than judgmental.

A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

13. The feedback from my evaluator was focused on district teaching standards and criteria.

A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

Please describe these attributes of the evaluation context (questions 14-18).

Resources available for professional development (questions 14-15).

14. An adequate amount of time was allocated during the teaching day for professional development.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

15. My district makes many training programs/models of best practice available to teachers.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

District values and policies in evaluation (questions 16-18).

16. The district has clear policy statements regarding the purpose(s) of evaluation.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

17. The evaluation(s) is(are) intended to place a high emphasis on teacher accountability.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

18. The evaluation(s) is(are) intended to place a high emphasis on teacher professional growth.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

Personal Teaching Attributes (questions 19-27).

19. In terms of my professional expectations, I demand a great deal from myself.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

20. In terms of my professional orientation, I consider myself an instructional "risk-taker."
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

21. As it relates to my classroom, I consider myself relatively open to change.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

22. I consider myself to be a teacher who frequently engages in instructional experimentation in the classroom.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

23. I consider myself to be open for constructive criticism.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

24. I have a great deal of knowledge about the technical aspects of teaching.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

25. I have a great deal of knowledge about the subject matter than I am responsible to teach.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

26. I am quick to implement suggestions I have received from my evaluator.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

27. I embraced the district teaching standards and criteria as appropriate for my classroom.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

Perceptions of my evaluator (questions 28-38).

28. My evaluator is a credible source of meaningful feedback.
A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

29. My evaluator is helpful in promoting my ongoing professional growth.
A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

30. My evaluator is trustworthy.
A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

31. My evaluator's interpersonal manner is non-threatening.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

32. My evaluator demonstrated flexibility throughout the evaluation process.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

33. My evaluator is knowledgeable about the technical aspects of teaching.
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

34. My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s).
A. 2001-02

Strongly Disagree
Agree
Disagree
Strongly Agree
Neither Agree nor Disagree.

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

35. My evaluator is familiar with my classroom.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

36. My evaluator is a content area specialist in the field in which I teach.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

37. The suggestions my evaluator provides are useful for my professional improvement.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

38. My evaluator is able to provide a persuasive rationale for suggestions for improvement.

A. 2001-02

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

B. 2002-03

Strongly Disagree
Disagree
Neither Agree nor Disagree.
Agree
Strongly Agree

The intent of the rest of this survey is to draw overall comparisons between your first year of teaching (2001-02) and your reflections on your second year of teaching (2002-03).

39. Please rate your overall experience related to your district's formal teacher evaluation process. Take into account the entire evaluation process, including planning for the evaluation, classroom observations, feedback from classroom observations, and the summative evaluation.

A. 2001-02

Very Poor Poor Fair Good Excellent

B. 2002-03

Very Poor Poor Fair Good Excellent

40. Please rate the **influence** of your experiences with your district's formal evaluation process as it relates to your teaching practices, attitudes about teaching, and/or understanding of teaching.

A. 2001-02

No Impact
Minimal Impact
Fair Amount of Impact
Moderate Impact
Strong/Profound Impact

B. 2002-03

No Impact
Minimal Impact
Fair Amount of Impact
Moderate Impact
Strong/Profound Impact

41. What one piece of additional feedback would you give to groups who may have responsibility for potential changes to legislation regarding teacher quality, specifically in the area of teacher evaluation?

42. In what ways, if any, has the focus on the Iowa Teaching Standards and Criteria helped you to improve the quality of teaching?

43. Are there any specific personal experiences, as they relate to teacher evaluation, that you would have liked to include in this survey? If so, please share them below.

44. What additional comments do you have?

Demographic information

45. Gender

Female Male

46. Ethnicity (please check all that apply):

White Native-American African-American Hispanic Asian Other

47. Gender of **your evaluator.**

Female Male

Send Answers Clear All

APPENDIX C. HUMAN SUBJECTS APPROVAL

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office of Research Compliance
Vice Provost for Research and
Advanced Studies
2810 Beardshear Hall
Ames, Iowa 50011-2036
515 294-4566
FAX 515 294-7288

TO: Bradley A. Buck

FROM: Ginny Austin, IRB Administrator

RE: IRB ID # 04-116
DATE REVIEWED: February 27, 2004

The project, *"A Comparative Study of the Perceptions of Novice Iowa Public School Teachers Toward the Effects of the IEATP"*, has been declared exempt from Federal regulations as described in 45 CFR 46.101(b)(2) according to the review and decision made by the IRB Committee.

