TEACHER EVALUATION: ASSESSING PRINCIPALS’ PERCEPTIONS IN THE
STATE OF NEW JERSEY

A Dissertation

Presented to

The Faculty of the School of Education

Liberty University

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

by

Robert J. Fisicaro

February 2010
Teacher Evaluation: Assessing Principals’ Perceptions in the State of New Jersey

by Robert J. Fisicaro

APPROVED:

COMMITTEE CHAIR, GRADUATE STUDIES

Mark Angle, Ed.D.

Charles Schneider, Ed.D.

Thomas Power, Ph.D.

Scott B. Watson, Ph.D.
Abstract

Robert J. Fisicaro. TEACHER EVALUATION: ASSESSING PRINCIPALS’ PERCEPTIONS IN THE STATE OF NEW JERSEY. (Under the direction of Dr. Mark Angle) School of Education, January, 2010. This dissertation describes a study assessing New Jersey school principals’ attitudes towards various characteristics of teacher evaluation and compares perceptions among sub groups. Four-hundred sixty-two building principals completed a survey which measured perceptions of four constructs of teacher evaluation that were selected from current educational theory: teacher evaluation should be founded in a partnership, differentiated for individuals, ongoing, and considerate of student learning outcomes. Principals were examined as sub groups according to gender, level of school, and years of experience. Descriptive statistics indicated that principals agreed that evaluation systems should be part of an ongoing cycle. Principals were neutral to agreeable on two of the constructs measured – student learning and partnership. Principals were neutral to the construct that evaluation procedures should be differentiated for teachers. Participants were not consistent in their responses to questions that were grouped together to measure a common construct. A MANOVA was completed to examine different perceptions among sub groups. Principals in the sub group of 16-20 years of experience had higher mean scores for the construct of teacher evaluations as an ongoing process. No additional differences by construct were identified among the sub groups. Cronbach’s alpha was utilized to measure the reliability of the survey instrument.
Dedication

This dissertation is dedicated to my family. My beautiful wife Jen, son Michael, and daughter Anna who inspire me daily to be the best husband and father that I can be.
Acknowledgements

I stand in awe of a great God who has blessed me beyond measure. Without the inspiration and assistance of the following people, this dissertation and degree would not have become a reality. To my mother and father, Galina and Robert Fisicaro, who always offered endless love and encouragement. Your guidance through the years has made all of the difference. Mom, you are remembered and loved as much now than when you were with us.

To Tim Lenahan, after all these years I can still hear your voice, “Winners do what losers won’t.” To Bill Lange and Pete Carrera for instilling confidence in me in the pursuit of dreams. To Ed Canzanese for believing in me, those years at Rosa will always be special. To two mentors, Joe Manzione and Donna Delrossi, who offered the insight and encouragement necessary to be successful in the classroom. To John Herbst, who permitted me to “get my sea legs” as an administrator and who never wavered in his support. To Dr. Schneider and Dr. Power, who served diligently on my committee and provided valuable insight. Charlie, our time together in 741 was truly inspirational. To my dissertation chair, Dr. Mark Angle, who provided me with so much consistent and ongoing support.

To my secretary, Candy Chiaradonna, who offered so much assistance with this dissertation. To the administrators, staff, students, and board of education in Logan Township School District for whom I am so thankful, and the faculty of the School of Education at Liberty University. My time here has been a life changing experience for which I will be forever grateful. Thank you!
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.1</td>
<td>65</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>74</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>75</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>75</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>76</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>77</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>79</td>
</tr>
<tr>
<td>Table 4.7</td>
<td>82</td>
</tr>
<tr>
<td>Table 4.8</td>
<td>86</td>
</tr>
<tr>
<td>Table 4.9</td>
<td>88</td>
</tr>
<tr>
<td>Table 4.10</td>
<td>89</td>
</tr>
<tr>
<td>Table 4.11</td>
<td>92</td>
</tr>
<tr>
<td>Table 4.12</td>
<td>94</td>
</tr>
<tr>
<td>Table 4.13</td>
<td>100</td>
</tr>
<tr>
<td>Table 4.14</td>
<td>101</td>
</tr>
<tr>
<td>Table 4.15</td>
<td>103</td>
</tr>
<tr>
<td>Table 4.16</td>
<td>103</td>
</tr>
</tbody>
</table>
List of Figures

Figure 4.1 ........................................................................................................90
Figure 4.2 ........................................................................................................99
Figure 4.3 .........................................................................................................101
Figure 4.4 .......................................................................................................104
Table of Contents

Abstract ........................................................................................................... iii

Dedication ....................................................................................................... iv

Acknowledgements .......................................................................................... v

List of Tables .................................................................................................... vi

List of Figures ................................................................................................... vii

Table of Contents .............................................................................................. viii

Chapter 1: Introduction

  Background ..................................................................................................... 1

  Statement of the Problem ............................................................................. 3

  Research Questions ....................................................................................... 4

  Hypotheses ..................................................................................................... 5

  Target Population .......................................................................................... 5

  Definition of Key Terms ............................................................................... 6

  Professional Significance of the Study ......................................................... 7

Chapter 2: Literature Review

  Traditional Evaluation .................................................................................... 10

  An Absence of Leadership .......................................................................... 11

  Thinking Win-Win ......................................................................................... 14

  Views of Local School Boards ..................................................................... 16

  Perceptions of Administrators ..................................................................... 17

  Missing the Mark .......................................................................................... 18
A View of Separation .......................................................... 18

Teachers as Widgets ..................................................... 20

Historical Changes ....................................................... 22

Modern Evaluation Theory ........................................... 25

A Call for Change .......................................................... 25

The Value of Feedback .................................................. 26

An Ongoing Partnership ................................................ 28

A Link to Student Achievement ..................................... 37

Differentiated Approaches ............................................. 44

Dual Purposes Served .................................................... 46

Comprehensive Frameworks .......................................... 47

Summary of the Literature ............................................. 56

Biblical Integration ....................................................... 58

Chapter 3: Methodology

General Perspective ...................................................... 60

Research Design .......................................................... 60

Research Context and Participants ................................. 61

Research Questions ..................................................... 61

Hypotheses ................................................................. 62

Survey Development ................................................... 63

Development of Questions .......................................... 64

Panel of Experts .......................................................... 67
Chapter 4: Results

Descriptive Statistics ................................................................. 74
Research Question 1 ................................................................. 76
Research Question 2 ................................................................. 87
Research Question 3 ................................................................. 90
Research Question 4 ................................................................. 100
Research Question 5 ................................................................. 102

Chapter 5: Summary and Discussion

Restatement of the Problem ..................................................... 105
Relationship of Current Research to Previous Research .................. 106
Review of Methodology ............................................................. 107
Summary of Results ................................................................. 108
Discussion of Results ............................................................... 110
Limitations of Study ............................................................... 111
Recommendations for Further Research........................................112

Conclusions.................................................................109

References..............................................................................113

Appendix A............................................................................121
Chapter 1: Introduction

Background of the Study

Since 2002 and the commencement of No Child Left Behind, the pressures on public schools to improve the quality of teaching and learning have increased substantially. Summative state assessment scores are being scrutinized in an effort to examine student growth and achievement across the country and create higher accountability measures for schools. Educators, teachers, and administrators alike have scrambled to redefine, restructure, and refocus their efforts on best instructional practices and have placed an increased emphasis on school improvement. School boards and superintendents are placing increased demands on building principals who are attempting to perform as instructional leaders, rather than managers. Building principals are concerning themselves daily with not only structure and order, but also accountability and process as they pertain to teaching and learning. School administrators and principals are referring to themselves more than ever as instructional leaders and as change agents in the current culture of No Child Left Behind. A significant amount of research has been conducted throughout the last thirty years and has demonstrated that leadership, specifically instructional leadership, was one of several defining characteristics of successful schools. Because of the efforts of organizations like the American Association of Curriculum and Development (ASCD) and the Mid-Continent Research for Education and Learning (McREL), evidence was collected and revealed that what good school administrators do can make a difference in teacher quality and student
performance (Waters, Marzano, & McNulty, 2003). Such studies have evolved into modern educational theory and are challenging building principals and administrators to engage in supervisory practices as educational leaders to improve the quality of teaching and learning rather than using evaluation as a management tool. Under the pressures and mandates of the federal legislation, local school boards and the constituents who depend on them are also depending on school leaders and teachers to employ instructional practices that will positively influence student achievement. Research on teacher evaluation declared that an important function of appraisal is to promote the formation of effective teaching practices (Danielson, 2001). In other words, the feedback generated from administrators to teachers positively influences teachers’ practices. Principals are responsible first as instructional leaders to assist teachers in improving their craft, not merely to judge them against a predetermined set of criteria. School building administrators are no longer recognized primarily as managers but are viewed as a catalyst for necessary improvement (Sergiovanni, 1994). In the 21st Century, school improvement is the focal point for educational leadership, and the school principal is recognized as the catalyst for this necessary improvement (Kersten, 2005).

Supervision and evaluation are the leadership functions on which this study will be based. Schmoker (2006) stated, “If teachers teach the good stuff, in the right way, and on most days, schools can achieve miraculous results now in education” (p. 30). An important part of education is making sure that principals impart the type of instructional leadership that is frequently discussed but rarely practiced by providing teachers with targeted feedback aimed at improving practices and influencing learning.
Statement of the Problem

There is a disconnect between what modern educational theory states about teacher evaluation and what is practiced and perceived as valuable by New Jersey school principals. The purpose of this study is to assess the levels to which principals’ beliefs are consistent with what modern educational theory states regarding the characteristics of teacher appraisal. In a former era of education, teachers were mostly responsible for raising student achievement, while principals worked primarily as managers to maintain order and operations in school buildings. However, in the current era of education, principals are being called upon as instructional leaders to share in the responsibility of raising student achievement. Research is declaring that if supervision and evaluation are completed in a manner that is consistent with modern education theory, they have the potential to elevate current levels of expertise among teachers (Danielson & McGreal, 2000). The concept of instructional leadership has become a vogue term and a buzz phrase that is frequently used in the field. One function of instructional leadership that principals engage in is the supervision and evaluation of teaching staff. Modern research demonstrates that evaluation instruments and procedures can be used as tools to help foster growth in teachers. However, while principals are attempting to make the transition from building managers to instructional leaders, many of the evaluation instruments and procedures that are utilized are left over from a previous era of education that emphasized school management over instructional leadership. Frequently, administrators seem to be giving lip service to practicing instructional leadership while continuing with evaluation practices that have been identified as a mutual waste of time. Mike Schmoker (1992) stated,
Research has finally told us what many of us suspected all along: that conventional evaluation, the kind the overwhelming majority of American teachers undergo does not have any measurable impact on the quality of student learning (p. 24).

The most widely used teacher evaluation models have been labeled as inadequate. In today’s culture, formative evaluation should be emphasized over summative evaluation, and new, comprehensive approaches should incorporate content valid instruments, conferencing, and reliable indicators (Heafele 1993).

**Research Questions**

The study aimed to measure principals’ attitudes and perceptions of the content and process of teacher evaluation through a lens of four constructs. The following research questions were explored:

1. What are principal’s perceptions of teacher evaluation according to four constructs of educational theory (partnership, ongoing, student learning, differentiated)?

2. Do differences in perceptions according to construct exist among principals who serve at different school levels?

3. Do differences in perceptions according to survey items exist among principals who serve at different school levels?

4. Do differences exist according to construct exist among male and female principals?

5. Do differences in perceptions according to construct exist among principals who have different amounts of experience in the field?
Hypotheses

The following null hypotheses for each of the research questions were developed.

\( H_{oa} \): Principals perceptions of teacher evaluation are not consistent with the four constructs of educational theory.

\( H_{ob} \): There are no differences in perceptions according to the constructs among principals who serve at different school levels.

\( H_{oc} \): There are no differences in perceptions according to survey items among principals who serve at different school levels.

\( H_{od} \): There are no differences in perceptions according to the constructs among principals of different genders.

\( H_{oe} \): There are no differences in perceptions according to the constructs among principals who have different amounts of experience in the field of education.

Target Population

Each of the 2,105 building principals in the state of New Jersey was targeted to participate in the study. Participants were selected by the use of public data from the New Jersey Department of Education website, which identifies principals by name, school, level, and title. Assistant principals, supervisors, and other administrators were not contained in the sample size. Emails sent to 160 of the 2,105 principals originally selected to survey were returned and marked as undeliverable. The unreachable principals had retired or were no longer employed in the school districts that were specified by the New Jersey Department of Education data. After subtracting the 160 unreachable participants, the sample size of principals that were surveyed was 1,945.
Definition of Terms

The following definitions are provided to ensure uniformity and understanding throughout the study. All definitions not otherwise noted have been developed by this researcher.

*Educational Leader* - One serving in a high ranking educational position with the mission of elevating others to higher levels of expertise, motivation, and morality.

*Formative Evaluation* - Feedback as part of an evaluative process that is designed to assist professionals to perform to higher levels of mastery.

*Modern Evaluation Theory* - Research stating that the primary goal of teacher evaluation is for the attainment of individual growth. The research states that effective evaluation systems consider a comprehensive view of teachers’ abilities and performance and are characteristic of the following: are formed in a partnership, are ongoing, include measures of student learning, and are differentiated for individuals.

*Summative Evaluation* - Feedback in the form of an evaluation that is provided to a teacher from an administrator for the purpose of judging levels of competence.

*Teacher Evaluation* - The process by which teachers are observed in a school setting and provided with feedback that is reflective of their performance. Evaluations occur during single lessons and are performed throughout the duration of a school year.

*Traditional Evaluation* - The processes by which teachers have historically been judged by administrators with the use of checklist instruments designed to rate the levels of observed or non-observed behaviors. May also be referred to as *teacher appraisal*. 
**Professional Significance of the Study**

One step to creating positive organizational change is to challenge a process (Kouzes & Posner, 2003). Implications for such a study concern local school boards and central office personnel who maintain local control over evaluation procedures in the state of New Jersey. Those who serve in such positions have negotiated with local teacher unions and the New Jersey Educational Association (NJEA) to utilize evaluation instruments that judge classroom lessons with checklists and rating scales. Board members often rely on evaluation instruments in the form of checklists that serve as easy reading and are marked with judgments regarding individual teachers and their performance on a given day.

The key point here is twofold: First, since board members are typically not educators, feedback in narrative form may not be easy for them to clearly understand. Evaluation instruments that read like checklists and rate lessons as Outstanding, Satisfactory, or Needs Improvement are easier to understand than prescriptive narratives or detailed rubrics. Secondly, instruments that quantify a teacher’s performance as Unsatisfactory simply have greater strength if the need arises to file tenure dismissal charges on that staff member. School board members often rely on evaluation processes as a management tool, despite current research that demonstrates the positive influence such procedures can have on teacher growth.

This study and others like it could potentially provide insight on how principals perceive evaluation in an attempt to assist stakeholders in adapting to more fruitful systems. The primary goal of teacher evaluation should be to assist teachers in elevating their practices to higher levels of expertise. Local school boards and superintendents in
the state of New Jersey have an opportunity to negotiate and implement improved and mandated supervision instruments and procedures. Another important thing such a study could reveal is that principals and administrators in New Jersey are in need of increased training in the area of supervision and evaluation.

Learning more about what modern educational theory indicates about supervision and evaluation could produce enhanced leadership skills in those who serve or aspire to serve as building principals and educational leaders. Applying research can be most difficult when attempting to impart change in an organization. The change process pertaining to teacher evaluation and observation could face resistance from multiple parties such as teachers, central office personnel, school board members, and building principals who are accustomed to the traditional process. However, to begin the change process, collective stakeholders may need to revisit the common educational mission. If the common mission is to improve the quality of teaching and learning in our schools, then a commitment should be made to implement processes and procedures that can better serve that mission.

District superintendents are positioned to work with local school boards, building principals, and teacher unions to implement comprehensive staff evaluation systems that can reflect accurate measures of performance while also initiating valuable prescriptive feedback that teachers need to improve. The most important step in improving the process by which teachers are evaluated and feedback is received is with the appointment of competent administrators who value the instructional leadership role of building principals. School principals must do more than pay lip service to the practice of instructional leadership while most of their time is dedicated to management tasks. This
study could potentially encourage local school boards and superintendents to increase the number of observations required for each teaching staff member while altering the templates and styles for evaluation. Principals who seek to serve as instructional leaders should be equipped with evaluation tools that can help them perform as catalysts for good instruction. For staff evaluation to be utilized as valuable procedure, the content of what is evaluated and the process of how it is completed must be focused on assisting teachers in the improvement of their craft.
Chapter 2: Review of Literature

Characteristics of Traditional Teacher Evaluation

Traditional supervision and evaluation call building principals to observe classroom lessons and record copious notes. Next, principals translate their notes to an evaluation instrument which prompts the evaluator to rate the lesson against a previously determined scale. Lessons are judged based on whether the teacher clearly stated the objective to students, whether the teacher correlated activities to learning objectives, and other observed or non-observed behaviors. After rating the teacher’s lesson in each category, the principal then offers some affirmative and prescriptive feedback as an addendum to the ratings and judges the lesson in its entirety with an overall rating such as Outstanding, Proficient, Needs Improvement, or Unsatisfactory. Perhaps the two constituents may meet to discuss the lesson and the evaluation. This practice and others like it occur in thousands of schools across of the country and have become vogue for at the least the past thirty years.

According to Marshall (1996) this process is largely ineffective and typically has little impact on the quality of teaching and learning in schools (p. 338). Traditional teacher evaluation procedures often cast teachers in the role of passive participants, who have little input into their evaluation beyond one or two brief meetings with the principal (Milanowski & Heneman, 2001). Research by Milanowski and Heneman (2001) described traditional evaluation procedures as an outdated system that is cumbersome and places little emphasis on improving instruction. Milanowski and Heneman (2001) stated,
“The single annual observation traditionally used to assess most teachers is more of a check to ensure minimally acceptable performance than a formative process” (p. 198). Rubrics are seldom designed to provide teachers with criteria referenced data as an assessment of teaching. As a result, ratings of Outstanding, Proficient, Needs Improvement, and Unsatisfactory are often arbitrarily given out by principals and reflect large amounts of human subjectivity (Medley & Coker, 1987). Despite the evidence that shows that adults respond primarily to positive reinforcement and desire to operate in a collegial environment, traditional teacher evaluation often violates these understandings while rendering teachers passive participants in the process (R. Brandt, 1996).