2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

To be in compliance with ISU's Federal Wide Assurance through the Office of Human Research Protections (OHRP) all projects involving human subjects, must be reviewed by the Institutional Review Board (IRB). Only the IRB may determine if the project must follow the requirements of 45 CFR 46 or is exempt from the requirements specified in this law. Therefore, all human subject projects must be submitted and reviewed by the IRB.

Because this project is exempt it does not require further IRB review and is exempt from the Department of Health and Human Service (DHHS) regulations for the protection of human subjects.

We do, however, urge you to protect the rights of your participants in the same ways that you would if IRB approval were required. This includes providing relevant information about the research to the participants. Although this project is exempt, you must carry out the research as proposed in the IRB application, including obtaining and documenting (signed) informed consent, if applicable to your project.

Any modification of this research should be submitted to the IRB on a Continuation and/or Modification form to determine if the project still meets the Federal criteria for exemption. If it is determined that exemption is no longer warranted, then an IRB proposal will need to be submitted and approved before proceeding with data collection.

cc: ELPS

HSRO/OCR 9/02

APPENDIX D. PAPER SURVEY

Perceptions of Iowa Novice Teachers on the Effects of the Implementation of a State-Mandated Teacher Evaluation Framework

In recent years, the teaching profession has been marked by rapid change and the emergence of a number of issues and concerns. One of the areas receiving a great deal of attention is teacher evaluation. As you are aware, the state of Iowa has implemented the Iowa Teaching Standards and Criteria and requires novice teachers to participate in a mentoring and induction program prior to being recommended for their standard licenses.

This survey is designed to allow you to describe your experiences with teacher evaluation. Your responses will be combined with those of other teachers from Iowa to provide a picture of the key elements in an effective teacher evaluation experience.

This research is intended to determine if and how evaluation can be improved to serve relevant and useful purposes and to determine the effects of the evaluator approval training for Iowa educators. As a member of the first group of teachers to complete the provisional licensure requirements in Iowa, your input is especially important to this research.

This survey will take approximately 20 minutes to complete. **Your participation is very much appreciated**.

Informed Consent

By logging into the electronic survey you are providing consent to use the data as has been described in this document as well as in the cover letter for the survey. Please remember that the data results will be kept strictly confidential.

Did you have **the same evaluator (typically the principal, but not always) for both the 2001-02 and 2002-03 school years?

☐ Yes ☐ No

Did you complete your first two years of teaching in 2001-02 and 2002-03 **and transition from a provisional license to a standard teaching for the 2003-04 school year?

☐ Yes ☐ No

If you selected "No" for **either or both of these questions**, please scroll to the bottom of the survey to "submit survey" and thank you for your time in completing the survey.

If you chose "Yes" **for both of these questions** (most people will answer "yes" for both), please continue with the survey, starting with question number 1, below.

The following questions will ask for your views regarding your attributes as a teacher, your perceptions of your evaluator, the procedures used during your observations, the feedback you received, and the evaluation process within the context of the school year. You will be asked to make comparisons between your first year of teaching in 2001-02 and your most recently completed year of teaching (2002-03).

Please use the following descriptors in answering questions 1-6:

1 = Strongly Disagree
2 = Disagree
3 = Neither Agree nor Disagree
4 = Agree
5 = Strongly Agree

A. Questions 1-3 address attributes of the procedures used during your observation(s) in 2001-02 and your observation(s) in 2002-03, in particular, the procedures used to address the dimensions of your teaching/(standards) to be evaluated?

1. The district teaching standards and criteria were communicated to me in detail.
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
2. The content of the district teaching standards and criteria was clear to me.
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
3. The district teaching standards and criteria were differentiated to meet my unique learning needs.
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Questions 4-6 relate to the procedures that were used to address the dimensions of your teaching that were evaluated.

4. Direct observation of my classroom performance was used extensively in my evaluation(s).
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
5. A review of classroom or school records (lesson plans, etc.) was used extensively in my evaluation(s).
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
6. A review of student achievement data was part of my evaluation(s).
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Questions 7-8 relate to the extent of observation in your classroom.
Note: In these items, FORMAL refers to observations that were pre-announced and followed by a conference with the evaluator; INFORMAL refers to unannounced drop-in visits that were accompanied by some type of written and/or oral feedback.

7. Number of FORMAL observations:
   A. 2001-02 more than 4 ☐ 4 ☐ 3 ☐ 2 ☐ 1 ☐ 0 ☐
   B. 2002-03 more than 4 ☐ 4 ☐ 3 ☐ 2 ☐ 1 ☐ 0 ☐

8. Approximate frequency of INFORMAL observations:
   A. 2001-02
      Daily ☐
      Once per week ☐
      Once per month ☐
      Less than once per month ☐
      None ☐

   B. 2002-03
      Daily ☐
      Once per week ☐
      Once per month ☐
      Less than once per month ☐
      None ☐

Please use the following descriptors in answering questions 9-38:

1 = Strongly Disagree
2 = Disagree
3 = Neither Agree nor Disagree
4 = Agree
5 = Strongly Agree

B. Please describe these attributes of the feedback you received (questions 9-13).

9. I received a great deal of feedback from my evaluator that was directly applicable to my classroom.
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

10. The ideas and suggestions contained in the feedback from my evaluator were of high quality.
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

11. The information provided by my evaluator was very specific
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

12. The nature of the information provided by my evaluator was descriptive rather than judgmental.
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

13. The feedback from my evaluator was focused on district teaching standards and criteria.
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

C. Please describe these attributes of the evaluation context (questions 14-18).

i. Resources available for professional development (questions 14-15).

14. An adequate amount of time was allotted during the teaching day for professional development.
   A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
   B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

15. My district makes many training programs/models of best practice available to teachers.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

ii. District values and policies in evaluation (questions 16-18).

16. The district has clear policy statements regarding the purpose(s) of evaluations.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

17. The evaluation(s) is/are intended to place a high emphasis on teacher accountability.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

18. The evaluation(s) is/are intended to place a high emphasis on teacher professional growth.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

D. Personal Teaching Attributes (questions 19-27):

19. In terms of my professional expectations, I demand a great deal from myself.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

20. In terms of my professional orientation, I consider myself an instructional "risk-taker."
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

21. As it relates to my classroom, I consider myself relatively open to change.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

22. I consider myself to be a teacher who frequently engages in instructional experimentation in the classroom.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

23. I consider myself to be open to constructive criticism.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

24. I have a great deal of knowledge about the technical aspects of teaching.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

25. I have a great deal of knowledge about the subject matter that I am responsible to teach.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

26. I am quick to implement suggestions I have received from my evaluator.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

27. I embraced the district teaching standards and criteria as appropriate for my classroom.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

E. Perceptions of Your Evaluator (questions 28-38):

28. My evaluator is a credible source of meaningful feedback.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

29. My evaluator is helping in promoting my ongoing professional growth.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

30. My evaluator is trustworthy.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

31. My evaluator's interpersonal manner is non-threatening.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

32. My evaluator demonstrated flexibility through the evaluation process.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

33. My evaluator is knowledgeable about the technical aspects of teaching.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

34. My evaluator has the capacity to demonstrate or model needed improvements in my classroom practice(s).
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

35. My evaluator is familiar with my particular classroom.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

36. My evaluator is familiar with classrooms in general within my school.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

37. The suggestions my evaluator provides are useful for my professional improvement.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

38. My evaluator is able to provide a persuasive rationale for suggestions for improvement.
    A. 2001-02 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
    B. 2002-03 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

The intent of the rest of this survey is to draw overall comparisons between your first year of teaching (2001-02) and your reflections on your second year of teaching (2002-03).

39. Please rate your overall experience related to your district's formal teacher evaluation process. Take into account the entire evaluation process, including planning for the evaluation, classroom observations, feedback from classroom observations, and the summative evaluation.

A. 2001-02

- ☐ 1 Very Poor
- ☐ 2 Poor
- ☐ 3 Fair
- ☐ 4 Good
- ☐ 5 Excellent

B. 2002-03

- ☐ 1 Very Poor
- ☐ 2 Poor
- ☐ 3 Fair
- ☐ 4 Good
- ☐ 5 Excellent

40. Please rate the **influence** of your experiences with your district's formal evaluation process as it relates to your teaching practices, attitudes about teaching, and/or understanding of teaching.

A. 2001-02

- ☐ 1 No Impact
- ☐ 2 Minimal Impact
- ☐ 3 Fair Amount of Impact
- ☐ 4 Moderate Impact
- ☐ 5 Strong/Profound Impact

B. 2002-03

- ☐ 1 No Impact
- ☐ 2 Minimal Impact
- ☐ 3 Fair Amount of Impact
- ☐ 4 Moderate Impact
- ☐ 5 Strong/Profound Impact

41. What one piece of additional feedback would you give to groups who may have responsibility for potential changes to legislation regarding teacher quality legislation, specifically in the area of teacher evaluation?

42. In what ways, if any, has the focus on the Iowa Teaching Standards and Criteria helped you to improve the quality of your teaching?

43. Are there any specific personal experiences, as they relate to teacher evaluation, that you would have liked to have included in this survey? If so, please share them below.