**An Absence of Leadership**

After his inspection of the current conditions of public schools, Richard Elmore (1999) published *Building a New Structure for School Leadership*. In it, he described how teachers are protected by an invisible barrier that discourages constructive scrutiny of instruction and supervision from outside inspection, interference, or disruption (Elmore, 2000). Elmore (1999) concluded that the school classrooms are also protected from supervision and instructional leadership, even in a culture which emphasizes the role of a school principal as an instructional leader. Hence, the problem was brought into a clearer focus. In the Midwest Region studies have recently shown that teacher evaluations frequently amount to summative reports that are used to support decisions about retaining teachers and granting tenure, rather than for professional development (Brandt, Mathers, Oliva, Brown-Sims, & Hess, 2007).

In a 1998 report, *Inside the Black Box*, which highlighted the importance of formative feedback for students, Paul Black and Dylan William asked how anyone could
be sure that a particular set of new inputs could produce better outputs if one has not at least studied what happened inside of the classroom (Black & Wiliam, 1998). In spite of this, administrators were found to formally evaluate teachers with minimal compliance while teachers continually focused on ratings and judgments rather then the prescriptive feedback for improvement (Danielson & McGreal, 2000). The conditions in Boston circa 1990 were described as preventative for school principals to imparting instructional leadership, and the conditions were reported to actually do harm to the school culture.

According to Marshall (2005), a condition of emptiness can be unintentionally created in the professional relationship between teachers and school leaders. Elmore (2000) commented on the isolated condition that exists between principals and teachers when he stated, “Direct involvement in instruction is among the least frequent activities performed by administrators of any kind at any level and those who do engage in instructional leadership activities on a consistent basis are a relatively small proportion of the total administrative force” (p. 17). Danielson and McGreal (2000) stated, “When people perceive that an environment is conducive to professional learning, then they see it as profoundly different from one that yields evaluative judgments” (p. 42). The researchers concluded that traditional systems need to be revamped and described them as burdensome and unhelpful for teachers who seek to improve their practice or for administrators who have to make difficult decisions regarding teachers’ performance (Danielson & McGreal, 2000). They asserted that teacher evaluation systems are erroneous in multiple layers and described what has come to be known as the Lake Wobegon Effect, “where most expert teachers expect to receive ratings of outstanding on their evaluations and that anything less, especially for experienced teachers would signal
a serious deficiency” (Danielson & McGreal, 2000, p. 4). Questions were also raised regarding administrator competence as the authors indicated that on average principals may have less content knowledge pedagogy to a specific subject compared to the teacher that they are actually responsible for observing (Danielson & McGreal, 2000). Additionally, some administrators may judge teachers unethically by providing them with favorable observations in a hope that the positive ratings will assist marginal teachers in transferring to another school. This unethical practice is known as the “dance of the lemons” (Danielson & McGreal, 2000, p. 5). The research also raises questions as to why novice teachers are held accountable to the same standards as veteran teachers and deems such conditions a convergence of pollution (Danielson & McGreal, 2000). The body of research and professional opinions clearly indicate disconnect between the type of educational leadership that is emphasized in major higher institutions and the actual imparted instructional leadership that is provided in schools by building principals. Schmoker (2006) named this condition a “leadership illusion” (p. 118). Despite massive evidence to the contrary, the prevailing assumption is that teachers learn most of what they need to know about how to teach before they enter the classroom. This limited view of what teachers need to know and do demands little educational leadership from administrators. Moreover, “when administrative work currently has little to do with the content of teaching, much less its improvement, it may actually act to protect teachers from various external intrusions on their isolated work” (Elmore & Fuhrman, 2001, p. 70). Simply stated, the position of many experts is that administrators are failing to provide teachers with the type of feedback that improves the quality of teaching and learning in schools.
**Thinking Win-Win**

Stephen Covey (1992), who authored *The Seven Habits of Highly Effective People*, dedicated an entire chapter to the habit of “Thinking Win-Win.” The idea is to develop systems, lines of communication, business strategies, and solutions that are beneficial for more than one party (Covey, 1992). However, if one were to compare Covey’s “habit” to the traditional teacher evaluation system, one could easily see how it was more often than not a “lose-lose.” For example, if a principal enters a teacher’s classroom and observes an instructional lesson that is weak in certain areas, the principal usually rates the lesson with numbers or symbols that correspond poorly to a pre-established checklist that illustrates the components of an effective lesson. The administrator then judges the lesson in its entirety as either Needs Improvement or Unsatisfactory. Following the completion of the rating scale, principals usually offer an additional narrative by commenting on what the teacher could do to improve the lesson the next time. The teacher focuses on the ratings when viewing the evaluation, becomes upset with the judgment, and either dismisses the comments as subjective or chalks the poor performance up to a bad day. The evaluation then gets filed away and the teacher’s instructional approach often remains unchanged (Marshall, 1996). If these teachers are tenured, typically the principal will not observe them again until the following year when the cycle is repeated. According to Covey, this cycle would be categorized as a “lose” because no progress or teacher growth was accomplished by the process. If one were to consider the opposite outcome of traditional evaluation, a similar lack of influence on performance could be observed. A principal enters a classroom and observes sound teaching that is closely aligned to the components of an effective lesson. The teacher
demonstrates good classroom management and is prepared. The principal judges the overall lesson as Outstanding and marks high numbers which correspond to the rubric. However, the principal also notes in the comments section of the observation some strategies that the instructor could employ to make such a lesson even better. In this case, the teacher usually reads the evaluation, and revels with, “Whew, I got an overall rating of Excellent,” and files the evaluation away (Marshall, 1996, p. 338). This case also demonstrates a “lose” because again the system and its process do not yield any improvement in future teaching practice. Principals are aiming for the mission of instructional leadership, but the current evaluation instruments and procedures simply fail to assist them in accomplishing that mission. Nevertheless, school boards and teacher unions cling to traditional evaluation processes that have little impact on teachers’ growth. The tools used by principals for evaluation help to foster the traditional role of a principal as a school manager. With the use of checklists and rating scales, principals measure teachers’ performance similar to the way an umpire would call balls and strikes in a baseball; no instruction is provided for the pitcher on how actually to throw strikes (form, balance, delivery, etc.). However, a qualifier for educational leaders is the ability to not only evaluate, but also elevate the practice of their constituents. Educational leaders should aim to be more closely aligned with the role of a coach who still holds the pitcher accountable for a high standard of performance, but also provides continuous and constructive feedback that aides the pitcher in accomplishing that task. If principals are to align themselves with the role of leader rather than manager, then the evaluation process and the tools that are used to complete the process must be changed in a way that can create a Win-Win, rather than the current Lose-Lose.
Views of Local School Boards

Human resource officials have argued that the following purposes should first be served by evaluation systems: screen out unqualified teachers, produce constructive feedback to practitioners, reinforce outstanding service, provide direction for staff development, provide evidence that will survive scrutiny, aid institutions in terminating incompetent persons, and unify teachers and administrators in their collective efforts to educate students (Haefele, 1993). Evaluations also aim to provide teachers with prescriptive feedback to guide their practices (Haefele, 1993). However, if one were to hold traditional teacher evaluation systems and current practices against Haefele’s criteria for supervision, most would fail miserably. Many teacher evaluation systems serve neither the accountability nor the professional development function (Danielson & McGreal, 2000, p. 9).

Local school boards have not readily adopted evaluation systems that link student learning on state assessments with teacher evaluation procedures. However, modern research has yielded more progressive theories and systematic rubrics that can have more profound effects on teaching practices. Nevertheless, school boards and policy makers demand technical evaluations based on a set of previously determined standards for teaching that often correspond to a checklist (Habermas, 1970). In a culture that emphasizes educational leadership, school leaders need to rethink the way communication occurs with teachers both inside and outside of the formal evaluation process which has proven to have little impact on the quality of teaching in schools.
Perceptions of Administrators

Conditions of teacher evaluation and feedback directly involve building principals who have the potential to understand what strategies might make a difference in improving teaching and student achievement (Senge & Lannon-Kim, 1991). This can be partly attributed to the interactions that administrators have with faculty and students, as well as additional internal and external stakeholders. However, delivery of feedback as part of the evaluation process has long been a point of contention. Educators have observed how teacher evaluation processes have evolved over time periods from simple end of the year checklists and summative narratives to more sophisticated clinical teacher evaluation models (Kersten & Israel, 2005).

Researchers Thomas Kersten and Marla Isreal (2004) surveyed 102 building principals in an effort to determine if principals perceived certain evaluation approaches to be more effective than others. Building administrators were asked to record the number of teachers they evaluated in a year and the average amount of time they spent per year on non-tenured versus tenured evaluations. They were also asked to rate the effectiveness of particular evaluation tools including summative checklists, summative narratives, pre-observation conferences, observation checklists, post-observation conferences, and portfolio reviews. Principals were surveyed on perceived benefits and impediments to such practices. The data indicated that principals believed that such evaluation systems are inordinately time intensive and preclude many other opportunities for school building leaders to work with faculty to improve classroom instruction (Kersten & Israel, 2005). The study also revealed an underlying problem with the culture of public schools which impedes the evaluation process as a tool for professional growth.
Administrators noted that teachers typically expected to receive excellent evaluations and resisted evaluation methods that deviated from the status quo. Some administrators indicated that they did not perceive school cultures as likely to embrace something new in evaluation systems and did not value the process as a tool for improvement, but rather something that the teacher and administrator were required to endure (Kersten & Isreal 2005).

**Missing the Mark**

Kedian (2006) claimed that the nature and extent of any learning that occurs as a result of teacher evaluation is uncertain. In theory, appraisal should be a dynamic interaction which involves ongoing reflection, exploration, risk taking, consultation, observation, and feedback. The process should result in negotiated goals that are linked to professional development. However, Kedian (2006) stated that the process rarely results in the desired purpose, and “evaluation becomes a high stakes activity which is characterized in many schools by being threatening and stressful” (p. 13). Contrary to being characteristic of a collegial relationship, evaluation often only consists of brief conversations that follow infrequent observations. Kedian (2006) contended that appraisal generally seeks to accurately ascertain a teacher’s level of competence while also leading to professional learning. However, in an attempt to achieve both purposes, traditional evaluation systems often succeed with neither (Danielson & McGreal, 2000).

**A View of Separation**

The conditions that Kedian (2006) described are similar to what has also been labeled as a dysfunctional marriage between formative and summative evaluation (Stanley & Popham, 1988). Although the functions of formative and summative teacher
evaluation are often closely linked, some have stated that “the blending of formative and summative teacher evaluation represented a grave conceptual error” (Stanley & Popham, 1988, p. 58). The researchers theorized that although formative and summative evaluations are important functions, “these two teacher evaluation tasks must be carried out separately by different individuals” (Stanley & Popham, 1988, p. 59). Marshall (1996) seemed to concur with this point of view and offered that “the basic problem is that teacher evaluation combines two conflicting tasks: Improving instruction and judging performance” (p. 338). Some have equated this separation between summative and formative evaluation as the difference between evaluation and supervision (Glanz, 2005). Glanz (2005) described how clinical supervision should be aimed for instructional improvement as a separate process that engages teachers in dialogue for the purpose of improving teaching and promoting student achievement. He concluded that evaluation should serve a different purpose of quality assurance (Glanz, 2005). To be effective, Glanz (2005) stated that clinical supervision should be divorced from the evaluation process. More progressive theory states that the incompatibility between the two goals can be overcome by designing evaluation systems characterized by clear evaluative criteria, the citing and weighing of evidence, neutralization of bias, and development of shared values about what constitutes good teaching (Danielson & McGreal, 2000).

Similar traditional theories also distinguish between supervision and appraisal. Mitchell, Scott, Hendrick, and Boyns (1998) observed, “Supervision as a process is aimed at teacher improvement and seeks to provide support for teachers unconditionally; appraisal, though also seeking teacher improvement, may also lead to termination, promotion, or transfer” (p. 115). Teachers could be confused if administrators performed
summative and formative evaluations as separate functions. Hypothetically, teachers could be left to wonder the purpose of a principal’s visits to their classrooms if the role of the principal as an observer has not been clearly established. Some assert that not only should appraisal and supervision be viewed differently, but each should actually be carried out by separate administrators so that the lines of communication and individuals are understood long before the evaluation procedures begin (Stanley & Popham, 1988).

**Teachers as Widgets**

Weisberg, Sexton, Mulhern, and Keeling (2009) authored *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, a report that reflected survey responses from 15,000 teachers, 1,300 administrators, 12 districts, and four states and painted a grim picture of traditional evaluation. The report concluded that current evaluation procedures are ineffective and that information on teacher performance is almost exclusively used for decisions related to teacher remediation and dismissal (Weisberg, Sexton, Mulhern, & Keeling, 2009). This is contrary to what research (Danielson, 2000) informed us about the primary purpose of evaluation. Danielson (2000) asserted that fostering individual growth among teachers should be the primary goal of evaluation systems. Weisberg, et al. (2009) stated that current teacher evaluation procedures were deficient in multiple layers including that most teachers (99%) received ratings of good or great, authentic excellence often went unrecognized, evaluation was not linked to professional development, probationary teachers received no special attention, and ineffective teaching was not being documented. The conditions described in *The Widget Effect* correspond to what has been previously labeled as a “leadership illusion” (Schmoker, 2006 p. 22). Despite research
that informed principals of the value of formative evaluation, many are not engaging in such practices consistently (Elmore, 1999). Evaluation procedures are failing to differentiate performance among teachers (Weisberg, Sexton, Mulhern, & Keeling, 2009). *The Widget Effect* (2009) described the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher:

This decade old fallacy fosters an environment in which teachers cease to be understood as individual professionals, but rather as interchangeable parts. In its denial of individual strengths and weaknesses, it is deeply disrespectful to teachers; in its indifference to instructional effectiveness, it gambles with the lives of students. A teachers’ effectiveness, the most important factor for schools in improving student achievement is not measured, recorded, or used to inform decision making in any meaningful way (p. 4).

Compared to other areas such as alternative certification, licensing exams, and charter schools, educational reformers have overlooked the power of revamping teacher evaluations which often comprise a single, fleeting classroom visit by a principal (Toch, 2008). Teachers are also evaluated on outdated criteria that bear little relevance to the learning process. Evaluations can often amount to checklists that contain items such as “Is presentably dressed” and “Begins class on time.” Following evaluations, many principals do not take the time to discuss the results with classroom teachers (Kennedy, 2008). Toch (2008) stated:

A host of factors—lack of accountability for school performance, staffing practices that strip school systems of incentives to take teacher evaluation seriously, teacher union ambivalence, and public education’s practice of using
teacher credentials as a proxy for teachers’ quality have produced superficial and capricious teacher evaluation systems. (p. 34)

**Historical Changes in Evaluation**

Such checklist type instruments can underestimate the complexity of teaching while evaluating teachers solely on observable behaviors. These traditional evaluation tools have roots in the 1940s, when administrators frequently relied upon the traits approach to teacher evaluation. Educators of this era believed that teachers who possessed certain traits were more likely to perform effectively, so their traits became the centerpiece items in a checklist format of teacher evaluation criteria (Danielson & McGreal, 2000). Such traits included teacher appearance, enthusiasm, emotional stability, and tone of voice. Modern research clearly indicates that there is no link between such traits and student learning outcomes.

The emphasis on supervision and evaluation as management and accountability functions was commonly practiced in the beginning of the Twentieth Century. As schools became larger in size and complexity, specialists were needed to supervise the greater range of subject areas to be taught, such as music, home economics, languages, and science. The hallmark of this period, from around 1900 through the 1920s, was the transfer of scientific principles of business such as control, accountability, and efficiency (Tanner & Tanner, 1987). The emphasis on measurement led to increased attention on direct classroom observation and data gathering, particularly through the use of an observation checklist, a tool commonly used today. Lucio and McNeil (1969) stated that this type of supervision was aimed to create some order from the chaos of the educational practices of that period. During this period, principals and supervisors were focusing on
evaluation and supervision as management functions (Lucio & McNeil, 1969). There was hardly any reference to the concept of principals as educational or instructional leaders during this period. Moreover, efficient management was viewed as synonymous with school leadership.

Following the 1940s, oversight of instruction began to be viewed as a form of guidance rather than direction of instruction (Tracy, 1995). This marked the first shift in evaluation from being solely viewed as managerial to slightly formative in nature. Some assert that this shift in supervision represented movement into the beginning of a human relations phase of administration. Beginning in the 1960s, a resurgence of the application emphasized control and accountability. Tracy (1995) described this time period as characterized by the use of complex observation systems to measure effective and ineffective teaching behaviors.

The popularity of Madeline Hunter’s (1994) model of supervision that is still practiced today demonstrates that the principles of this time period are still intact. The skills needed by principals to evaluate teachers in this phase were technical in nature as classroom observation was a point of emphasis (Tracy, 1995). However, sometimes the need for face to face interaction between principal and teacher was diminished (Sergiovanni, 2001). The emergence of clinical supervision focused on sustaining teacher and supervisor interactions in order to mutually solve classroom problems. Tracy (1995) stated:

The primary purpose of this was to assist pre-service and in-service teachers by having the supervisor and teacher analyze the teacher’s performance together. The assumptions were that a sustained cycle of assistance is necessary for
teaching to improve and that the analysis of teaching behavior patterns can lead to useful insights. Additionally, a positive teacher/supervisor relationship is viewed as important for effective supervision. The supervisor was required to be highly skilled in data collection, providing feedback, and relating to people (p. 320).

Goldhammer’s (1970) clinical supervision approach once again shifted the view of a supervisor to that of a coach more than a judge. This shift represented a major change in evaluation because for the first time teacher evaluation systems attempted to serve two purposes. Evaluation systems were developed in an attempt to measure teachers’ performance while assisting staff members in achieving growth. This humanistic view of instructional leadership, as opposed to the scientific method of evaluation, influenced other researchers such as Danielson and McGreal (2000) who later asserted that the primary function of evaluation should be to inspire growth, while making judgments should be secondary.