44. What additional comments do you have?

Demographic Information:

45. Gender:

☐ Female ☐ Male

46. Ethnicity (please check all that apply):

☐ White
☐ Native-American
☐ African-American
☐ Hispanic
☐ Asian
☐ Other

47. Gender of **your evaluator**:

☐ Female ☐ Male

48. Ethnicity of **your evaluator** (please check all that apply):

☐ White
☐ Native-American
☐ African-American
☐ Hispanic
☐ Asian
☐ Other

49. Please provide your current age:
    - ☐ 0-20
    - ☐ 21-23
    - ☐ 24-26
    - ☐ 27 or older

50. Please list the grade level(s) and subject(s) you taught in 2002-03:

Grade	Subject(s)
Grade	Subject(s)
Grade	Subject(s)

51. Please estimate the approximate number of hours you spent on the evaluation process overall, including the completion of relevant paperwork, any conferencing that occurred, etc., – in:
    - A. 2001-02
    - B. 2002-03

52. Would you be willing to participate in an on-site follow-up interview?
    - ☐ Yes ☐ No

Thank you for your thoughtful responses!

BIBLIOGRAPHY

Achilles, C. (1988). Unlocking some mysteries of administration and administrator preparation: A reflective prospect. In D. Griffiths, R. Stout, & P. Forsyth (Eds.), *Leaders for America's schools: Final report and papers of the national commission on excellence in educational administration* (pp. 41-67). San Francisco: McCutchan.

Agresti, A. & Finlay, B. (1997). *Statistical methods for the social sciences*. Upper Saddle River, NJ: Prentice Hall.

Allix, N. M. (2000). Transformational leadership democratic or despotic? *Educational Management and Administration*, *28*(1), 7-20.

Amos, A. (1988). *Differences in attitudes toward post evaluation conferences for proficient and nonproficient teachers in Mississippi*. (ERIC Document Reproduction Service No. ED302942)

Armato, N. (1990). Career stages in the teaching profession: A description of distinguishing attitudes and behaviors. *Dissertation Abstracts International, 51,* (04), 1200. (UMI No. 9023510)

Amendt, B. C. (2005). A study of administrator perceptions of state-mandated teacher evaluation: The student achievement and Iowa teacher quality law. *Dissertation Abstracts International, 65*(05), 1597. (UMI No. 3133102)

Backus, J. M. (1989). Teacher evaluation methods and attitudes toward evaluation in metropolitan Milwaukee public schools. *Dissertation Abstracts International 51*(03), 688. (UMI No. 9021142)

Barber, L. W. (1985). *Improving teacher performance: Formative evaluation*. Bloomington, IN: Phi Delta Kappa.

Bartky, J. A. (1956). *Administration as educational leadership*. Stanford, CA: Stanford University Press.

Bass, B. M., Farrow, D. L., Valenzi, E. R., & Solomon, R. J. (1975). Management styles associated with organizational, task, personal, and interpersonal contingencies. *Journal of Applied Psychology*, *60*(6), 720-729.

Berger, J. (2002). *Updated information regarding the revised teacher evaluator approval training program development.* Retrieved March 12, 2004, from http://www.state.ia.us/educate/ecese/tqt/tc/teacheval.html

Blase, J., & Blase, J. (1997). The micropolitical orientation of facilitative school principals and its effects on teachers' sense of empowerment. *Journal of Educational Administration*, *35*(2), 138-157.

Blase, J., & Blase, J. (1999a). Principals' instructional leadership and teacher development: Teachers' perspectives. *Educational Administration Quarterly*, *35*(3), 349-378.

Blase, J., & Blase, J. (1999b). Implementation of shared governance for instructional improvement: Principals' perspectives. *Journal of Educational Administration*, *37*(5), 476-498.

Bolman, L., & Deal, T. (1992). Images of leadership. *The American School Board Journal*, *179*(4), 36-39.

Bourisaw, D. M. (1989). Anticipatory set bias: Effects on teacher observation in evaluation. *Dissertation Abstracts International, 50*(01), 31. (UMI No. 8909129)

Brown, P. R. (1990). Accountability in public education. *Policy Briefs, Far West Laboratory, 14*, 1-6.

Brown, S. L. (1997). Perceptions of principals and teachers regarding the West Virginia mandated performance evaluation of teachers. *Dissertation Abstracts International, 58*(03), 656. (UMI No. 9727667)

Brown, N. M. (2004). Improving the first year: How opportunities for faculty collaboration, support from administrators, reduced workloads and formal induction experiences impact first-year teachers. *Dissertation Abstracts International, 65*(05), 1737. (UMI No. 3134471)

Brown, D. M. (2005). Teacher perceptions of African-American principal leadership. *Dissertation Abstracts International, 66* (05), 1569. (UMI No. 3175813)

Bulach, C., & Peterson, T. (1999). *Levels of openness and trust: Do principals "walk the talk"?* (ERIC Document Reproduction Service No. ED428443)

Burden, P. (1982). *Developmental supervision: Reducing teacher stress at different career stages.* Paper presented at the annual conference of the Association of Teacher Educators, Phoenix, AZ.

Burke, P., Christensen, J., Fessler, R., McDonnell, J., & Price, J. (1987, April). The teacher career cycle: Model development and research report. Paper presented at the annual meeting of the American Educational Research Association, pp. 1-59.

Burns, J. M. (1979). Two excerpts From *Leadership. Educational Leadership*, *36*(6), 380-383.

Buttram, J. L., & Wilson, B. L. (1987). Promising trends in teacher evaluation. *Educational Leadership*, *7*(7), 4-6.

Casey, J. A. (1994). Career stages in the teaching profession: A description of distinguishing characteristics. *Dissertation Abstracts International 55,* (04), 809. (UMI No. 9425276)

Cioci, M. (1991). *Teachers and principals: Gender related perceptions of leadership and power in secondary school. Revised.* (ERIC Document Reproduction Service No. ED332364)

Clabaugh, G. K. (2001). Looking for leadership: Longing for a führer. *Educational Horizons, 79*(3), 106-107.

Cogan, M. L. (1973). *Clinical supervision.* Boston: Houghton Mifflin.

Conley, D. T. (1987). Critical attributes of effective evaluation systems. *Educational Leadership, 44*(7), 60-64.

Cuban, L. (1988). *The managerial imperative and the practice of leadership in schools.* Albany: State University of New York Press.

Cunningham, J. M. (2004). The effect of gender-role identity of female high school principals in Texas on teacher perceptions of instructional leadership behaviors. *Dissertation Abstracts International 65,* 10, 3635. (UMI No. 3149098)

Cunningham, L.L. (1971). *Governing schools: New approaches to old issues.* Columbus, OH: Charles E. Merrill.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2001). New trends in teacher evaluation. *Educational Leadership, 58*(5), 12-15.

Danielson, C., & McGreal, T. (2000). *Teacher evaluation: To enhance professional practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

Dawson, V., & Acker-Hocevar, M. (1998). *Contradictions within a state-mandated teacher evaluation system: A historical perspective, personal reflections and principal interviews.* Paper presented at the 1998 Annual Mid-South Educational Research Conference, New Orleans, LA.

Day, C., Harris, A., & Hadfield, M. (2001). Challenging the orthodoxy of effective school leadership. *International Journal of Leadership in Education, 4*(1), 39-56.

De Bevoise, W. (1984). Synthesis of research on the principal as instructional leader. *Educational Leadership, 41*(5), 15-20.

DeMoulin, D., & Guyton, J. (1987). *Phase II: Examination of principal's perceptions in identifying instructional stages associated with teacher output.* Paper presented at annual meeting of Mid-south Educational Research Association, pp. 1-23.

Dowdle, J. M. (1991). Evaluator training: One response to a legislative mandate for improved evaluation of professional educators. *Dissertation Abstracts International 52*(02), 360. (UMI No. 9119809)

Duffy, F. M. (1985, March). *Analyzing and evaluating supervisory practice.* Paper presented at the meeting of the Association for Supervision and Curriculum Development, Chicago, IL.

Duke, D. L., & Stiggins, R. J. (1986). *Teacher evaluation: Five keys to growth.* Washington, DC: NEA Professional Library.

Eberhart, M. C. (1990). The teaching profession: An evaluation of career stages. *Dissertation Abstracts International 51,* (09), 2956. (UMI No. 9105923)

Edmunds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15-24.

Ellett, C. D. (1987). Emerging teacher performance assessment practices: Implications for the instructional supervision role of school principals. In W. Greenfield (Ed.), *Instructional leadership concepts, issues, and controversies* (pp. 304-309). Boston: Allyn and Bacon.

Farrar, E. (1990). Reflections on the first wave of reform: Reordering America's educational priorities. In S. L. Jacobson & J.A. Conway (Eds.), *Educational leadership in an age of reform* (pp. 30-31). New York: Longman.

Fessler, R., & Christensen, J. (1992). *The teacher career cycle.* Boston: Allyn and Bacon.

Fink, E., & Resnick, L. B. (2001). Developing principals as instructional leaders. *Educational Leadership, 82*(8), 598-610.

Franseth, J. (1961). *Supervision as leadership.* Evanston, IL: Row, Peterson and Company.

Frase, L. E., & Streshly, W. (2000). *Top 10 myths in education: Fantasies Americans love to believe.* Lanham, MD: The Scarecrow Press.