From the mid 1980s to the present, the supervision and evaluation process has evolved to become what is now referred to as the human development phase. Prior to this period, views of teacher evaluation swung like a pendulum over different time periods. Each side of the pendulum emphasized evaluation as either a management tool for accountability or a clinical tool for growth. Beginning in the 1980s and carried through to the present, evaluation systems have aimed to serve the purposes of both measuring teachers’ performance and simultaneously assisting teachers’ growth. During this time period, the idea that teacher evaluation should be differentiated for individuals because of their life stages and cognitive, conceptual, and personality factors became popular (Glickman, Gordon, & Ross-Gordon, 2001). In the human development phases,
there was no one method of evaluation that was viewed as the most effective. Moreover, evaluations are no longer viewed by modern theory as solely scientific or collaborative, but instead as comprehensive approaches that consider multiple factors. The development phase combines the concern for a teacher’s personal needs with the concern for the productivity of the organization (Tracy, 1995).

A variety of models have been developed to address teachers’ needs and provide performance judgments. An understanding of the history of teacher evaluation has influenced modern theory on the topic. Within the area of teacher evaluation, views about evaluation tools, along with the amount of time, format, and feedback, have changed dramatically from the past when the building administrator was viewed primarily as the school manager (Kersten 2005). The functions of teacher evaluation have evolved during the span of American education, with each new phase borrowing from the previous ones and adding its own contribution (Tracy, 1995). Analyzing these phases and the historical influences can assist in helping shape evaluation and supervision practices for the next century.

**Modern Evaluation Theory: A Call for Change**

The demand for improving the quality of teaching in our schools is stronger today than at any other point of our nation’s history. Research supports the idea of a school principal working as the catalyst for the necessary improvement. Principals are no longer recognized as managers, though they are often responsible for managerial duties. Principals and building administrators are now responsible as instructional leaders (Glickman, et al., 2001). The role of public school teachers is to partner with students to facilitate the learning process in a way that will help students close the gap between
novice levels of understanding and mastery levels, in reference to a predetermined body of criteria known as the state standards. To help maximize advancements in students’ learning, teachers choose to employ teaching strategies which aim to assist in the transfer of knowledge while helping pupils in the development of new skill sets and thinking processes. To support their constituents, school principals attempt to impart leadership in schools by providing teachers with the feedback that they need to refine and improve their craft. Feedback is often crafted by principals in the context of evaluation. When performance appraisal is utilized as a professional development tool, it can serve to provide staff members and principals with the necessary information to address situations that hinder performance (Kersten, 2009).

**The Value of Feedback**

The importance of feedback is not a new concept in the field of education and the manner in which students receive feedback has been established as critical in the field (Black & Wiliam, 1998). Although feedback is commonly known to influence performance, evidence shows that the type of feedback and the way it is given can be differentially effective (Hattie, 2007). After a quantitative meta-analysis was conducted, the data revealed that the effect sizes of formative feedback on the learning cycle could range from .26 to 1.35 (Black & Wiliam, 1998). Generally, feedback with the largest effect size was corrective in nature, meaning that its context provided students with an explanation of what they were doing that was correct and what they were doing that was incorrect (Marzano, Pickering, & Pollock, 2001). Absent from the published work was any reference to letter grades as a type of feedback. However, administrators, partly because of negotiated contracts between school boards and teacher unions,
continue to engage in formal evaluation procedures that do not encourage the exchange of important feedback between principals and teachers (Schmoker, 2006). Such procedures often rely on teacher ratings in various categories but are frequently absent of constructive and meaningful feedback.

Principals who have attempted to transition in their roles from traditional organizational managers to school leaders have not followed suit in the purpose of providing ongoing feedback for teachers (Elmore, 2000). Ironically, principals have directed teachers to provide their students with increased amounts of corrective feedback that can help bridge the gap between where students are and where they need to be concerning learning targets, while placing less emphasis on letter grades. However, principals are not doing the same for teachers. Principals observe non-tenured teachers only three times in the previously described process and tenured teachers only once in the state of New Jersey. However, principals consistently instruct teachers to provide students with continuous feedback relative to learning targets (Marzano, et al., 2001).

Research studies in this area point to the importance of prescriptive feedback in improving performance. Formative feedback has been identified as the most powerful single modification that enhances achievement (Black & William, 1998). Writing from the perspective of teacher-student feedback, Brookhart (2008) stated:

Students are less likely to pay attention to descriptive feedback if it is accompanied by judgments, such as a grade or an evaluative comment. Some students will even interpret “judgment” when the teacher intended description. Teachers should give students lots of opportunities to practice and receive feedback without a grade being involved. (p. 24)
If Brookhart (2008) is correct, couldn’t the same principles be applied to principal-teacher feedback? Are teachers not also learners in larger bodies? Could one not look at what research has established and the opinions of experts and conclude that students and teachers who receive high amounts of corrective feedback will make significant improvements? A teacher evaluation system should provide teachers with useful feedback on classroom needs, the opportunity to learn teaching techniques, and counsel from the principal on how to make classroom improvements (Boyd, 1989).

**An Ongoing Partnership**

One characteristic of effective evaluation systems is the presence of a partnership between observing principals and the teachers being observed. Historically, teachers were passive participants in the evaluation process. However, reform and restructuring initiatives have called for the changing of roles, responsibilities, and relationships between teachers and administrators (Danielson, 2000). A system’s procedures and practices allow or encourage what happens between teachers and administrators. Successful supervision and evaluation depend on the quality of what happens when the principal and the teacher get together (McGreal, 1983). Many of the variables necessary to make this one-one relationship productive revolve around the type of training given to participants and the attitudes they hold and display during their involvement in evaluation procedures (Darling-Hammond & et al., 1983). Thus, the degree to which the evaluator and teacher trust one another can ultimately determine the success of the evaluation process. Evaluation procedures that can produce teacher growth are rooted in relationships among professionals. To facilitate partnerships, principals should create structures for collaboration in schools and promote results-driven learning among staff.
(DuFour, 2002). The human relations between teachers and principals should not be underestimated (Sergiovanni, 2001). For teacher evaluation to be effective, it must be ongoing and consider much more than what happens in a classroom during a single lesson. Boyd (1989) stated:

Building administrators should review lesson plans and classroom records, and expand the number of people involved in evaluation. Most often principals or department supervisors conduct evaluations. Again, many state laws and collective bargaining agreements specify that teachers’ supervisors evaluate their performance. This system works well if the only goal of evaluation is to determine competence. If the goal of the evaluation is to promote growth, however, other evaluators should participate. Self-evaluations give teachers perspective on their work. Surprisingly, few school systems require self-evaluations. Peer and student evaluations, if schools administer them properly, can also benefit teachers. (p. 2)

Research indicates that if teacher evaluation systems are to be successful, teachers should be involved as active participants (Duke & Stiggins, 1986). Teachers must feel a sense of involvement within the internal workings of an evaluation system (Darling-Hammond, 1984). Donaldson (2001) stated that by creating an Action-in-Common, administrators and teachers can nurture shared beliefs, reinforced by shared experience and action, and together groups and organizations can act to accomplish goals more successfully than individuals can alone. Popham and Stanley (1986) wrote:

The most logical and practical method is to construct processes that encourage more administrator-teacher cooperation. Teachers should become actively
involved in data collection and feedback via collegial supervision, peer coaching, and teacher mentoring. Mechanisms should be built within a system that allow for more teacher participation as they work with administrators (p. 16).

Donaldson (2001) described evaluation as a collaborative process, focusing on the teacher’s professional growth designed to increase student success. Drawing on the work of teacher evaluation approaches in the Pajaio Valley Unified School District in California, Donaldson stated that a school leader’s purpose is to build relationships, clarify purposes, and facilitate Action-in-Common so that all people train their energies and talents on learning. Many structured collaborative opportunities existed in the school district. Probationary teachers collaborated with the building administrator and focused on developing and documenting the teaching described in the CSTP (California Standards and Teaching Profession) through classroom observation, teacher delivered portfolios, and face-to-face feedback. Donaldson (2001) stated that individuals who aspire to be leaders must engage purposefully in the web of relationships within the school building. Knowing one another well enough to establish basic trust, openess, and affirmation is a precondition for forming the relationships that can mobilize people for professional improvement and personal support. Donaldson suggested not only that partnerships between administrators and teachers must be created, but that partnerships among groups of teachers must exist to allow for the free exchange of ideas. On the topic of relationships and fostering connections among others, Donaldson stated:

The leader’s daily actions convey to others the belief that “we are in this together; your challenges and successes are ours and ours are yours.” By visibly
connecting with people and putting them in touch with others, the leader asserts an invitational, collaborative norm that says we depend on each other (p. 110).

Danielson (2001) advocated for teacher evaluation processes to foster the involvement of mentor teachers and peer coaches. Mentor teachers and peer coaches can play an important role in the larger evaluation system since teachers are more unlikely to be cordial in their professional discussions if they fear that the information could be used against them (Danielson 2001). Thus, partnerships must develop among groups of teachers in addition to teachers and administrators working together for growth.

Concerning the principal and teacher relationship, there has been a shift away from the traditional model of supervision in exchange for a more collaborative approach (Glickman, et al., 2001). Donaldson (2001) stated:

Leaders initiate this process by bringing to each individual and group a predisposition to trust and respect. They enter into conversations, meetings, and conferences believing that others will reciprocate if they are trusted and respected to begin with. Leaders who believe in the importance of working interdependently can, through conviction and persuasion, carry others toward similar belief and to the relationship that lies at its core (p. 116).

This belief stands in direct contrast to traditional supervision and evaluation models that have frustrated both principals and teachers by creating superior-subordinate relationships that emphasize conformity rather than growth and produce checklist data that is irrelevant to the curriculum (Downey, Steffy, English, Frase, & Poston, 2004).

When teacher appraisal is linked to professional development and a school improvement plan, the process can create a culture of trust and collaboration where each
individual understands his or her part in achieving established goals (Reddehopp, 2007). Carolyn Downey (2004), the creator of the Downey Walk Through Model of Supervision, emphasized the idea of a partnership between building principals and teachers by utilizing reflective dialogue. Her model changed the role relationship between teachers and principals by the use of reflective questions and dialogue to create an exchange of ideas between equals (Downey, et al., 2004). As dialogue develops among teachers and principals, relationships evolve from bureaucratic and legalistic to that of two professionals as partners in mutual pursuit of critical reflection regarding current practice.

There are other models that also emphasize the formative purpose of evaluation. Brandt (1996) has called for a new generation of formative evaluation models focused on teams of teachers evaluating their teaching and developing group instructional improvement plans consistent with school goals. This movement toward team evaluation and team-based instructional improvement is consistent with research findings that successful schools are categorized by collegiality and collaboration centered on discussion, critique, and improvement at teaching (Glickman, Gordon, & Ross-Gordan, 1998). Principal participation and leadership are keys to establishing a school-wide Professional Learning Committee (DuFour, 2002). Evidence is emerging that effective schools need collaborative principals. Deutschman (2007) demonstrated that the traditional combination of evidence, authority, and fear is insufficient to lead constituents. In corporate America, the results of top-down supervisors have not resulted in improved organization, but rather in wasted resources and burned out, scarred, or frustrated employees.
Engaging teachers as partners in the evaluation process has roots in what is known as clinical evaluation. Such a process elevates the formative and prescriptive components of evaluation over the summative and judgmental components. The term clinical supervision gained national prominence in the 1960s through the writings of Robert Goldhammer (Hughes & Ubben, 1994). The earliest roots of clinical supervision included a five-step process that occurred between the principal and teacher: a pre-conference observation, the classroom observation, analysis of the lesson, the post observation conference, and the post observation analysis. Clinical supervision elevates the cooperative problem solving component of school leadership, rather than an aversive inspection for staff deficiencies (Tanner, 1987).

This structure of clinical supervision influenced the modern work and research of Charlotte Danielson (1996) who emphasized the practice of teachers and administrators working together for growth. However, Danielson and McGreal (2003) have attempted to improve the traditional rating scales and appraisal designs to make them more useful as clinical tools. Danielson (1996) developed a framework that identifies aspects of a teacher’s responsibilities that have been documented through empirical studies and theatrical research. In the Danielson framework, the activity of teaching is divided into 22 components that are clustered into four domains: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities (Danielson, 1996). The comprehensive framework is designed to help teachers in each stage of their careers improve in their effectiveness while supplying evaluators with criteria for each domain and its components (Danielson, 2001). Instead of rating specific behaviors on a checklist,
the criteria help to build a common language and common understanding between evaluators and teachers that reflect or support actions (Danielson, 1996).

For supervision and evaluation purposes, the framework for professional practice can guide principals and teachers in conducting conversations about where to focus improvement efforts within the context of shared definition and values (Danielson, 1996). Connected to instructional supervision is the evaluation process that allows principals to lead their faculty through formative processes and also provides principals with a basis for personnel decisions (Danielson, 1996).

Research also demonstrates that in order for evaluation to be effective, it must encompass more than classroom observations. Wiggins (1989) stated that evaluation should rely on multiple sources that are collected over time and in diverse contexts. Kim Marshall (2008) elaborated:

Principals must guard against getting an inaccurate impression of teachers’ performance in glamorized lessons put on for the principals’ benefit. Year end teacher evaluations should never be based on a single classroom observation; to write fair and accurate evaluations, principals must make frequent unannounced classroom visits and draw on multiple sources of data to get a sense of what’s happening in classrooms during the 99.5% of the time when teachers are on their own with students (p. 2).

Evaluators can often have a difficult time understanding the rationale for what they observed in class in the absence of previously established open lines of communication between principals and teachers. Without procedures to discuss teachers’ rationale for classroom decisions, principals’ judgments of lessons could be rendered
without considering important information. Without dialoguing about lessons through the use of pre and post conferencing, the supervisors are forced to make evaluative decisions with a high degree of speculation and subjective interpretation. When these perceptions color principals’ overall evaluation of lessons, the condition is known as the *ghost behind the blackboard* (Wajnryb, 1991). However, when done systematically with open lines of communication, evaluations can serve as a powerful professional development tool for both teacher and supervisor (Dudney, 2002). Instead of being conducted as isolated classroom visits, the evaluation process should include regular class visits as the foundation for formative and individualized long-term planning of teacher development. The teacher evaluation process should be part of an ongoing cycle with ongoing communication between teachers and principals (Glickman, et al., 2001). Dudney (2002), while serving as the Polish Department Chair and Dean of the European and Latin American School at the Defense Language Institute, studied the conditions needed to establish a meaningful link between classroom observations and teacher development. He concluded:

Observations should always be announced and last for the duration of a teaching hour. The observation cycle should consist of a pre-observation session with the teacher, an observation, and a post-observation conference. The pre-observation should be used to negotiate the observation focus or the specific aspect of the teacher’s teaching on which the observer will primarily concentrate. Secondly, it should be used to decide on a data gathering method best suited to capturing information in support of the focus. The post-observation conference should take place no later than two or three days after the observation and should be devoted
to interpreting the collected data, providing the teacher with opportunities to expand on his or her decision making processes during the lesson, and begin to work out an action plan in preparation of the next observation (p. 4).

Dudney (2002) advocated for no written documentation of the evaluation to be finalized or distributed until after the three part process of pre-observation, observation, and post-observation had occurred. Documentation from the observed cycle can be carried out on an observation form capturing the main points from the process. Such a process allows for a partnership to develop among teachers and administrators. It allows for the pair to collaboratively define specific areas of focus that are based on individual teachers’ needs and abilities. The observed hour of instruction is not the end of the observation cycle but serves as a springboard for continued development (Dudney, 2002).

Mitchell (1998) contended that the evaluation process required long-term commitments to relationships, data gathering as a basis for decision making, and improvement of the process of teaching. Donaldson (2001) stated that effective principals understand how to properly engage in the web of relationships in school buildings to assert leadership (Donaldson, 2001). Engaging in such relationships may be the first step in building partnerships upon which supervision and evaluation can be founded. For teachers and principals to grow and develop the collegiality, confidence, and trust needed to overcome the fears and negative feelings regularly associated with appraisal and supervision, schools need to establish a framework focused on teaching (Dinham & Scott, 1998). Danielson’s (1996) efforts in Enhancing Professional Practice: A Framework for Teaching attempted to serve the purpose of striking the right balance that focused on growth while not losing sight of the supervisory function.
A Link to Student Achievement

Bill Sanders, formerly of the University of Tennessee’s Value Added Research and Assessment Center, concluded something that school principals have known for a long time: teachers make a difference in student achievement. After observing student growth over a three year period, he concluded that the most important factor affecting student learning is the teacher (Sanders & Horn, 1998). The immediate and clear implication of this finding is that seemingly more can be done to improve education by improving the effectiveness of teachers (Sanders & Horn, 1998). In the study, Dr. Sanders examined student growth in the area of mathematics from 3rd to 5th grade. He found that when students were placed with three high performing teachers in a row beginning in third grade, students scored an average of the 96th percentile on Tennessee’s statewide mathematics assessment at the conclusion of 5th grade. On the contrary, when students were placed with low performing teachers three years in a row, their average achievement on the same state assessment was a dismal 44th percentile (Sanders & Horn, 1998).

Teaching contains both artistic and scientific components to facilitate the transfer of learning. Marzano, Pickering, and Pollock (2001) established through a meta-analysis that specific instructional strategies, when implemented properly, can dramatically increase student achievement. The implication of both the Sanders study and the Marzano meta-analysis is that supervising and evaluating educators should consider much more than the act of teaching, namely the results of teaching (Tucker & Stronge, 2005). When comparing value added growth model traditional evaluation systems, results indicate that growth-orientated systems that facilitate a flow of performance
information back to the teacher are more likely to have a positive effect on teacher quality (Stiggins, 1986). However, traditionally this has not been common practice for principals.

The Educational Research Service (1988) concluded that 99.8 percent of principals rely on classroom observation as the primary source when evaluating teachers. An over reliance on classroom observation can be problematic because it often represents only a small sample of teaching performance (Medley & Coker, 1987). Studies have shown that four hours of observation would equal less than one half of one percent of a teacher’s time during a given year. The true fundamental flow in such an approach is the assumption that the presence of good practice during the classroom observation equates to the academic success of students (Tucker & Stronge, 2005). The research is not directing principals to discontinue the practice of classroom observation, but to consider measures of student growth as important criteria. Such methods create a balanced approach to teacher evaluation and involve an assessment of the act of teaching as well as the results of teaching (Tucker & Stronge, 2005).

The use of standardized student assessments enables schools to measure the impact that instruction is having on student performance. The use of value added models and the application of growth models that measure longitudinal change in student progress over time are becoming more widely relied on in the United States. Several case studies demonstrate how schools are taking advantage of this approach (Milanowski, 2004). While the quality of state and local assessments differ widely, the items on a well developed standardized assessment have been field tested for fairness by the application of statistical models (Oliva, Mathers, & Laine, 2009). School districts can have the
opportunity to examine the relationship between changes in student achievement gains, teachers, and schools (Sanders & Horn, 1998).

Examining summative standardized assessment scores is not the only way the school districts are linking teacher evaluation with student performance. The analysis of student work samples is an alternative method to serve the same purpose (Gearhart & Osmundson, 2009). The purpose of this method is to measure student progress and performance over time. By implementing a portfolio review of student work, administrators can make finer distinctions about the quality of teacher performance (Stronge & Tucker, 2003). The use of student work samples in a portfolio can aide administrators in identifying which elements of teaching directly impact student learning (Mathers, Oliva, & Laine, 2008). Wolf, Lichtenstein, and Stevenson (1997) identified important features of a portfolio. They asserted that a portfolio should contain examples of both student and teacher work, paired with captions and written commentary that explain and reflect on the content of the portfolio. Most school districts that require portfolios with the evaluation system view them primarily as professional development tools. By adding an assessment component with clear criteria, they can also be used for summative and evaluative purposes (Danielson, 2000).

In addition to providing meaningful feedback for instructional improvement, student achievement data can provide encouragement and a sense of gratification (Tucker & Stronge, 2005). Schmoker (1998) stated that examining data was useful in helping teachers generate intrinsic motivation to improve. Tucker and Stronge (2005) declared that teachers were responsible for not only teaching but also, to some extent, learning outcomes.
Given the research base, school districts across the country are reaching for ways to link student learning with teacher evaluation fairly. Fairness is an important thread of the process since students must share in the responsibility of learning. Without student participation, the learning process is not possible (Frymies, 1998). Because of the dual party responsibility, school districts need to carefully consider the complexity of using evidence of student learning for evaluative purposes. Much of the time, assessment data serves as a beginning point rather than an end point as a tool for evaluation. Tucker and Stronge (2001) stated:

The information standardized tests provide seem to be a good starting point for identifying students who have difficulty learning material or teachers who have difficulty teaching specific content. Diagnosing the precise problem and providing the needed assistance require professional understanding of the dynamics of teaching and learning. Standardized testing should not be used as a final judgment of failure or success, but as an indicator or source of information that educators can systematically analyze for patterns of strengths and weaknesses. (p. 34)

The Oregon Teacher Work Sample Methodology (TWSM) sought to find more authentic ways to assess teacher performance by examining student learning. TWSM was designed to portray the learning program of pupils on the outcome desired by a teacher and taught by a teacher over a sufficiently long enough period of time for a program in learning to occur. Similar to value added models (Sanders, 2000), TWSM required teachers to reflect on their own teaching by considering the learning achieved by students. However, if evidence on student learning is to be used for purposes of teacher
and school evaluations, measures used to collect this evidence must meet technical and ethical criteria (Sanders, 2000). Reliability, validity, freedom from bias, and fairness are concerns for connecting teacher assessment to student assessment (Andrews & Wheeler, 1994). Multiple measures of student performance should be considered (Stronge & Tucker, 2000). Student test scores should serve as one element and as part of multiple measures for teacher evaluation because research points to a variety of influences on standardized test scores other than teacher performance (Darling-Hammond, 1984). Oregon’s teacher licensure system requires prospective teachers to provide evidence of students’ learning during monthly instructional units they have designed (Schalock, 1998).

Virginia state law requires that the performance evaluation of instructional personnel include measures of student academic progress. School boards are responsible for developing procedures for use by division superintendents and principals in evaluating instructional personnel in the state of Virginia. Evaluation procedures aim to assess student academic progress and the skills and knowledge of instructional personnel, including, but not limited to instructional methodology, classroom management, and subject matter knowledge. Taking the lead from the Commonwealth of Virginia, the Alexandria School District attempted to revamp its current teacher evaluation procedures. Administrators and teachers sought to create an authentic portrait of a teacher’s work. Two of the components of the Alexandria City teacher evaluation system focused on goal setting and student achievement. When attempting to articulate desired goals both in the short and long term, teachers shifted in their thinking from what they attempted to accomplish as practitioners and focused on developing goals for their students. The
alternative strategy of goal setting placed less emphasis on the teaching and more emphasis on the learning. The academic goal setting process explicitly focused on student academic progress (Tucker & Stronge, 2005). The superintendent in Alexandria reported that the school system observed a paradigm shift in how teachers and evaluators thought about evaluation (Alexandria City Public School, 2000).

Danielson (2000) also advocated for teachers and teams of teachers to develop goals to create assessments that can measure and describe student learning. Products should include rationale, desired student outcomes, necessary materials, recommended teaching practices, and a plan for assessing student learning and evaluating the merit of the activity (Danielson, 2000). Eaker (2002) asserted that such reflective thinking could be beneficial when work groups meet together for a common purpose. When teachers engage in the use of self evaluation by maintaining a weekly journal and responding to written prompts, they are able to reflect on specific issues regarding teaching and learning that are valuable in guiding their practice (Baker & Shahid, 2003). Based on a collaborative analysis of the results, teachers should be searching for what they can do to improve student learning (Eaker, DuFour, & Burnette, 2002). Because teachers have often been left alone when planning instruction, executing the plan, and analyzing student work, a culture of dependency has frequently disempowered teachers and become a barrier to productive collaboration (Fullan & Hargreaves, 1996). One promising structure for facilitating the collaborative analysis of student work has been the formation of professional learning communities where teachers working together as a team can tap into existing capabilities and potential which make them more apt to flourish while working in a unit compared to working with an external trainer (Schmoker, 2006).
In addition to group analysis and reflection, a teacher engaging in self evaluation is also a required ingredient for achieving growth (Reeves, 2004). The reflective process is a necessary part of accountability. Teachers can distinguish between the popularity of teaching techniques and their effectiveness through the use of reflective thinking (Reeves, 2004). Such a process aids a teacher in becoming more diagnostic as a coach and mentor while focusing on student learning as a measure of accountability. Although the idea is controversial because students must take active roles in the learning process, measuring student learning outcomes and progress can actually increase teacher motivation. Accountability for learning enhances intrinsic motivation by commingling a sense of meaningfulness with a sense of competence and progress, the keys to maintaining that motivation (Wetherill, Burton, Calhoun, & Thomas, 2002).

Jacob and Lefgren (2008) stated that principals can distinguish the teachers who produce the largest and smallest standardized achievement gains, but demonstrate less ability to distinguish between teachers in the middle of the distribution. The research implies that judgments made by principals regarding teaching performance may not be as subjective as implied by teacher union officials who are generally opposed to differentiated evaluation procedures and performance pay. Some argue that teaching is too complex to be assessed; others contend that the evaluation tools are too subjective to be worthwhile or that the process is too haphazard to be meaningful (Mitchell, et al., 1998). Such objections are stated despite the fact that most parents and community stakeholders know clearly who the effective and ineffective teachers are. Mitchell, et al. (1998) asserted that evaluation procedures and system procedures play only a partial role
in making the evaluative process effective and claimed that the environment and the context in which evaluation occurred were equally as critical. They explained:

When we consider the differences that exist among schools and communities, between teachers and students, and between students and students, we begin to see how critical it is to understand the environment within which appraisal occurs. Discounting the environment may lead us to misinterpret our data (p. 24).

The rationale for implementing value added growth models rather than uniform proficiency standards as a measure of teacher performance is that students begin school years with different achievement levels, and these must be accounted for (Harris, 2005). Proponents of value added growth systems assert that student learning must be the touchstone by which teachers and teacher educators are gauged (Schalock, 1998).

**Differentiated Approaches**

There are also increasing amounts of literature that emphasize that evaluation procedures should be differentiated for individuals. Supervision in successful schools is a developmental function that increases teacher choices, stimulates teacher thinking, and encourages collective action (Ham, et al., 1994). In 1991, The Colchester Vermont Board of Education and the Colchester Education Association jointly initiated a Teacher Evaluation Study Committee who aimed to differentiate evaluation procedures for teachers. School based meetings were organized to introduce all teachers to the models of five differentiated components: (1) focused assistance, adapting to new contents; (2) focused assistance, improving current practice; (3) administrator consultation; (4) colleague consultation; and (5) self directed enhancement (Ham, et al., 1994). One goal of the study was to enhance staff growth while promoting increased collegiality, peer
consultation, and self-reflection. While this one school district in Vermont attempted to differentiate evaluation procedures, similar practices across the country have not become common. For the most part, when administrators evaluate novice or veteran, highly competent or incompetent, motivated or unmotivated teachers, they often do so with the same checklist or evaluative form for each of them. Danielson (2000) pointed out that among many professions, only teaching makes the same demands on novices as on experienced practitioners. The moment first year teachers enter their first classroom, they are held to the same standards and subjected to the same procedures as their more experienced colleagues (Danielson, 2000). Danielson (2003) stated the following:

Most other professions build in a period of apprenticeship. No one would expect a prospective surgeon, straight from medical school, to take charge of a complex operation. Nor would an architect be asked to design, single-handedly, a large office building. Yet the job of teacher for a novice is identical to that of a seasoned veteran, and the procedures used to evaluate them are identical (p. 5).

Danielson (2000) described how evaluation systems could be designed to differentiate procedures for teachers who were in different stages of their career or for those who demonstrated different needs to achieve growth. The beginning teacher program, the professional development track, and the teacher assistance track were proposed as separate evaluative tracks by which teachers could be supervised (Danielson & McGreal, 2000). To reverse the Widget Effect, Weisberg, et al. (2009) advocated for improving evaluation systems and a commitment to changing the culture of indifference to classroom effectiveness. To do so, a four part improvement plan it was suggested: (1) Adopt comprehensive performance evaluation systems that have clear performance
standards, monitor administrator judgments, and include frequent feedback for teachers; 
(2) Provide rigorous training and ongoing support for administrators so that they can 
make fair and consistent assessment of performance; (3) Increase the amount of influence 
that evaluations have by driving decisions about what teachers get tenure, how teachers 
are assigned, and how they are compensated; (4) Provide low stakes options for 
ineffective teachers to leave their positions without being exiled (Weisberg, et al., 2009).

**Dual Purposes Served**

While all of the research does not concur as to whether summative and formative 
evaluation should be blended, or whether supervision and evaluation are separate or 
joined functions, there is common theme regarding the theoretical background. 
Transformational leaders attempt to move colleagues and followers to higher levels of 
expertise and stimulate constituents to view issues from new perspectives. Effective 
principals instill in workers the desire to strive to make improvements toward the benefit 
of the entire organization. Educational leaders reframe problems so they may view them 
through a different lens, question assumptions, and in the process, stimulate followers to 
become more innovative and creative (Hoy & Miskel, 2002).

One of the problems that principals are continuing to focus upon is the application 
of the theory that instructional leaders can successfully engage in the clinical supervision 
process, while still fairly evaluating staff for accountability and quality assurance 
purposes. Principals are attempting to strengthen the link between teacher appraisals, 
instructional improvement, and student achievement. Regardless of whether the literature 
emphasizes the primary goal of evaluation as either summative or formative, researchers 
have searched for systems that could improve the quality of teaching. The key to
educational improvement lies in upgrading the quality of teachers (Darling-Hammond 1983). This responsibility often rests at the local level with building principals as instructional catalysts. However, teachers must function as willing partners (Glickman, 1985). Supervision that is characterized by collaboration, participative decision making, and reflective practice is necessary as part of a school improvement program that aims to promote teaching and learning (Glanz, 2005).

**Comprehensive Frameworks**

Comprehensive models of teacher evaluation use explicit standards, are based on multiple measures, and involve multiple evaluations (Toch, 2008). Danielson and McGreal (2000) have created alternative forms of teacher evaluation instruments in an effort to provide a more comprehensive approach. They have revamped the teacher evaluation process and procedure by dividing it into different domains (Danielson & McGreal, 2000). When adapting to such methods supervisors can engage in evaluation practices that are more formative and useful in shaping teachers’ practice. Screening out unsuitable candidates, dismiss ing incompetent teachers, and providing legally defensible evidence all are summative functions that can be met by the traditional instruments that emphasize accountability and quality assurance (Haefele, 1993). Providing constructive feedback, recognizing and reinforcing outstanding practices, providing direction for staff development, and unifying teachers and administrators around student learning are all formative functions. Such purposes were simply not being met by the traditional systems (Marshall, 2005).

Danielson’s attempt to revamp evaluation procedures has inspired more carefully crafted scales and more detailed rubrics with clearly defined criteria. Her work yielded
an improved formative function which corresponded to detailed rubrics which are classified in different domains: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities (Danielson, 1996). The modern system calls for principals to consider much more than the delivery of a lesson in isolation when evaluating performance and emphasizes ongoing evaluation throughout the teaching cycle of a school year. Because her work on teacher evaluation has clearly shown to better serve the educational mission of improving teaching and learning, her contributions make up much of what is referred to as modern educational theory in this study.

Danielson’s (2000) work in teacher evaluation is more progressive than traditional teacher assessments that are often technical in nature, infrequent, and yield judgments paired with little constructive feedback and limited ability to measure or monitor results. Danielson (2000) differentiated summative and formative evaluations and proposed that the two types have traditionally been in direct conflict with one another. Legislators and policy makers tend to value the summative purposes, and educators tend to think that the teacher evaluation should be designed for the improvement of teaching (Danielson & McGreal, 2000). Traditional teacher evaluation systems, to which many school districts are clinging, have been identified to actually serve each of the abovementioned goals rather poorly (Danielson, 2000). Danielson (2000) declared that a synergistic relationship between summative and formative evaluation can be developed.

Another promising method for measuring teaching performance and providing feedback has been the establishment of value-added models where gains from individual students are compared to the gains made in the previous year (Sanders & Horn, 1998). In the current accountability culture, state departments are producing value-added data that
can be clustered to measure and evaluate teacher performance. Average student gains are calculated at the teacher level to determine if expected student learning was achieved (Stronge, Tucker, & Hindman, 2004). The information can then be used to produce targeted feedback in the form of professional growth plans for teachers.

Marshall (1996) advocated for frequent classroom observations by building principals followed by focused conversations with teachers to provide feedback. Paying closer attention to teaching practices and their effects on student learning is a goal that many principals are now prioritizing without a clear understanding of how best to perform functions that will yield the desired improvements in teaching and learning (Reddekopp, 2007). Many corporations use the 360 degree feedback model to evaluate colleagues because this feedback process attempts to improve organizational performance by increasing the range of data included in employee appraisal (McFarland, 2001). Corporate managers view the 360 degree feedback structure as a welcome solution to the problems that plague traditional performance appraisal. The 360 degree approach utilizes data collected from individuals with whom the employee interacts, both vertically and horizontally, as well as data collected from self evaluation. Educational institutions may very well be able to derive benefit from models such as the 360 degree feedback structure. Such systems could improve appraisal by creating more of an ongoing process than an appraisal system based on isolated observations.

In Enhancing Professional Practice: A Framework for Teaching (1996), Danielson broke teaching down into four major categories, which she labeled as domains: Planning and Reaparation, Classroom Environment, Instruction, and Professional Responsibilities (Danielson, 1996). Accompanying these domains are 22 themes that
range from demonstrating knowledge to motivating students and providing feedback. Danielson (1996) also created rubrics for evaluations that detailed the performance behaviors needed to earn Distinguished, Proficient, Basic, and Unsatisfactory ratings in each skill category. Such a model illustrated more productive ways that teachers could be evaluated while providing a richer picture of teachers’ performance. Danielson’s work has been influential for educators and policy makers who desire to revamp the teacher evaluation process. In 1999, the Teacher Advancement Program (TAP) was created by using the work of Charlotte Danielson. The organization is currently operated by the California-based National Institute for Excellence in Teaching. TAP has made intensive instructional evaluations the centerpiece of a comprehensive program to strengthen teaching. Now in 180 schools in five states and the District of Columbia, the program encompasses 5,000 teachers and 60,000 students (Toch, 2008). In addition to Danielson’s criteria and standards, TAP also encompasses multiple measures, differentiation, and partnerships. Teachers are evaluated at least three times each year against a hybrid of Danielson’s teaching standards by alternate evaluators, including master and mentor teachers that are trained in the use of rubrics. The results from different evaluators and evaluations are utilized to determine an annual performance rating. Some procedures that principals can use as part of a comprehensive ongoing evaluation system are frequent visits to classrooms, the review of lesson plans and classroom artifacts, and the expansion of the number of people involved in the evaluation process (Boyd, 1989). Despite the demonstrated benefits of more comprehensive evaluation systems, many local, state, and national union leaders have not pressed for more rigorous evaluation systems for fear that such systems may result in the dismissal of
additional teachers for poor performance and may strengthen the case for performance pay at the expense of a single salary schedule (Toch, 2008). The single salary schedule has been labeled as the most significant barrier to improved evaluation systems. Kate Walsh, president of the National Council on Teacher Quality, stated, “If there are no consequences for rating a teacher at the top, the middle, or the bottom, if everyone is getting paid the same, then why would a principal spend a lot of time doing careful evaluation? I wouldn’t bother” (Toch, 2008, p. 34).

Many teacher unions make the counterpoint by arguing that the evaluations seriously require the single salary schedule (Toch, 2008). Principals are often left to sort out the tension between developing caring relationships while engaging in clinical and formative supervision and delivering on high levels of accountability. By developing a good understanding of pedagogy and curriculum along with good consulting skills, principals can deliver a combination of active listening, problem solving, and support (Donaldson, Marnik, MacKenzie, & Ackerman). Being equipped with effective supervision and evaluation procedures can assist in this process. Donaldson, Marnik, MacKenzie, and Ackerman (2009) advocated for principals working alongside teams of teachers prior to evaluation time. Principals need to sit with teachers as they analyze assessments, engage in professional development, and plan instruction (Donaldson, Marnik, MacKenzie & Ackerman, 2009).

Cohen, Raudenbush, and Ball (2003) defined teaching as what teachers do, say, and think with learners, concerning content in particular organizations and other environments over time. The classroom environment is created for students for the transfer of knowledge (Douglas, 2009). The extent to which teachers structure
cognitively demanding learning activities is also an important measure (Rowan, Jacob, & Correnti, 2009). Evaluation models should consider that some aspects of instruction are experienced differently by students in the same classroom (Connor, Jakobsons, Crowe, & Meadows, 2009). Therefore, comprehensive evaluation models that have the ability to include a full range of classroom dimensions are needed (Danielson, 1996).