Fullan, M. (2002). The change leader. *Educational Leadership*, 59(8), 16-20.

Fuller, F. (1969). Concerns of teachers: A developmental conceptualization. *Educational Research Journal, 6*, pp. 207-226.

Glatthorn, A. A. (1996). *Quality teaching through professional development. Principals taking action series.* Thousand Oaks, CA: Corwin Press.

Glatthorn, A. A. (1997). *Differentiated supervision. 2nd Edition.* Alexandria, VA: Association for Supervision and Curriculum Development.

Glickman, C. D. (2002a). *Leadership for learning: How to help teachers succeed.* Alexandria, VA: Association for Supervision and Curriculum Development.

Glickman, C. D. (2002b). The courage to lead. *Educational Leadership, 59*(8), 41-44.

Gougeon, T. D., & Hutton, S. I. (1993). Effects of culture and gender on principal-teacher communication in schools. *Human Resource Development Quarterly, 4*(3), 277-290.

Gregoric, A. (1973, December). Developing plans for professional growth. *NASSP Bulletin, 57*, 377, 1-8.

Hallinger, P. (1992). The evolving role of American principals: From managerial to instructional to transformational leaders. *Journal of Educational Administration, 30*(3), 35-48.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences.* Boston: Houghton Mifflin Company.

Holland, P. E., & Adams, P. (2002). Through the horns of a dilemma between instructional supervision and the summative evaluation of teaching. *International Journal of Leadership in Education, 5*(3), 227-247.

Huberman, M. (1989, Fall). The professional life cycle of teachers. *Teacher College Record*, 91, 31-57.

Iowa Department of Education. (2002a, Fall). *Data Driven Leadership Participant's Manual.* Des Moines.

Iowa Department of Education. (2002b, Fall). *Iowa Evaluator Approval Training Program Participant's Manual.* Des Moines.

Iowa Department of Education. (2005). *The Annual Condition of Education Report.* Des Moines.

Jacobson, S. L. (1990). Reflections on the third wave of reform: Rethinking administrator preparation. In S. L. Jacobson & J.A. Conway (Eds.), *Educational leadership in an age of reform* (30-31). New York: Longman.

Jellinek, M. (1985). Phases of a teaching career. *Independent School, 45*(2), 7-11.

Johnson, J. (2002). Staying ahead of the game. *Educational Leadership, 59*(8), 26-30.

Katz, L. (1972). Developmental stages of preschool teachers. *Elementary School Journal, 73*, 50-54.

Kelly, T. F. (1999). Why state mandates don't work, *Phi Delta Kappan*, *80*(7), 543-546.

Kemmerer-Fehr, S. J. (2001). The role of the educational supervisor In United States public schools from 1970 to 2000 as reflected in the supervision literature. *Dissertation Abstracts International 62,* (07), 2327. (UMI No. 3020446)

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*. Pacific Grove: Brooks/Cole.

Knowles, M. S. (1967). Program planning for adults as learners. *Adult Leadership*, *15*(8), 267-268, 278-279.

Knowles, M. S. (1973). *The adult learner: A neglected species*. Houston, TX: Gulf Publishing Company.

Lambert, L., Walker, D., Zimmerman, D. P., Cooper, J. E., Lambert, M. D., Gardner, M., et al. (2002). *The constructivist leader*. New York: Teachers College Press.

Lawler, D. W. (1992). An examination of teacher evaluation in Iowa. *Dissertation Abstracts International 53,* (03), 677. (UMI No. 9223940)

Leithwood, K. (1992). The move toward transformational leadership. *Educational Leadership*, *49*(5), 8-12.

Leithwood, K. (1994). Leadership for school restructuring. *Educational Administration Quarterly*, *30*(4), 498-518.

Leithwood, K. & Duke, D. (1999). A century's quest to understand school leadership. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (pp. 45-72). San Francisco: Jossey-Bass.

Leithwood, K., Jantzi, D., & Steinbach, R. (1999). *Changing leadership for changing times*. Buckingham, U.K.: Open University Press.

Leypoldt, M. M. (1967). The teaching/learning process with adults. *Adult Leadership*, *16*(6), 212-213, 233.

Manatt, R. P. (1976). Developing a teacher performance evaluation system as mandated by Senate File 205 (Iowa General Assembly, 1976). In Iowa Association of School Boards (Eds.). Ames: Iowa State University, Department of Educational Administration.

Mazzarella, J., & Grundy, T. (1989). Portrait of a leader. In S. C. Smith & P. K. Piele (Eds.), *School leadership: Handbook for excellence* (pp. 9-27). Eugene, OR: ERIC Clearing House on Educational Management.