In addition to Danielson’s (1996), other models are emerging that are complete with similar comprehensive evaluative structures. The ISI Classroom Observation System model references student characteristics such as language, self-regulation, and social interaction in addition to foundational dimensions of instruction such as management, emotional climate, and teacher knowledge (Connor, et al., 2009). Pianta and Hamre (2009) constructed an evaluative model that was closely aligned with ISI. Rowan and Corretti (2009) used a similar approach but added layers to highlight academic content and specific teaching practices such as cognitive demand and explicit instruction. In the model, teachers are required to keep teaching logs to report on time and content of instruction. The utility of logs for gathering more complex information, such as how teachers encourage students to engage in activities with high cognitive demand, has not yet been demonstrated. Since current evaluation procedures often are limited to the scope of a single classroom visit increasing the number of time points that data is collected is desperately needed (Marshall, 2000). Croninger, Valli, and Walters (2007) concluded that a given teacher had different levels of success using the same cognitive strategy for the same group of students from one day to the next. These conclusions clearly illustrated the need for ongoing evaluation (Valli, Croninger, & Walters, 2007).
Other research claimed that supervision and appraisal must be blended together in order to better achieve the higher purpose of improving teaching and learning in schools. Parkay, Stanford, and Gougeon (1996) stated that supervision and appraisal should be an integrative vehicle that considers staff development, curriculum and lesson planning, and teachers’ performance over time. They expounded:

An appraisal plan sets the focus for the whole process. The plan will establish the purposes, criteria, procedures, time line, and schedule for the evaluation process. It is important that relationships are established, understanding and acceptance of the process developed, and anxiety and apprehension relieved. Teachers have opportunities to raise questions and concerns about the process; and the evaluator lays the groundwork for positive interaction with the teacher and discusses the nature, time and frequency of his or her visits and the role he or she will play in helping the teacher. The manner in which feedback will be provided is also agreed upon.

The appraisal plan helps to develop two key conditions that teacher evaluation must be founded upon: (1) a partnership between evaluator and teacher and (2) an ongoing process. The appraisal plan is unlike a pre-observation conference in that the focus of the pre-observation conference is a teacher’s intention for a situational context (Dinham & Scott, 1998). Appraisal plans establish a partnership developing a common understanding of what the evaluative process will look like throughout the school year. Incorporating a time to communicate an appraisal plan and a pre-conference with teachers prior to a classroom visit are not the only elements that need to be utilized to develop strong partnerships. Following an observation period, evaluators should conduct
a feedback conference in which the teacher and appraiser discuss the data collected and the evaluator’s analysis of it (Scott, 1998). Scott (1998) stated:

Appraisers/supervisors must indicate not only weaknesses but also strengths, and they should offer specific assistance to help teachers improve. Where this approach is not taken, a golden opportunity is missed for improving rapport and for jointly exploring solutions to perceived instructional problems and developing focuses for staff development are lost. Together, appraisal and supervision represent a systematic approach to working with teachers in the teachers’ professional environment. One should not be considered without the other. The approach focuses on teaching as the major element for improving classroom practice. If administrators accept this focus, then staff development plans and teaching improvement will go hand in hand and administrators’ relationships with teachers will improve as teachers see the appraisal and supervision process as open, helpful, and democratic. (p. 169)

Scott (1988) asserted the importance of not only the necessary partnership between administrators and teachers, but also the link between appraisal and supervision. Both formative and summative components can work together within a comprehensive teacher evaluation framework. This view represents a shift in focus during the last twenty years. Prior to that, evaluations were used only to measure teacher competence (Mayo, 1997). Contrary to the traditional use of one-size-fits-all checklists, current theory calls for the development of mutually constructed goals on which the principal and the teacher agree. Mayo (1997) stated:
Developing criteria for goals in which the teacher and principal agree is the beginning of a needs assessment and necessary front end work for teacher development. The use of a pre-conference can serve in the development of such goals. A pre-conference and creation of clearly defined and agreed upon criteria, followed by a classroom observation can assist in the post-observation feedback process. The evaluation system at this point should provide teachers with useful feedback on student needs, the opportunity to learn new teaching techniques, along with counsel from principals on how to make changes in the classroom (p. 270).

The process of pre-observation discussion also helps to demystify observation for assessment and gives a sense of teacher and observer working together (Hughes, 2008). Hughes (2008) described the dangers of principals using generic observation forms accompanied by checklist type rating scales. Given the use of these instruments, teachers will teach to suit the score sheet rather than teaching to suit the students (Hughes, 2008). Dudney (2002) stated that effective teacher evaluations should be part of an ongoing cycle that includes a pre-observation session to negotiate the observation focus, an announced classroom visit, and a post-observation conference within two to three days to discuss the data that was collected during the observation. This approach to teacher evaluation aims to be systematic and purposeful while ensuring a meaningful link between observations and professional development based on the individual teacher’s needs and abilities.
Summary of Literature

Research in the area of teacher evaluation demonstrates a historical and ongoing progression of various views. During the first part of the 20th Century and into the 1950s, the primary purpose of evaluation was for teacher accountability. Some of the checklists and procedures from this era are still utilized today. Beginning in the 1970s, the idea of clinical supervision and evaluation was emphasized. The primary purpose of supervision became the growth of a teachers’ practice over time. During the 1970s and 1980s many school leaders embraced the idea and practiced clinical supervision. However, a disconnect developed between the ideas of clinical evaluation, which emphasized the formative function of improvement, and the previously used summative evaluation procedures, which emphasized accountability and judgment. During this time period much of the research indicated that the popular view was that both summative and formative evaluation were necessary, but should exist as separate components that function for different purposes.

Among the principals’ most powerful tools for school improvement and effectiveness are program and personal evaluation (Stronge, 1995). Charlotte Danielson (1996) suggested that a synergistic relationship could be attained for comprehensive systems to include summative and formative components that work together. This view is what separates the modern education theory on teacher evaluation from traditional views. More research and application of teacher evaluation procedures began to search for models that could serve both purposes since Danielson’s (1996) work, Enhancing Professional Practice: A Framework for Teaching. School districts are attempting to construct comprehensive evaluation systems that can produce both quality assurance and
professional growth. Modern research demonstrates that such systems should consist of an ongoing evaluative cycle and be founded in a partnership between teachers and principals. Additionally, evaluative models should be differentiated for teachers at different stages of their careers and be considerate of student learning outcomes. These are the characteristics upon which this study will be based.

Applying these principles and moving them from theory to application is a challenging endeavor for principals. Moving in a more efficient direction for teacher evaluation is much needed, due to the fact that despite all of the modern research, even beginning teachers today are typically evaluated two or three times per year, and experienced teachers are only evaluated once every two or three years (Brandt, 2007).

Ludwig and Raddear (1987) stated that mankind has been in search of three elusive goals: the Fountain of Youth, the Holy Grail, and the perfect evaluation system. Because schools are social systems with hierarchical structures, the context in which evaluation occurs must not only take into account the instructional setting, but also the social, psychological, and socioeconomic settings as well. There may not be a one-size-fits-all approach. Principals must contemplate the modern elements of evaluation theory in an effort to move each teacher forward while elevating overall school quality as the building educational leader. Data suggests that high levels of direct and continuing professional development can lead to observable differences in teacher practices (Hansen, 2001). Modern evaluation approaches include differentiated systems and multiyear cycles. They also call for teachers to take active roles through the use of portfolios, professional conversations, and student achievement evidence (Danielson, 2001). Modern theory
regarding teacher evaluation declares that a well designed system can effectively merge professional growth with quality assurance.

**Biblical Integration**

As Christians, we are called to be both stewards of the Lord and highly competent in our craft. God is looking for leaders who are both faithful to him and experts in their field. Psalm 78:72 illustrates this point when it references the life of David. It says that he led his people with “both a pure heart and skillful hands,” and the message is that how we serve, and the levels to which we serve are both important for Christians who are working for the harvest. Colossians 3:23 tells us, “Whatever you do, work at it with all your heart as if working for the Lord, not men.” Rethinking and revamping the way principals perceive and perform teacher evaluations will require much contemplation and hard work. However, principals must lead schools by challenging, not preserving, the status quo while aiming to help teachers on their journey in continuous improvement.

Similarly to Paul and his pressing towards his calling, principals must strive for excellence for themselves and for others under their guidance and supervision. Elevating others to higher levels of expertise, motivation, and morality is not an easy mission. To succeed, principals must have the courage and willingness to apply the passage of Ephesians 4:15 and “speak the truth in love.” Once principals are equipped with the knowledge and skill set to serve as school leaders, those tools need to be utilized to provide others with thoughtful, constructive, and valuable feedback. Engaging in these practices will assist others in closing the gap between how they are applying their talents and how God would want them to do so. Each teacher is equipped with God ordained talents. The Bible is filled with examples of leaders who were responsible as “fishers of
“men,” who were responsible for others under the direction of God and the Holy Spirit. God expects building principals and educational leaders to maximize teacher potential and growth, who in turn, can more positively influence students. Engaging in such a process may require a disruption to the status quo and a challenge to traditions. Nevertheless, as Christians we must understand that our purpose is not to conform to the world, but to seek to apply eternal truths that God reveals through his word and his nature through the process of education. Before attempting to impart positive change and improvement in schools, principals and others must seek and serve God as the top priority while being sensitive to the direction of the Holy Spirit. Psalm 2:8 implies that God’s desire to use us is linked to our desire to be whom God wants to be.
Chapter 3: Methodology

General Perspective

The problem statement that guided this study sought to assess the levels to which New Jersey principals’ beliefs are consistent with what modern educational theory states regarding the characteristics of teacher appraisal. A 20 question survey instrument was developed to determine principals’ perceptions. Chapter 3 will explain the research design, research questions, development of the survey instrument, administration of the survey, description of the population, and the methods of data analysis.

Research Design

This exploratory study utilized a non experimental quantitative approach and has yielded descriptive statistics. Non experimental research is common in the field of education and social sciences (Johnson, 2001). While principals’ perceptions of teacher evaluations could be examined using either qualitative or quantitative approaches, a quantitative approach was selected due to the type of survey instrument that was developed. The study focused on exploring principals’ perceptions of four constructs of teacher evaluation that were selected from educational theory. The constructs upon which this study is founded are that teacher evaluation should be founded in a partnership, ongoing, considerate of student learning, and differentiated for individuals.
Research Context and Participants

Each of the 2,105 building principals in the state of New Jersey were surveyed electronically. Participants were selected by the use of public data from the New Jersey Department of Education website which identifies principals by name, school, level, and title. Email addresses were entered manually into an electronic address book. Assistant principals, supervisors, and other administrators were not contained in the sample size. Emails sent to 160 of the 2,105 principals originally selected to survey were returned and marked as undeliverable. After inquiring with a few school districts, I discovered that these principals had retired or were no longer employed in the school district that was specified by the New Jersey Department of Education data. After subtracting the 160 unreachable participants, the sample size of principals that were surveyed was 1,945. After receiving a preliminary email that described the survey and the study and a second email as a reminder, 462 principals completed the electronic survey. The 23.8% response rate of principals participating in the survey ensured that the data collected could be interpreted as valid (Gay, 1996).

Research Questions

The study aimed to measure principals’ attitudes and perceptions of the content and process of teacher evaluation through a lens of four constructs. The following research questions were explored:

1. What are principal’s perceptions of teacher evaluation according to four constructs of educational theory (partnership, ongoing, student learning, differentiated)?
2. Do differences in perceptions according to construct exist among principals who serve at different school levels?

3. Do differences in perceptions according to survey items exist among principals who serve at different school levels?

4. Do differences exist according to construct exist among male and female principals?

5. Do differences in perceptions according to construct exist among principals who have different amounts of experience in the field?

**Hypotheses**

The following null hypotheses for each of the research questions were developed. For research questions 2 and 3, hypotheses were rejected at the .05 alpha level. For research question 1, conclusions were formed from examination of the mean scores and standard deviations for each of the constructs.

- **H_{oa}**: Principals perceptions of teacher evaluation are not consistent with the four constructs of educational theory.

- **H_{ob}**: There are no differences in perceptions according to the constructs among principals who serve at different school levels.

- **H_{oc}**: There are no differences in perceptions according to survey items among principals who serve at different school levels.

- **H_{od}**: There are no differences in perceptions according to the constructs among principals of different genders.

- **H_{oe}**: There are no differences in perceptions according to the constructs among principals who have different amounts of experience in the field of education.
Survey Development

No appropriate instrument for this study was found in the literature, so I sought a valid and reliable survey instrument to measure principals’ perceptions of the four constructs that were chosen from the theoretical background. The modern research that was reviewed in Chapter 2 suggested that teacher evaluation should be formed in a partnership, part of an ongoing cycle, considerate of student learning, and differentiated for individuals. These are the constructs by which the survey was created. Constructing clear questions that measure each of the constructs was an important task in order to create a valid survey (Ary, 2006; Fowler, 2002). Both of these authors provided guidelines for the development of a survey instrument. Ary (2006) specified 11 guidelines for the construction of survey questions. The following criteria were carefully considered as part of question construction and selection. Ary (2006) stated:

1. Questions should be short, simple and direct.
2. Questions should be understood by all respondents. Avoid technical terms.
   Asking a pilot group of respondents similar to the main study group to evaluate the meaning of questions is recommended.
3. Avoid questions that lead to ambiguous answers.
4. Avoid bias in the question wording.
5. Avoid questions that assume traits that might not be present in the sample.
6. Avoid leading questions.
7. Avoid psychologically threatening questions.
8. Avoid double-barreled questions that ask two questions in one.
9. Answer choices should provide all possible responses to a question for closed answers.

10. Keep the questionnaire as brief as possible. Respondents are more likely to answer completely and honestly if the survey takes a minimum of time to complete.

11. Ensure that respondents are appropriately knowledgeable to answer the questions.

Fowler (2002) identified five criteria about the instrument itself that complemented the issues raised by Ary et al. (2006). Assuming that the questions meet the preceding guidelines, self-administered surveys should also meet the following conditions:

1. The questionnaire should be self-explanatory.

2. The items should mainly involve closed answers.

3. Only a few forms of questions should be used.

4. The instrument should be visually uncluttered.

5. Cues for respondents to inform them of the next steps in the survey should be provided.

**Development of Questions**

After selecting the constructs from the review of research, questions were developed to measure various aspects of each construct. The questions were formulated after considering the criteria provided by Ary (2006). Table 3.1 presents an overview of the constructs and questions related to the literature.
Table 3.1

*Construct and Question Identification*

<table>
<thead>
<tr>
<th>Construct</th>
<th>Questions</th>
<th>Support from Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership</td>
<td>1. Teachers have a clear picture of the criteria that is used to evaluate them.</td>
<td>Danielson, 2000</td>
</tr>
<tr>
<td></td>
<td>11. Teachers should be provided with advanced notice of evaluative classroom visits.</td>
<td>Glickman, 2001 and Dudney, 2002</td>
</tr>
<tr>
<td></td>
<td>8. Principals should conduct pre-conferences with staff members prior to evaluative visits.</td>
<td>McGreal, 1983</td>
</tr>
<tr>
<td></td>
<td>10. The role of a principal in the teacher evaluation process should be more closely aligned with a coach than a judge.</td>
<td>Danielson, 2000</td>
</tr>
<tr>
<td></td>
<td>6. Principal-teacher relationships are impacted by the way feedback is presented to a teacher following a classroom observation.</td>
<td>Sergiovanni, 1994</td>
</tr>
<tr>
<td></td>
<td>14. Traditional evaluation practices have cast teachers and administrators into adversarial rather than cooperative roles.</td>
<td>Marshall, 1996</td>
</tr>
<tr>
<td>Ongoing</td>
<td>7. The degree to which teachers are affirming to students over time should be included in evaluations.</td>
<td>Cohen, Raudenbush, &amp; Ball, 2003</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td></td>
<td>9. Information gained through regular contact with staff should be used in evaluations.</td>
<td>Schmoker, 2003</td>
</tr>
<tr>
<td></td>
<td>2. Teacher evaluations include multiple sources of data.</td>
<td>Wiggins, 1989</td>
</tr>
<tr>
<td></td>
<td>4. Teachers should not be evaluated on the basis of one or two classroom visits.</td>
<td>Marshall, 1996</td>
</tr>
<tr>
<td>Student Learning</td>
<td>16. A focus on student learning should pervade teacher evaluation.</td>
<td>Danielson, 2000</td>
</tr>
<tr>
<td></td>
<td>15. Individual student growth should be measured as part of teacher evaluations.</td>
<td>Sanders, 1998</td>
</tr>
<tr>
<td></td>
<td>13. Teachers should be asked to provide evidence of student learning for evaluative purposes.</td>
<td>Sanders, 1998</td>
</tr>
<tr>
<td></td>
<td>19. Student assessments should be utilized as a measure of teacher effectiveness.</td>
<td>Tucker &amp; Stronge, 2005 and Oliva, Mathers, &amp; Laine, 2009</td>
</tr>
</tbody>
</table>
17. Student motivation should be measured as part of teacher evaluations. Wolf, Lichtenstein, & Stevenson, 1997

18. Teachers should have options within supervision and evaluation systems. Danielson, 2000 and Mayo, 1997

12. Evaluations procedures should be different for tenured and non-tenured teachers. Danielson, 2000

5. Teacher evaluation should be tailored to fit the individual. Danielson, 2000 and Weisberg, Sexton, Mulhern, and Keeling, 2009

20. Mentor teachers can be utilized as alternative evaluation personnel. Danielson, 2000

3. Different procedures for evaluation should exist for tenured teachers at different stages of their career. Danielson, 2000

Panel of Experts

A panel of experts was consulted to establish validity of the survey instrument. Patricia L. Haney is the Superintendent of Schools in the Logan Township, New Jersey School District. In a previous position as Director of Curriculum and Instruction in the Harrison Township, New Jersey School District, Ms. Haney played a role in assisting the district in transitioning from a traditional checklist style evaluation instrument to a criterion referenced and research based evaluation tool. Ms. Haney is well grounded in
research on the topic of teacher evaluation and was selected to serve on the panel of experts to provide feedback on the content of the questions.

Sharon Campanese is a retired college professor and high school English teacher. She previously taught in the College of Education at Rowan University and at South Philadelphia High School. Ms. Campanese was employed on the panel of experts to provide feedback on the wording of questions to ensure that they would be easily understood by principals. The survey was also submitted to the dissertation committee members comprising of Dr. Charles Schneider, headmaster of Lifeway Christian School in Centerton, Arkansas and adjunct professor at Liberty University, and Dr. Thomas Power, professor in the School of Psychology in Pediatrics at the Children's Hospital of Philadelphia. Dr. Power has experience in scale development and in the administration of psychometric surveys.

Kim Marshall, author of *Rethinking Teacher Evaluation and Supervision* (2009) and several other journal articles that were utilized as part of the review of research, was also consulted via electronic communication to review the constructs and questions. The expert review panel addressed the content of the survey, support of research, and criteria set forth by Ary (2006). The panel agreed that the four constructs were valid and the questions chosen as a measure of each construct were reasonable. The panel of experts did make suggestions concerning the wording and ordering of questions. More than one expert on the panel suggested that a question be added to *Construct 1, Partnership*. This addition is detailed below.
Survey Revised

The panel agreed that the original draft of the survey was solid and that only minor changes were needed. However, they suggested that a question be added to Construct 1 to capture the element of communication as a function of principal-teacher partnerships. As a result, the following question was added:

1. Teachers have a clear picture of the criteria that is used to evaluate them.

This question is supported by literature which states that evaluation should be based on mutually agreed upon and understood criteria (Danielson, 2000).

Pilot Study

Flink and Kosecoff (1998) recommend a pilot study, which allows for a final opportunity to refine survey questions for clarity prior to full administration. After the panel of experts reviewed the survey and provided feedback, revisions were completed, and the survey was administered to a consortium of Gloucester and Camden County, New Jersey administrators. Eighteen administrators, all of whom have the task of evaluating teachers in their local school districts, took the survey as a way to check for the clarity of the questions and predict variability in responses. Each of the administrators reported that the questions were relevant to teacher evaluation and clearly understood. The responses also demonstrated variability for most items.

Second Revision

Following the pilot study and analysis of responses, I decided to delete a question that did not gather much variability in principals’ responses. Question 2 of Construct 2 originally was formulated to read, “Principals should conduct frequent, informal walkthroughs as a way to evaluate teachers.” This question was deleted because each of
the principals in the pilot group consortium marked this question with a response of either “Agree” or “Strongly Agree.” Also, a few of the administrators in the pilot group were confused by the interpretation of the word “frequently.” I decided not to replace this question with another, partly due to the fact that a question had been recently added to another construct and the deletion and non-replacement of the question returned the total questions on the survey to the intended number of 20.

**Survey Administration**

Following the approval of the survey by the Internal Review Board at Liberty University, the survey questions were entered into the web based software program of SurveyMonkey and were accompanied by three additional questions that collected demographic information that would be utilized in the descriptive statistics and data analysis. The entire population of New Jersey building principals was surveyed electronically. Of the 1,945 New Jersey principals, 462 completed the survey. These surveys were fully anonymous and involved no interviewer bias, as the surveys were self administered via an email link. The data was collected by SurveyMonkey and entered into a Microsoft Excel Spreadsheet. The data from the Excel spreadsheet was then transferred into SPSS for conducting the statistical analysis.

**Data Analysis**

Theoretical constructs that are consistent with modern theory were incorporated into the development of the questionnaire. Four Constructs that were chosen as theoretical background from research on teacher evaluation were:

- Teacher evaluation should be formed in a partnership.
- Teacher evaluation should be ongoing.
- Teacher evaluation should consider student learning.
Teacher evaluation should be differentiated for individuals. For each construct, statements were developed using a Likert scale. Data analysis aimed to measure principals’ perceptions about evaluation as it related to school leadership. The data analysis also sought to compare principals’ perceptions about characteristics of evaluations from the perspective of different levels. Using a MANOVA, the mean scores collected compared perceptions of principals that served in four different levels of education: elementary, middle school, high school, and other. The fourth level, “other,” existed as an option for those who identified themselves in specialty levels in addition to the elementary, middle, and high school roles. Principals who identified themselves as “other” provided the following information in the open ended section of the demographics question: K-8, Pre-School, Special Education, Adult Education, 7-12th grade, Pre-K-8th grade, Post Secondary, Principal of two levels, Principal and Superintendent.

For each construct, statements were developed by using a Likert scale and were formulated with five response choices per statement. Each response choice from the Likert scale was paired with a corresponding numerical value with 5 representing the choice that was most reflective of the construct and 1 corresponding to the choice that was least representative of the construct. When participants completed the survey, the constructs were hidden and the order of statements was randomized. Data analysis sought to reveal whether or not principals agreed with, disagreed with, or understood key aspects of teacher evaluation. The dimensions of the teacher evaluation process that were measured were within the parameters of content, or what is evaluated, and process, or how it is completed.
Statistical Support

Two experts in the field of statistics were consulted to assist with the procedures involved in Chapter 4. Dr. James Lani earned a Ph.D., from Miami University of Ohio, in the field of Clinical Psychology and consults regularly with dissertation candidates for statistics support. Dr. Lani was employed to review the statistics procedures utilized, and assisted me in understanding the technical procedures involved regarding the ANOVA and MANOVA outcomes. Ms. Jeanine Delaney has a Masters Degree in Business Administration from James Madison University, and has career experience conducting survey marketing research. Ms. Delaney teaches college level marketing research at Rowan University in Glassboro, New Jersey. She assisted me with troubleshooting a few applications of the SPSS computer software program and reviewed the data and procedures used to check for accuracy. Ms. Delaney also provided consulted with me in making conclusions from the statistical analysis.

Reliability

The five point scaled scores for the 20 questions were averaged to generate an overall score. A Cronbach’s alpha was calculated for each of the constructs to compare the consistency of responses for each of the questions that were assigned to each construct. The alpha scores revealed how consistent principals were in their answers. For this reason, they indicated how well principals understood some of the key elements of teacher evaluation.
Summary

This purpose of this chapter was to state the hypothesis and to describe the procedures for the survey development, gathering data, description of participants, and data analysis. The following chapter will describe the results in the analysis.
Chapter 4: Results

Descriptive Statistics

Four-hundred sixty-two participants were involved in the study. Of these, 238 (51.5%) were male and 223 (48.3%) were female; frequencies and percents are provided in Table 4.1. The majority of participants, 137 (29.7%), had 6-10 years of administrative experience; frequencies and percents are provided in Table 4.2. The majority of participants, 246 (53.2%), worked at the elementary school level; frequencies and percents are provided in Table 4.3. Cronbach’s alphas for the four subscales of the Principal Survey are displayed in Table 4.4. Preliminary analysis KS tests were conducted to assess the assumption of normality for each of the constructs (partnership, ongoing, student learning, and differentiated) in questions 1-20 from each group (level, years of experience, and gender). The results of the KS tests revealed that dependent variables were not normally distributed. However, the MANOVA test is powerful enough that the results should not have been affected by the violation (Stevens 2002).

Table 4.1 Frequency and Percent for Participant Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>238</td>
<td>51.5</td>
</tr>
<tr>
<td>Female</td>
<td>223</td>
<td>48.3</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>Total</td>
<td>462</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Table 4.2  

*Frequency and Percent for Participant Years of Experience*

<table>
<thead>
<tr>
<th>Years of Administrative Experience</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5 years</td>
<td>94</td>
<td>20.3</td>
</tr>
<tr>
<td>6-10 years</td>
<td>137</td>
<td>29.7</td>
</tr>
<tr>
<td>11-15 years</td>
<td>97</td>
<td>21.0</td>
</tr>
<tr>
<td>16-20 years</td>
<td>48</td>
<td>10.4</td>
</tr>
<tr>
<td>More than 20 years</td>
<td>86</td>
<td>18.6</td>
</tr>
<tr>
<td>Total</td>
<td>462</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 4.3  

*Frequency and Percent for School Level of Present Employment*

<table>
<thead>
<tr>
<th>School Level</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary</td>
<td>246</td>
<td>53.2</td>
</tr>
<tr>
<td>Middle</td>
<td>79</td>
<td>17.1</td>
</tr>
<tr>
<td>High School</td>
<td>86</td>
<td>18.6</td>
</tr>
<tr>
<td>Other</td>
<td>48</td>
<td>10.4</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>0.6</td>
</tr>
<tr>
<td>Total</td>
<td>462</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Table 4.4

*Cronbach’s Alpha’s for Research Variables*

<table>
<thead>
<tr>
<th>Research Variables</th>
<th>α</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership</td>
<td>.468</td>
<td>6</td>
</tr>
<tr>
<td>Ongoing</td>
<td>.540</td>
<td>4</td>
</tr>
<tr>
<td>Differentiated</td>
<td>.581</td>
<td>5</td>
</tr>
<tr>
<td>Student Learning</td>
<td>.672</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4.4 illustrates that the Cronbach’s alpha reliability scores were calculated at less than the desirable level of .70. The alpha scores revealed that principals were not consistent in the way that they answered questions that were grouped within a similar construct. The manner in which principals responded to questions inconsistently affected my conclusions regarding research question 1. These conclusions are discussed in Chapter 5 in addition to the alpha scores for the ANOVA and MANOVA analysis for research questions 2, 4, and 5.

**Research Question 1**

What are principals’ perceptions of teacher evaluation according to four constructs of educational theory (partnership, ongoing, student learning, differentiated)?

**H₀**: Principals perceptions of teacher evaluation are not consistent with the four constructs of educational theory.

To examine research question 1, descriptive statistics were conducted on the survey responses to understand principals’ perceptions regarding teacher evaluations. Participants responded to items with a five-choice rating that were coded to a Likert scale for analysis: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5=
Strongly Agree. Means and standard deviations for all of the sample responses by item are presented in Table 4.5; means and standard deviations separated by level (elementary, middle, high, other) are presented in Table 4.6; means and standard deviations separated by years of experience (1-5, 6-10, 11-15, 16-20, > 20) are presented in Table 4.7; and means and standard deviations separated by construct (partnership, ongoing, differentiated, and student learning) are presented in Table 4.8.

**Table 4.5 Means and Standard Deviations on Survey Item for all Participants**

<table>
<thead>
<tr>
<th>Survey Item</th>
<th>N</th>
<th>Min.</th>
<th>Max.</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q01. Teachers have a clear picture of the criteria that is used to evaluate them.</td>
<td>434</td>
<td>1.00</td>
<td>5.00</td>
<td>4.16</td>
<td>0.81</td>
</tr>
<tr>
<td>Q11. Teachers should be provided with advanced notice of evaluative classroom visits.</td>
<td>421</td>
<td>1.00</td>
<td>5.00</td>
<td>2.80</td>
<td>1.05</td>
</tr>
<tr>
<td>Q08. Principals should conduct pre-conferences with staff members prior to evaluative visits.</td>
<td>431</td>
<td>1.00</td>
<td>5.00</td>
<td>3.55</td>
<td>1.02</td>
</tr>
<tr>
<td>Q10. The role of a principal in the teacher evaluation process should be more closely aligned with a coach than a judge.</td>
<td>431</td>
<td>1.00</td>
<td>5.00</td>
<td>4.08</td>
<td>0.96</td>
</tr>
<tr>
<td>Q14. Traditional evaluation practices have cast teachers and administrators into adversarial rather than cooperative roles.</td>
<td>418</td>
<td>1.00</td>
<td>5.00</td>
<td>3.17</td>
<td>1.06</td>
</tr>
<tr>
<td>Q07. The degree to which teachers are affirming to students over time should be included in evaluations.</td>
<td>425</td>
<td>1.00</td>
<td>5.00</td>
<td>3.99</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Q09. Information gained through regular contact with staff should be used in evaluations.  
Q02. Teacher evaluations should include multiple sources of data.  
Q04. Teachers should not be evaluated on the basis of one or two classroom visits.  
Q06. Principal-teacher relationships are impacted by the way feedback is presented to a teacher following a classroom observation.  
Q16. A focus on student learning should pervade teacher evaluation.  
Q15. Individual student growth should be measured as part of teacher evaluations.  
Q13. Teachers should be asked to provide evidence of student learning for evaluative purposes.  
Q19. Student assessments should be utilized as a measure of teacher effectiveness.  
Q17. Student motivation should be measured as part of teacher evaluations.  
Q18. Teachers should have options within supervision and evaluation systems.  
Q12. Evaluation procedures should be different for tenured and non-tenured teachers.
Q03. Evaluators should use different evaluation procedures for tenured teachers at different stages of their career.

Q05. Teacher evaluation should be tailored to fit the individual.

Q20. Mentor teachers can be utilized as alternative evaluation personnel.

Table 4.6

*Means and Standard Deviations on Survey Item by School Level*

<table>
<thead>
<tr>
<th>Survey Item</th>
<th>Elem.</th>
<th>Middle</th>
<th>High</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q01. Teachers have a clear picture of the criteria that is used to evaluate them.</td>
<td>4.14</td>
<td>0.84</td>
<td>4.20</td>
<td>0.95</td>
</tr>
<tr>
<td>Q11. Teachers should be provided with advanced notice of evaluative classroom visits</td>
<td>2.98</td>
<td>1.00</td>
<td>2.57</td>
<td>1.10</td>
</tr>
<tr>
<td>Q08. Principals should conduct pre-conferences with staff members prior to evaluative visits</td>
<td>3.66</td>
<td>0.97</td>
<td>3.51</td>
<td>1.17</td>
</tr>
<tr>
<td>Q10. The role of a principal in the teacher evaluation process should be more closely aligned with a coach than a judge</td>
<td>4.10</td>
<td>0.96</td>
<td>4.13</td>
<td>0.85</td>
</tr>
<tr>
<td>Question</td>
<td>Statement</td>
<td>Mean</td>
<td>Standard Deviation</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td>Q14.</td>
<td>Traditional evaluation practices have cast teachers and administrators into adversarial rather than cooperative roles</td>
<td>3.13</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>Q07.</td>
<td>The degree to which teachers are affirming to students over time should be included in evaluations.</td>
<td>3.99</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Q09.</td>
<td>Information gained through regular contact with staff should be used in evaluations.</td>
<td>3.84</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>Q02.</td>
<td>Teacher evaluations should include multiple sources of data.</td>
<td>4.42</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>Q04.</td>
<td>Teachers should not be evaluated on the basis of one or two classroom visits.</td>
<td>4.06</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>Q06.</td>
<td>Principal-teacher relationships are impacted by the way feedback is presented to a teacher following a classroom observation.</td>
<td>4.17</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Q16.</td>
<td>A focus on student learning should pervade teacher evaluation.</td>
<td>4.18</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Q15.</td>
<td>Individual student growth should be measured as part of teacher evaluations.</td>
<td>3.52</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Q13.</td>
<td>Teachers should be asked to provide evidence of student learning for evaluative</td>
<td>4.16</td>
<td>0.66</td>
<td></td>
</tr>
</tbody>
</table>
purposes.

Q19. Student assessments should be utilized as a measure of teacher effectiveness.  

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q17. Student motivation should be measured as part of teacher evaluations.  

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q18. Teachers should have options within supervision and evaluation systems.  

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q12. Evaluation procedures should be different for tenured and non-tenured teachers.  

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q03. Evaluators should use different evaluation procedures for tenured teachers at different stages of their career.  

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q05. Teacher evaluation should be tailored to fit the individual.  

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q20. Mentor teachers can be utilized as alternative evaluation personnel.  

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.7

*Means and Standard Deviations on Survey Item by Years of Experience*

<table>
<thead>
<tr>
<th>Survey Item</th>
<th>1 to 5</th>
<th>6 to 10</th>
<th>11 to 15</th>
<th>16 to 20</th>
<th>More than 20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Q01. Teachers have a clear picture of the criteria that is used to evaluate them.</td>
<td>4.14</td>
<td>0.78</td>
<td>4.08</td>
<td>0.90</td>
<td>4.23</td>
</tr>
<tr>
<td>Q11. Teachers should be provided with advanced notice of evaluative classroom visits.</td>
<td>3.00</td>
<td>1.13</td>
<td>2.74</td>
<td>1.04</td>
<td>2.63</td>
</tr>
<tr>
<td>Q08. Principals should conduct pre-conferences with staff members prior to evaluative visits.</td>
<td>3.68</td>
<td>0.99</td>
<td>3.50</td>
<td>0.96</td>
<td>3.48</td>
</tr>
<tr>
<td>Q10. The role of a principal in the teacher evaluation process should be more closely aligned with a coach than a judge.</td>
<td>4.40</td>
<td>0.75</td>
<td>4.03</td>
<td>0.93</td>
<td>3.89</td>
</tr>
<tr>
<td>Q14. Traditional</td>
<td>3.32</td>
<td>1.06</td>
<td>3.37</td>
<td>1.05</td>
<td>3.05</td>
</tr>
</tbody>
</table>
evaluation practices have cast teachers and administrators into adversarial rather than cooperative roles.

Q07. The degree to which teachers are affirming to students over time should be included in evaluations.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.84</td>
<td>0.71</td>
<td>3.94</td>
<td>0.76</td>
<td>3.98</td>
<td>0.73</td>
<td>4.27</td>
</tr>
</tbody>
</table>

Q09. Information gained through regular contact with staff should be used in evaluations.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.83</td>
<td>0.89</td>
<td>3.85</td>
<td>1.02</td>
<td>3.67</td>
<td>0.95</td>
<td>4.16</td>
</tr>
</tbody>
</table>

Q02. Teacher evaluations should include multiple sources of data.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.44</td>
<td>0.63</td>
<td>4.40</td>
<td>0.73</td>
<td>4.46</td>
<td>0.64</td>
<td>4.53</td>
</tr>
</tbody>
</table>

Q04. Teachers should not be evaluated on the basis of one or two classroom visits.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.07</td>
<td>1.00</td>
<td>4.09</td>
<td>0.91</td>
<td>4.04</td>
<td>0.84</td>
<td>4.27</td>
</tr>
</tbody>
</table>

Q06. Principal-teacher relationships are impacted

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.09</td>
<td>0.91</td>
<td>4.28</td>
<td>0.85</td>
<td>4.21</td>
<td>0.80</td>
<td>4.09</td>
</tr>
</tbody>
</table>
by the way feedback is
presented to a teacher
following a classroom
observation.

Q16. A focus on student
learning should pervade
teacher evaluation.

Q15. Individual student
growth should be
measured as part of
teacher evaluations.