McDonald, F. J. (1982). *A theory of the professional development of teachers.* Paper presented at the meeting of the American Educational Research Association, New York, March.

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

McKenna, B.H. (1973, February). A context for teacher evaluation. *The National Elementary Principal, 52,* 18-23.

Moir, E. (1999). The stages of a teacher's first year. In M. Scherer (Ed.), *A better beginning: Supporting and mentoring new teachers* (19-23). Alexandria, VA: Association of Supervision and Curriculum Development.

National Commission on Excellence in Education. (1993). *A nation at risk: The imperative for educational reform* (Stock No. 065-000-00177-2). Washington, DC: U.S. Government Printing Office.

National Education Association. (1973-74, Winter). *The evaluation of teachers.* Washington, DC: NEA Instruction And Professional Development Staff.

Nelson, J. (2001). *The art of focused conversations for schools: Over 100 Ways to guide clear thinking and promote learning.* Toronto, ON: The Canadian Institute of Cultural Affairs and New Society Publishers.

Newman, K. K. (1978). Middle-aged experienced teachers' perceptions of their career development. *Dissertation Abstracts International, 39*(08), 4885. (UMI No. 7902196)

North Central Regional Educational Laboratory. (1996). *Adult learning theory.* Retrieved May 30, 2003, from http://www.ncrel.org/sdrs/areas/issues/methods/technlgy/te10lk12.htm

Office of the Under Secretary. (2002). *No child left behind: A desktop reference.* Jesup, MD: ED Pubs.

Pajak, E. (2003). *Honoring diverse teaching styles: A guide for supervisors.* Alexandria, VA: Association for Supervision and Curriculum Development.

Petersen, A. R. (1978). Career patterns of secondary school teachers: An exploratory interview study of retired teachers. *Dissertation Abstracts International 39*(08), 4888. (UMI No. 7902204)

Peterson, D. W. (1988). Rating teacher performance to determine career ladder advancement: An analysis of bias and reliability. *Dissertation Abstracts International 49* (09), 2482. (UMI No. 8825436)

Peterson, K. D. (2000). *Teacher Evaluation: A comprehensive guide to new directions and practices.* Thousand Oaks, CA: Corwin Press.

Petrie, H. G. (1990). Reflections on the second wave of reform: Restructuring the teaching profession. In S. L. Jacobson and J.A. Conway (eds.), *Educational leadership in an age of reform* (30-31). New York: Longman.

Poplin, M. S. (1992). The leader's new role: Looking to the growth of teachers [special inset]. *Educational Leadership*, *49*(5), 10-11.

Rindler, B. H. (1994). *The attributes of teacher evaluation systems that promote teacher growth as perceived by teachers of intense English programs.* Boston University (Dissertation #AAT 9417322)

Rallis, S., & Highsmith, M. (1986). The myth of the "great principal": Questions of school management and instructional leadership. *Phi Delta Kappan*, *68*(4), 300-304.

Rinehart, J. S., & Young, I. P. (1996). Effects of teacher gender and principal gender on teacher performance. *Journal of Personnel Evaluation in Education*, *10*(4), 313-323.

Ryan, K., et al. (1979). *The stages in teaching: New perspectives on staff development for teachers' needs.* Paper presented at the ASCD Conference, Anaheim, CA, March.

Ryan, K. (1986). *The induction of new teachers.* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Sergiovani, T. J. (1990). Adding value to educational leadership gets extraordinary results. *Educational Leadership*, *47*(8), 23-27.

Smylie, M. A., & Hart, A. W. (1999). School leadership for teacher learning and change: A human and social capital development perspective. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (pp. 421-441). San Francisco: Jossey-Bass.

Steffy, B. (1987). How to develop a human development model. *School Administrator*, *44*(5), 14-15.

Steffy, B. E., Wolfe, M., Preston, P., & Enz, B. (Eds.). (2000). *Life cycle of the career teacher.* Thousand Oaks, CA: Corwin Press.

Stiggins, R. J., & Duke, D., (1988). *The case for commitment to teacher growth: Research on teacher evaluation.* New York: State University of New York Press.

Stronge, J. H., & Tucker, P. D. (2000). *Teacher evaluation and student achievement.* Washington DC: National Education Association.

Tishler, A., & Neely, A. (1987, November). *Perceptions about career ladders across levels of teaching socialization.* Paper presented at Mid South Educational Research Association, Mobile, AL.

Uchiyama, K., & Wolf, S. (2002). The best way to lead them. *Educational Leadership*, *59*(8), 80-83.

United States Department of Education. (2001). *Digest of education statistics*. Jesup, MD: ED Pubs.