Q13. Teachers should be
asked to provide evidence
of student learning for
evaluative purposes.

Q19. Student assessments
should be utilized as a
measure of teacher
effectiveness.

Q17. Student motivation
should be measured as
part of teacher
evaluations.
Q18. Teachers should have options within supervision and evaluation systems.

Q12. Evaluations procedures should be different for tenured and non-tenured teachers.

Q03. Different procedures for evaluation should exist for tenured teachers at different stages of their career.

Q05. Teacher evaluation should be tailored to fit the individual.

Q20. Mentor teachers can be utilized as alternative evaluation personnel.
Tables 4.5, 4.6, and 4.7 illustrate principals’ perceptions of the survey items numerically.

The data presented in Table 4.5 indicate the levels to which principals agreed or disagreed with individual survey items. The data in Table 4.5 considers all of the 462 principals’ responses. In Table 4.6, the principals’ responses are separated, and mean scores and standard deviations are presented to distinguish how principals of different levels perceived similar survey items. Table 4.7 separated the data according to principals’ years of experience so that comparisons could be made accordingly.

Table 4.8

*Means and Standard Deviations on Survey Constructs*

<table>
<thead>
<tr>
<th>Construct</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership</td>
<td>414</td>
<td>12.00</td>
<td>30.00</td>
<td>21.91/3.65</td>
<td>3.03</td>
</tr>
<tr>
<td>Ongoing</td>
<td>421</td>
<td>4.00</td>
<td>20.00</td>
<td>16.34/4.09</td>
<td>2.14</td>
</tr>
<tr>
<td>Differentiated</td>
<td>410</td>
<td>5.00</td>
<td>25.00</td>
<td>15.54/3.1</td>
<td>3.78</td>
</tr>
<tr>
<td>Student Learning</td>
<td>410</td>
<td>13.00</td>
<td>25.00</td>
<td>19.37/3.87</td>
<td>2.44</td>
</tr>
</tbody>
</table>

Table 4.8 displays data illustrating the levels to which the group of 462 principals agreed with the four constructs of this study. The data displayed indicate the levels to which the entire population of principals surveyed perceived constructs as agreeable or disagreeable. Principals’ responses were interpreted numerically and indicated that participants were agreeable to the construct of ongoing, ranged from neutral to agreeable in the constructs of partnership and student learning, and were neutral to the construct of differentiated. The larger mean score in each row represents the raw score mean score
that corresponds to the maximum number of points. Since the construct of partnership contained six questions, the total possible points were 30. The construct of ongoing contained four questions, and the total possible points were 20. Both differentiated and student learning contained five questions per construct, and the total possible points for each were 20. The smaller mean score reflects the raw mean score divided by the number of questions that were asked in each construct. This calculation was necessary to bring the mean scores into a similar scale so that inferences could be made. For the remaining tables, when the mean scores are displayed, both the raw mean scores and the adjusted means scores are included to avoid misleading statistics. The results from the Cronbach’s alpha reliability scales for each of the constructs indicated that within the constructs principals were not consistent in their responses to questions. The low alpha scores affirmed the theory that principals do not clearly understand construct characteristics nor do they understand the application of construct characteristics within the teacher evaluation process. Questions that were closely aligned to produce similar responses did not consistently yield the expected responses. For this reason and because the principals responses’ indicated only a mild agreement for three of the four constructs, this researcher failed to reject the null hypothesis for research question 1.

**Research Question 2**

Do differences in perceptions according to construct exist among principals who serve at different school levels?

\[ H_{ob}: \text{There are no differences in perceptions according to the constructs among principals who serve at different school levels.} \]
To examine research question 2, a Multivariate Analysis of Variance (MANOVA) was conducted to assess if differences existed on constructs (partnership, ongoing, student learning and differentiated) by level (elementary, middle, high school, and other). A MANOVA was appropriate because the four constructs of partnership, ongoing, student learning, and differentiated represented four dependant variables. The level of school at which principals served represented the independent variable for this question. The MANOVA calculated a total mean of the dependant variables (constructs), and determined if differences existed. The assumptions of homogeneity of covariance were assessed by Box’s M test of equality of covariance matrices. I assumed that the variance for each of the dependant variables was equal. To test this assumption, I utilized Box’s test and found the variances to be insignificant, so the assumption was not violated. The Wilk’s Lambda statistic was then used for the multivariate analysis. Wilk’s $\Lambda = .977$, $F(12, 995) = 0.73$, $p = .718$, multivariate $\eta^2 = .01$. Univariate ANOVAs are presented in Table 4.9 and suggest that no mean differences exist on an individual dependent variable by level. While the MANOVA sought to determine differences overall, the ANOVA was appropriate to help determine if differences existed according to each individual dependant variable. Means and standard deviations are presented in Table 4.10.

Table 4.9

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership</td>
<td>0.94</td>
<td>.420</td>
<td>0.01</td>
<td>0.26</td>
</tr>
<tr>
<td>Ongoing</td>
<td>0.41</td>
<td>.749</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Differentiated</td>
<td>0.54</td>
<td>.658</td>
<td>0.00</td>
<td>0.16</td>
</tr>
<tr>
<td>Student Learning</td>
<td>1.18</td>
<td>.317</td>
<td>0.01</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Table 4.9 is required because anytime that a MANOVA is conducted, presenting the results of the ANOVA is also appropriate. No differences were recognized and this researcher failed to reject the null hypothesis for question 2.

Table 4.10

*Means and Standard Deviations on Partnership, Ongoing, Student Learning and Differentiated by Level*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Elementary</th>
<th>Middle</th>
<th>High School</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>22.12/ 3.8</td>
<td>22.05/ 3.75</td>
<td>21.48/ 3.58</td>
<td>22.14/ 3.69</td>
</tr>
<tr>
<td>Ongoing</td>
<td>16.32/ 4.08</td>
<td>16.57/ 4.14</td>
<td>16.16/ 4.04</td>
<td>16.35/ 4.08</td>
</tr>
<tr>
<td>Student Learning</td>
<td>19.31/ 3.86</td>
<td>19.18/ 3.84</td>
<td>19.63/ 3.92</td>
<td>19.54/ 3.91</td>
</tr>
<tr>
<td>Differentiated</td>
<td>15.77/ 3.58</td>
<td>15.98/ 4.12</td>
<td>14.94/ 3.95</td>
<td>15.54/ 3.51</td>
</tr>
</tbody>
</table>

Table 4.10 is similar to Table 4.8 in that the data displayed relate to how principals perceived the four constructs. However, unlike Table 4.8, which considered the entire population of 462 principals, Table 4.10 desegregates the data according to the level of school at which the principals served.
Figure 4.1 displays that regardless of school levels, principals were consistent in their attitudes to the four constructs measured. No mean differences existed on individual constructs by level. The four shaded columns for each construct indicate that regardless of school level, principals responded with similar perceptions. This figure illustrates how different levels of principals responded for each construct. It should not be used to compare constructs because of the different scales for each construct. Partnership was a 30 point scale, ongoing was 20, and both differentiated and student learning were 25.

**Research Question 3**

Do differences in perceptions according to survey items exist among principals who serve at different school levels?

H₀: There are no differences in perceptions according to survey items among principals who serve at different school levels.
To examine research question 3, a Multivariate Analysis of Variance (MANOVA) was conducted to assess if differences existed on survey items 1-20 by level (elementary, middle, high school, and other). Research question 3 differs from research question 2 in that the four constructs are not considered. Each survey item in question 3 was examined independently. For this reason, Cronbach’s alpha scores do not limit the reliability of this question analysis. The survey items were examined independently and were not grouped according to a construct for research question 3. Similarly to the analysis for question 2, the assumptions of homogeneity of covariance were assessed by Box’s M test of equality of covariance matrices. Box’s test was not significant, and the assumption was not violated. The Wilk’s Lambda statistic was used for the multivariate analysis. Wilk’s Λ = .839, F (60, 1,075) = 1.09, p = .313, multivariate η² = .06. Univariate ANOVAs are presented in Table 4.11 and suggest that mean differences exist on Q2, Q5 and Q15 by level. I conducted a Scheffe post hoc test for the purpose of examining where the indicated differences were. The Scheffe post hoc test revealed that there were no differences revealed on Q2 (teachers should be provided with advanced notice of evaluative classroom visits) or on Q5 (traditional evaluation practices have cast teachers and administrators into adversarial rather than cooperative roles). However, principals in the “other” category did have a larger mean score compared to high school and middle school principals on Q15 (student motivation should be measured as part of teacher evaluations). This result was not helpful in making conclusions about the significance because principals who were contained in the “other” category represented schools of various types that did not fit into the traditional elementary, middle, or high school categories. Means and standard deviations are presented in Table 4.12. Because there
were no significant differences, this researcher failed to reject the null hypothesis for researcher question 3.

Table 4.11

*ANOVA*s on Survey Items 1-20 by Level

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q01. Teachers have a clear picture of the criteria that is used to evaluate them.</td>
<td>0.29</td>
<td>.835</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>Q11. Teachers should be provided with advanced notice of evaluative classroom visits.</td>
<td>3.76</td>
<td>.011</td>
<td>0.03</td>
<td>0.81</td>
</tr>
<tr>
<td>Q08. Principals should conduct pre-conferences with staff members prior to evaluative visits.</td>
<td>1.49</td>
<td>.218</td>
<td>0.01</td>
<td>0.39</td>
</tr>
<tr>
<td>Q10. The role of a principal in the teacher evaluation process should be more closely aligned with a coach than a judge.</td>
<td>0.29</td>
<td>.835</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>Q14. Traditional evaluation practices have cast teachers and administrators into adversarial rather than cooperative roles.</td>
<td>3.03</td>
<td>.030</td>
<td>0.02</td>
<td>0.71</td>
</tr>
<tr>
<td>Q07. The degree to which teachers are affirming to students over time should be included in evaluations.</td>
<td>0.06</td>
<td>.981</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Q09. Information gained through regular contact with staff should be used in evaluations.</td>
<td>0.47</td>
<td>.706</td>
<td>0.00</td>
<td>0.14</td>
</tr>
</tbody>
</table>
Q02. Teacher evaluations should include multiple sources of data.  0.72  0.543  0.01  0.20

Q04. Teachers should not be evaluated on the basis of one or two classroom visits.  1.04  0.373  0.01  0.28

Q06. Principal-teacher relationships are impacted by the way feedback is presented to a teacher following a classroom observation.  0.19  0.902  0.00  0.09

Q16. A focus on student learning should pervade teacher evaluation.  0.31  0.817  0.00  0.11

Q15. Individual student growth should be measured as part of teacher evaluations.  0.39  0.762  0.00  0.13

Q13. Teachers should be asked to provide evidence of student learning for evaluative purposes.  0.09  0.967  0.00  0.07

Q19. Student assessments should be utilized as a measure of teacher effectiveness.  0.93  0.427  0.01  0.25

Q17. Student motivation should be measured as part of teacher evaluations.  3.50  0.016  0.03  0.78

Q18. Teachers should have options within supervision and evaluation systems.  1.16  0.324  0.01  0.31

Q12. Evaluations procedures should be different for tenured and non-tenured teachers.  0.24  0.872  0.00  0.09
Q03. Different procedures for evaluation should exist for tenured teachers at different stages of their career.

Q05. Teacher evaluation should be tailored to fit the individual.

Q20. Mentor teachers can be utilized as alternative evaluation personnel.

Table 4.12

*Means and Standard Deviations on Survey Items 1-20 by Level*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Elementary</th>
<th>Middle</th>
<th>High School</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Q01. Teachers have a clear picture of the criteria that is used to evaluate them.</td>
<td>4.14</td>
<td>0.83</td>
<td>4.20</td>
<td>0.92</td>
</tr>
<tr>
<td>Q11. Teachers should be provided with advanced notice of evaluative classroom visits.</td>
<td>2.96</td>
<td>1.00</td>
<td>2.60</td>
<td>1.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
</tbody>
</table>

Q03. 1.23 0.297 0.01 0.33

Q05. 2.12 .098 0.02 0.54

Q20. 0.22 .885 0.00 0.09
Q08. Principals should conduct pre-conferences with staff members prior to evaluative visits.

Q10. The role of a principal in the teacher evaluation process should be more closely aligned with a coach than a judge.

Q14. Traditional evaluation practices have cast teachers and administrators into adversarial rather than cooperative roles.

Q07. The degree to which teachers are affirming to students over time should be included in evaluations.
Q09. Information gained through regular contact with staff should be used in evaluations.

Q02. Teacher evaluations should include multiple sources of data.

Q04. Teachers should not be evaluated on the basis of one or two classroom visits.

Q06. Principal-teacher relationships are impacted by the way feedback is presented to a teacher following a classroom observation.
Q16. A focus on student learning should pervade teacher evaluation.

Q15. Individual student growth should be measured as part of teacher evaluations.

Q13. Teachers should be asked to provide evidence of student learning for evaluative purposes.

Q19. Student assessments should be utilized as a measure of teacher effectiveness.

Q17. Student motivation should be measured as part of teacher evaluations.
Q18. Teachers should have options within supervision and evaluation systems.

Q12. Evaluation procedures should be different for tenured and non-tenured teachers.

Q03. Different procedures for evaluation should exist for tenured teachers at different stages of their career.

Q05. Teacher evaluation should be tailored to fit the individual.
Q20. Mentor teachers can be utilized as alternative evaluation personnel.

Table 4.12 illustrates the means and standard deviations that correspond to the ANOVAs in Table 4.11 for survey items 1-20. Including the means and standard deviations in the form of Table 4.12 is a necessary component of reporting an ANOVA and MANOVA that is consistent with APA style.

Figure 4.2. Bar graph on Survey Items 1-20 by Level

Figure 4.2 illustrates the manner in which principals from different school levels responded to individual questions. Both numerical and visual comparisons can be made from Figure 4.2.
Research Question 4

Do differences according to construct exist among male and female principals?

H₀: There are no differences in perceptions according to the constructs among principals of different genders.

To examine research question 4, a Multivariate Analysis of Variance (MANOVA) was conducted to assess if differences existed on constructs (partnership, ongoing, student learning and differentiated) by gender (male or female). Similarly to questions 2 and 3, the assumptions of homogeneity of covariance were assessed by Box’s M test of equality of covariance matrices. Box’s test was not significant, and the assumption was not violated. Wilk’s Lambda statistic was used for the multivariate analysis. Wilk’s Λ = .996, F (4, 378) = 0.34, p = .848, multivariate η² < .01. Univariate ANOVAs are presented in Table 4.13 and suggest that no mean differences exist on individual dependent variables by gender and this researcher failed to reject the null hypothesis for research question 4. Means and standard deviations are presented in Table 4.14.

Table 4.13

ANOVAs on Partnership, Ongoing, Student Learning, and Differentiated by Gender

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>F</th>
<th>Sig.</th>
<th>Partial η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership</td>
<td>0.16</td>
<td>.688</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Ongoing</td>
<td>0.26</td>
<td>.610</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Differentiated</td>
<td>0.59</td>
<td>.445</td>
<td>0.00</td>
<td>0.12</td>
</tr>
<tr>
<td>Student Learning</td>
<td>0.13</td>
<td>.722</td>
<td>0.00</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Table 4.14

Means and Standard Deviations on Partnership, Ongoing, Student Learning and Differentiated by Gender

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Partnership</td>
<td>21.09/3.52</td>
<td>22.03/3.67</td>
</tr>
<tr>
<td>Ongoing</td>
<td>16.28/4.07</td>
<td>16.39/4.10</td>
</tr>
<tr>
<td>Student Learning</td>
<td>19.47/3.90</td>
<td>19.28/3.86</td>
</tr>
<tr>
<td>Differentiated</td>
<td>15.67/3.13</td>
<td>15.53/3.85</td>
</tr>
</tbody>
</table>

Table 4.14 illustrates numerically that there were no differences perceived in the four constructs among male and female principals.

Figure 4.3. Bar graph on Partnership, Ongoing, Student Learning and Differentiated by Gender
Figure 4.3 visually represents that no differences among male and female principals were found according to construct.

**Research Question 5**

Do differences in perceptions according to construct exist among principals who have different amounts of experience in the field?

$H_{0e}$: There are no differences in perceptions according to the constructs among principals who have different amounts of experience in the field of education.