Unruh, A., & Turner, H. (1970). *Supervision for change and innovation*. Boston: Houghton Mifflin.

Vonk, J. (1989). *Becoming a teacher, brace yourself.* Unpublished paper. Amsterdam: Vrije University.

Ward, B., & Sistrunk, W. (1988). The relationship between teachers' perceptions of principals' actual supervisory behavior and teachers' preferred supervisory behavior in selected southern public junior high schools. (ERIC Document Reproduction Service No. ED302929)

Williams-Boyd, P. (2002). *Educational leadership: A reference handbook*. Santa Barbara, CA: ABC-CLIO.

Winter, P., et al. (1997). *The influence of work values on principal teacher selection decisions in schools undergoing reform.* (ERIC Document Reproduction Service No. ED435132)

ACKNOWLEDGMENTS

In a letter to the Philippians from a Roman prison, Paul demonstrates that he never lost his passion, his sense of mission, or his sense of direction. Admittedly, at one point early in the letter he mentions that he cannot choose between life or death; to him, both seem like positive options – staying on earth to serve Christ, or dying and going to Heaven in order to be with Him. In the end, he chooses the former rather than the latter as he knows he can be more useful to the cause while alive.

In Chapter 2, Paul discusses the servant's heart of a leader, illustrated through the humility of Christ. In Chapter 3, he addresses priorities, goal setting, and perseverance. At one point in the third chapter, Paul reminds the Phillipians of his trophies that he has discarded as trash compared to the priority of knowing Christ. In Chapter 4, Paul concludes with rejoicing and encourages the readers of the letter to maintain tough minds, fixing on what is right. To me, one of the most important verses in the entire letter is Phillipians 4:13: I can do all things through Christ who strengthens me (New King James Version).

While I do not mean to imply that the work completed as a part of this dissertation is equivalent to the trials that Paul went through, I am comforted by Christ's words through Paul, and inspired by the manner in which he approached life. In addition, I am confident that this research has come to fruition with the support I have in my relationship with a living God.

There are many people to thank at the end of an undertaking of this magnitude. Certainly, I would be remiss in not thanking the three major professors I had over the years: Drs. Bill Poston, Don Hackmann, and Tom Alsbury. I want to especially thank Dr. Poston for

his kind words and sage advice that I should not change jobs while working on this research. There were certainly times in which I wish I had followed that advice. My gratitude also goes to Dr. Hackmann for his efforts in assisting me to frame this work and in pushing me to always strive to make the product as great as it could be regardless of the stage of construction. I especially thank Dr. Alsbury for shepherding me through the end of the process, and for providing me with the encouragement that this actually could come to closure. I will always remember the patience and kindness he has shown me in completing this dissertation.

I am truly grateful to Dr. Mack Shelley for all his work in guiding me through the data analysis process. He is one of the most patient people with whom I have interacted and I certainly taxed his skills as he guided me through Chapters 3 and 4.

I work with many great people and they have supported me throughout this study. Thank you Drs. Stalker and Wilkerson for always believing that I would one day complete this process. Thank you Mike Blake and Juley Murphy-Tiernan for your extra work over the last several months while I took vacation to complete writing. Thanks also to Rox Cumings, for her support and sense of humor – there were certainly times where it felt better to laugh than to cry while realizing the enormity of this task.

To my great friends, Jeff Roberts, Mark Moss, and Nathan Lykkegaard, thank you for your caring support and belief in me. I also thank Dr. William Jacobson and the Scottish Rite Foundation for believing in me and providing financial incentive/support to help make this accomplishment attainable.

Most importantly, I am grateful to my family for their love, patience, and understanding throughout this work. I want to thank my parents, Ray and Karren Buck, for

raising me to value hard work and perseverance, and for instilling in me the desire to finish what has been started. These are the types of traits that have helped me see this research through to its completion. I also thank them for always loving me and encouraging me whether I was shooting a free throw in a junior high basketball game or completing the requirements for the Ph.D.

I want to especially thank my six great children – Gabrielle, Zeke, Grace, Jack, Sam, and Lucie. They have been patient while dad has been at the library, behind books, or sitting at the dissertation table and staring at the computer. I'm sure they grew tired of hearing that dad "…just has to get this done, and then this done, and then this…" Their patience and unconditional love were and still are appreciated. I'm especially looking forward to a summer of swimming and play!

Thank you to my beautiful wife, Traci. Your belief and encouragement in so many ways made the completion of the research and writing of the dissertation happen. I appreciate all of the nights and weekends you transported our children to various activities, put them to bed without seeing their dad for the day, and ran our home expertly without "dad" around very much. Your strength and resolve are nearly unshakeable. Simply, I love you!