To examine research question 5, a Multivariate Analysis of Variance (MANOVA) was conducted to assess if differences existed on construct (partnership, ongoing, student learning and differentiated) by years of experience (1 to 5, 6 to 10, 11 to 15, 16 to 20, and more than 20). The assumptions of homogeneity of covariance were assessed by Box’s M test of equality of covariance matrices. Box’s test was significant, and the assumption was violated. However, the violation was corrected by the Pillai’s Trace statistic, which was used for the multivariate analysis. The results of the MANOVA were significant; Pillai’s Trace = .079, $F(16, 1,516) = 1.91$, $p < .051$, multivariate $\eta^2 < .02$. Univariate ANOVAs are presented in Table 4.15, and suggest that no mean differences existed on partnership, student learning, and differentiated. A significant difference was revealed for ongoing by years of experience. I conducted a Scheffe post hoc test to examine where mean differences occurred for the construct of ongoing by years of experience and found that the 16-20 years of experience group had a larger mean compared to all other years of experience ranges. This researcher rejected the null hypothesis for research question 5. Means and standard deviations are presented in Table 4.16.
Table 4.15

ANOVA on Partnership, Ongoing, Student Learning and Differentiated by Years of Experience

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>F</th>
<th>Sig.</th>
<th>Partial η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership</td>
<td>1.30</td>
<td>.269</td>
<td>0.01</td>
<td>0.41</td>
</tr>
<tr>
<td>Ongoing</td>
<td>3.38</td>
<td>.004</td>
<td>0.04</td>
<td>0.90</td>
</tr>
<tr>
<td>Student Learning</td>
<td>2.22</td>
<td>.066</td>
<td>0.02</td>
<td>0.65</td>
</tr>
<tr>
<td>Differentiated</td>
<td>0.19</td>
<td>.942</td>
<td>0.00</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 4.16

Means and Standard Deviations on Partnership, Ongoing, Student Learning and Differentiated by Years of Experience

<table>
<thead>
<tr>
<th>Variable</th>
<th>1 to 5</th>
<th>6 to 10</th>
<th>11 to 15</th>
<th>16 to 20</th>
<th>More than 20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Partnership</td>
<td>22.67/ 2.85</td>
<td>21.84/ 3.10</td>
<td>21.67/ 2.74</td>
<td>21.76/ 3.50</td>
<td>21.88/ 2.88</td>
</tr>
<tr>
<td></td>
<td>3.8</td>
<td>3.64</td>
<td>3.60</td>
<td>3.63</td>
<td>3.65</td>
</tr>
<tr>
<td>Ongoing</td>
<td>16.26/ 1.96</td>
<td>16.20/ 2.39</td>
<td>16.05/ 2.07</td>
<td>17.58/ 1.90</td>
<td>16.28/ 1.86</td>
</tr>
<tr>
<td></td>
<td>4.1</td>
<td>4.1</td>
<td>4.01</td>
<td>4.4</td>
<td>4.2</td>
</tr>
<tr>
<td>Student Learning</td>
<td>18.89/ 2.29</td>
<td>19.22/ 2.21</td>
<td>19.61/ 2.41</td>
<td>20.21/ 3.00</td>
<td>19.42/ 2.40</td>
</tr>
<tr>
<td></td>
<td>3.8</td>
<td>3.9</td>
<td>3.92</td>
<td>4.04</td>
<td>3.9</td>
</tr>
<tr>
<td>Differentiated</td>
<td>15.78/ 3.75</td>
<td>15.58/ 3.48</td>
<td>15.51/ 3.36</td>
<td>15.92/ 5.51</td>
<td>15.37/ 3.53</td>
</tr>
<tr>
<td></td>
<td>3.13</td>
<td>3.12</td>
<td>3.10</td>
<td>3.18</td>
<td>3.07</td>
</tr>
</tbody>
</table>

Table 4.16 displays the mean scores and standard deviations for principals’ perceptions of each construct according to years of experience. In Table 4.16, the mean score of 17.58, and adjusted mean score of 4.4 indicates that principals with 16-20 years of experience were more agreeable to the construct of ongoing that principals in any of the other sub groups.
Figure 4.4. Bar graph on Partnership, Ongoing, Student Learning and Differentiated by Years of Experience

In the ongoing column of Figure 4.4, a visual representation of the statistically significant higher mean score for principals in the 16-20 year sub group is displayed.
Chapter 5: Summary and Discussion

Introduction

The summary, conclusions, and recommendations of this study are presented in Chapter 5. This chapter is split into eight different sections: (1) Restatement of the Problem, (2) Relationship of Current Study to Previous Research, (3) Review of Methodology, (4) Summary of Results, (5) Discussion of Results, (6) Limitations of Study, (7) Recommendations for Further Research, and (8) Conclusions.

Restatement of the Problem

There is a disconnect between what educational research states about teacher evaluation and what is understood, practiced, and perceived as valuable by school principals. Modern theory clearly states that the primary purpose of evaluation is to foster individual growth among teachers and provide teachers with the feedback needed to refine and improve their craft. To assist teachers, proper evaluation systems should be characteristic of four important criteria: teacher evaluation should be formed in a partnership between principal and teachers, evaluation should be part of an ongoing process, evaluation should encompass measures of student learning, and evaluation should be differentiated for teachers based on their individual needs. In the current era, principals are called to be instructional leaders who act as catalysts for school improvement. However, while principals are attempting to make the transition from building managers to educational leaders, many of the evaluation instruments and procedures that are currently utilized are outdated. The instruments and procedures are
remnants of an era in education that emphasized school management over instructional leadership. Before principals can apply the research and perform as instructional catalysts, they must first understand and value the elements and dimensions that modern evaluation systems require. Furthermore, principals’ attitudes toward the necessary components of evaluation will determine the levels to which principals commit to the application of the research. The levels to which principals value the four constructs of teacher evaluation and the examination of differences in perceptions among subgroups is the focus of this dissertation.

**Relationship of Current Study to Previous Research**

Teacher evaluation processes have evolved over time from simple end of the year checklists and summative narratives to more sophisticated clinical teacher evaluation models (Kersten & Israel, 2005). Researchers Thomas Kersten and Marla Isreal (2005) conducted a qualitative study that focused on principals’ perceptions of evaluative approaches. Administrators were asked to record the number of teachers they evaluated in a year and the average amount of time they spent per year on non-tenured versus tenured evaluations. They were also asked to rate the effectiveness of particular evaluation tools including summative checklists, summative narratives, pre-observation conferences, observation checklists, post-observation conferences, and portfolio reviews. Principals were surveyed on perceived benefits and obstacles. The study found that principals believed that evaluation systems are inordinately time intensive and hinder many other opportunities for school building leaders to work with faculty to improve classroom instruction (Kersten & Isreal, 2005).
The study also found an underlying problem with the evaluation process being perceived as a tool for growth. The established culture of many public schools impedes the evaluative process as principals noted that teachers expected excellent evaluation ratings and resisted evaluative methods that deviated from the status quo. Some administrators indicated that they did not perceive school cultures as likely to embrace something new in evaluation systems and did not value the process as a tool for improvement, but rather something that the teacher and administrator were required to endure (Kersten & Isreal, 2005). The current study is similar to the one performed by Kersten and Isreal (2005), in that the focus was to reveal the perceptions of building principals in the area of teacher evaluation. However, unlike the previous study, this dissertation was quantitative in nature and attempted to illustrate numerical values that corresponded to principals’ perceptions of four constructs of evaluation theory.

Review of the Methodology

This study was quantitative in nature and has yielded descriptive statistics. The purpose of this study was to assess the levels to which principals’ beliefs are consistent with research based constructs of the teacher evaluation process. The primary hypothesis for the dissertation was that New Jersey principals’ beliefs and perceptions about the purposes of teacher evaluation are not consistent with the research based constructs and purposes of the teacher evaluation system. I also attempted to measure if elementary, middle school, and high school principals and principals with different years of experience perceived key constructs pertaining to evaluation differently.

The state of New Jersey currently has 2,105 school building principals. Each of the 2,105 building principals in the state of New Jersey was contacted electronically and
asked to participate in this study by completing a survey constructed of my own creation. The survey contained closed ended questions that yielded an objective numerical reality of principals’ beliefs. Theoretical constructs that are consistent with educational research were incorporated into the development of the survey. Questions were designed to measure principals’ perceptions for each of the constructs and survey items that align with the research reviewed in Chapter 2. Statements were developed using a Likert scale. A MANOVA was utilized to compare the perceptions of principals who serve in the field with different years of experience and in different school levels: elementary, middle school, and high school. Each response choice from the Likert scale was paired with a corresponding numerical value with 5 representing the choice that is most strongly reflective of the construct and 1 corresponding to the choice that is least representative of the construct. A panel of experts was employed to review the construction of the survey and the formation of the questions. The survey was also piloted before being electronically distributed to each New Jersey Principal. Four-hundred sixty-two principals participated in the study. SurveyMonkey was used to collect that data. Subsequently, the responses were encoded into an Excel spreadsheet for SPSS analysis. A Cronbach’s alpha was calculated for each of the constructs to establish reliability.

**Summary of Results**

The results from the Cronbach’s alpha reliability scales for each of the constructs indicated that within the constructs principals were not consistent in their responses to questions. The low alpha scores affirmed the theory that principals do not clearly understand construct characteristics nor do they understand the application of construct
characteristics within the teacher evaluation process. Questions that were closely aligned to produce similar responses did not consistently yield the expected responses.

The first example of how principals contradicted themselves in their responses can be found in question 8 and question 11. Question 8 read, “Evaluators should conduct pre-conferences with staff members prior to evaluative visits.” The overall mean score of responses was 3.55 on a 5 point scale. Question 11 read, “Teachers should be provided with advanced notice of evaluative classroom visits.” The overall mean score of responses was 2.8. This indicated that participants responded negatively when the element of advanced notice was introduced. When interpreting the results, one is left to wonder how principals and teachers could conference about a lesson that was to be observed if the teacher was not informed of the date and time that the observation would occur. Principals seemed to be reversed in their thinking when the element of advanced notice was introduced to the construct of partnership. One explanation for this inconsistency in responses is that principals do not embrace all of the necessary research behaviors that must be present for the teacher evaluation process to be viewed as a partnership between the observer and the teacher.

Inconsistencies in responses yielded lower Cronbach’s reliability scores. These inconsistencies further affirmed the theory that principals are not cognizant of the related behaviors within constructs. For example, Question 12 read, “Evaluations should be different for tenured and non-tenured teachers.” The overall mean score of responses was 3.49. Question 5 read, “Teacher evaluations should be tailored to fit the individual.” The overall mean of responses was 2.93. Further inconsistencies were found when analyzing responses to question 18, which read, “Teachers should have options within evaluation
systems.” However, in question 20, which proposed including mentor teachers in the evaluative process, the overall mean score of the responses was 2.62. While the inconsistency of responses within constructs yielded lower reliability scores for the comparisons among sub groups, the theory was confirmed.

Further analysis of responses found that principals felt most strongly that teacher evaluation should be part of an ongoing process. However, the participants did not affirm that teacher evaluation procedures should be differentiated. For the total population sampled, the scores in this category indicted that principals’ attitudes toward this construct were neutral. The constructs of partnership and student learning yielded responses between the neutral and agree range. When the MANOVA was calculated to compare principals’ responses by level and by years of experience, I found that principals with between 16 and 20 years of experience had higher mean scores for the construct of ongoing than any other sub group.

**Discussion of Results**

Research questions 2 and 3 were formulated to measure differences in perceptions among principals who serve at different school levels and have different years of experience. Detailed data analysis was performed, and no differences, with the exception of principals with 16 and 20 years of experience for the construct of ongoing, could be determined. The fact that the different groups of principals did not indicate differences in perceptions indicated that attitudes across groups were generally consistent. This finding is significant as it indicates that regardless of school setting, years of experience, or gender, principals across the state of New Jersey share similar views pertaining to teacher evaluation. Within the construct of the teacher evaluation system as an ongoing process,
there was strong agreement among principals that teachers should receive frequent feedback as part of the process. Principals also mildly affirmed that evaluation should be formed in a partnership and considerate of student learning outcomes. These scores demonstrated that principals across the state perceived that the overall purpose of teacher evaluation was to promote growth in teachers. Principals did not affirm that teacher evaluation should be differentiated for teachers based on years of experience and individual needs, and were neutral when responding to this construct

Limitations of the Study

Cronbach’s alpha scores for each of the constructs were lower than the acceptable level of .7. While the low alpha scores assisted this researcher in inferring information about principals’ overall understanding of each of the constructs, the scores caused the reliability of the analysis pertaining questions 2, 4, and 5 to be lower than anticipated. With the alpha scores at lower than desirable levels, the finding of no differences among groups of principals by level, years of experience, and gender is one limitation of this study. While the alpha scores were calculated according to each construct there was no overall calculation of a Cronbach’s alpha score for the survey instrument as a whole. The absence of an overall reliability score to measure internal consistency is a limitation of the study. An additional limitation of the study is found when comparing the total responses from principals in different school settings. There were a higher number of elementary principals who completed the survey. Two-hundred forty-six of the total number of participants (53.2%) were elementary principals. Seventy-nine of the total number of participants (17.1%) were middle school principals, and 86 of the total number of participants (18.6%) were high school principals. The higher response rate from
elementary school principals was proportionate to the higher number of elementary school buildings in the state of New Jersey. According to the State Department of Education’s vital statistics, approximately 63% of all schools in New Jersey, or at least 1,375, are considered by the state to be elementary schools. These schools encompass the grade level ranges; K-5, K-6, 4-6, 2-8, PreK-2, PreK-K, 3-6, and K-8. Middle schools make up approximately 19% of the total population of schools while four-year high schools represent 17%. There are 429 middle schools and 345 high schools in the state of New Jersey. One percent of schools are alternative and specialty schools. An additional limitation of the study was that while the sample size of 462 principals from a population of 2,105 was considered sufficient, it was not largely representative of the entire population. One reason that more principals may have not have elected to participate is that the survey was sent out electronically shortly before winter break in a time when principals are typically committed to a large number of responsibilities.

For various reasons the 23.8% of principals who participated in the study also may have been more inclined to take part in a study on teacher evaluation than those whom did not participate. The participants may have either been more progressive or traditional in their views compared to the majority of the principals whom did not participate. This researcher was also not able to identify individual characteristics and backgrounds of the principals who participated in the study. Key characteristics such as the institutions where principals received their training or common past experiences with local teacher evaluation procedures were unknown. Such characteristics and commonalities of the respondents could have influenced the data and analysis.
Recommendations for Further Research

I recommend further research that attempts to measure principals’ beliefs corresponding to research based constructs of the teacher evaluation system. Survey questions should be reworked and field tested with principals in different school settings to increase reliability. For the purpose of conducting a replication study this researcher recommends that an exploratory factor analysis be conducted to determine how participants’ responses aligned with one another on the survey. The factor analysis could lead to the deletion or regrouping of questions according to each of the constructs and may result in higher reliability scores. It may also be beneficial to establish an overall reliability score for the entire survey in addition to the four constructs. Formulating an overall measure of reliability could improve internal consistency.

This study was founded on a random sample of all New Jersey principals. To better validate the results it is recommended that further research be conducted by targeting a smaller random sample size that will yield a higher response rate. Research findings could also become more insightful if teachers were included in the study. By including both principals and teachers and measuring perceived barriers in applying the constructs of the teacher evaluation, richer data could be yielded for analysis. Further research is also recommended to determine the extent to which principals apply the constructs in everyday supervision and evaluation practices and to determine barriers that exist, such as collective bargaining agreements and/or time constraints. This research could identify potential obstacles in designing differentiated teacher evaluation systems.
Conclusions

Principals are responsible for helping teachers attain high levels of effectiveness and are to be catalysts in school improvement efforts. Educational research has demonstrated that evaluation systems that are rooted in partnerships, ongoing processes, inclusive of measures of student learning, and differentiated for teachers can assist principals in these responsibilities.

This study explored principals’ perceptions of teacher evaluation characteristics. The responses affirmed in theory that principals were neutral to agreeable that the evaluative process in schools should be a part of an ongoing cycle, inclusive of student learning, and rooted in administrator-teacher partnerships. The participants did not feel strongly that teacher evaluation systems should be differentiated. In an era where principals routinely direct teachers to differentiate instruction for students, principals were neutral to the construct of differentiating evaluations for teachers based on distinguishing needs. These findings were consistent among principals of different genders, in different school settings, and who had varying levels of experience in the field.

The results of this dissertation also affirmed the hypothesis that principals do not clearly understand the daily application of educational research pertaining to teacher evaluation. Principals responded positively toward theoretical constructs, but when responding to questions aligned with applying the constructs, participants’ responses were interpreted as confused or disagreeable to specific elements that play a key role within a construct. For principals and other school administrators who have the task of
teacher evaluation, increased training that includes a thorough study of the constructs and practical application of the key elements involved is strongly recommended.

Principals are in an ideal position to inspire, support, and lead initiatives that will improve teaching and learning. By partnering with teachers to enhance the teaching and learning process, committing to frequent and ongoing lesson observation, tailoring feedback to specific teacher needs, and incorporating already established growth models that measure student learning, principals can become transformational school leaders who affect the quality of the teaching and learning in our nation’s classrooms. Principals must balance their managerial duties with instructional duties that call for principals to work with a laser-like focus on student achievement. Managerial duties may be delegated to other school personnel to free the principal to do the work of an instructional leader.

School district leaders, school board members, and teachers’ unions across the nation need to work together to create a common understanding of how and why improved teacher evaluation systems can enhance student achievement efforts. Many questions will need to be explored, including the following: Other than principals, what school officials should be included in the evaluation of teachers? How many observations are needed to glean meaningful information and make accurate inferences about the quality of learning? How can school districts achieve inter-rater reliability? What agreed-upon growth measures of student achievement would be used in the process? Will teacher unions accept measures of student learning for evaluative purposes as reliable and fair? Should teachers be financially compensated for students’ growth on measures of achievement? The extent to which educational stakeholders explore the aforementioned questions can be instrumental in transforming schools across the nation.
Charles Swindoll (1987) authored a book titled, *Living Above the Level of Mediocrity*. The author pointed out that everything we deal with in life begins in the mind, and without and ability to see beyond the majority, one can easily fall into a comfort zone known as mediocrity (Swindoll, 1987). Christians who aspire to become educational leaders must confront the status quo through the perspective of another kingdom that is ruled by Jesus Christ. With eyes of faith, we must have the ability to see beyond the majority, so that the next generation of school leaders will enter the profession with improved evaluation systems that will better support teachers in the noble task of explaining life to the next generation.
References


Dudney, G. M. (2002). Facilitating teacher development through supervisory class observations. ED469715.


Appendix A

Principal Survey

a. Gender

Male  Female

b. Years of Administrative Experience

1-5    6-10    11-15    16-20

c. Please mark the level of school that you presently are employed.

Elementary  Middle  High School  Other ______________

1. Teachers have a clear picture of the criteria that is used to evaluate them.

Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

2. Teacher evaluations should include multiple sources of data.

Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

3. Evaluators should use different evaluation procedures for tenured teachers at different stages of their career.

Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree
4. Teachers should not be evaluated on the basis of one or two classroom visits.

   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

5. Teacher evaluation should be tailored to fit the individual.

   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

6. Principal-teacher relationships are impacted by the way feedback is presented to a teacher following a classroom observation.

   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

7. The degree to which teachers are affirming to students over time should be included in evaluations.

   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

8. Principals should conduct pre-conferences with staff members prior to evaluative visits.

   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree

9. Information gained through regular contact with staff should be used in evaluations.

   Strongly Disagree  Disagree  Neutral  Agree  Strongly Agree
10. The role of a principal in the teacher evaluation process should be more closely aligned with a coach than a judge.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

11. Teachers should be provided with advanced notice of evaluative classroom visits.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

12. Evaluations procedures should be different for tenured and non-tenured teachers.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

13. Teachers should be asked to provide evidence of student learning for evaluative purposes.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

14. Traditional evaluation practices have cast teachers and administrators into adversarial rather than cooperative roles.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

15. Individual student growth should be measured as part of teacher evaluations.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>
16. A focus on student learning should pervade teacher evaluation.

Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

17. Student motivation should be considered as part of teacher evaluation.

Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

18. Teachers should have options within supervision and evaluation systems.

Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

19. Student assessments should be utilized as a measure of teacher effectiveness.

Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

20. Mentor teachers should be utilized as alternative evaluation personnel.

Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree