School Administrators’ Perceptions of the James Stronge Teacher Evaluation System

Jean Schoenlank
jeanhelen@optonline.net

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

Part of the Educational Administration and Supervision Commons

Recommended Citation
http://scholarship.shu.edu/dissertations/2429
School Administrators’ Perceptions of the James Stronge Teacher Evaluation System

Jean Schoenlank

Dissertation Committee
Barbara Strobert, Ed.D., Mentor
Anthony Colella, Ed.D.
Lorna Oates-Santos, Ed.D.
William Ward, Ed.D.

Submitted in partial fulfillment of the requirements for the degree of Doctor of Education
Department of Education Leadership, Management, and Policy
Seton Hall University
2017
SETON HALL UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN SERVICES
OFFICE OF GRADUATE STUDIES

APPROVAL FOR SUCCESSFUL DEFENSE

Jean Schoenlank, has successfully defended and made the required modifications to the text of the doctoral dissertation for the Ed.D. during this Fall Semester 2017.

DISSERTATION COMMITTEE
(please sign and date beside your name)

Mentor: Dr. Barbara Strobert 9/22/17

Committee Member: Dr. Anthony Colella 9/22/17

Committee Member: Dr. William Ward 9/22/17

Committee Member: Dr. Lorna Oates-Santos 9/22/17

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate’s file and submit a copy with your final dissertation to be bound as page number two.
Abstract

School Administrators’ Perceptions of the James Stronge Teacher Evaluation System

This qualitative study examined school administrators’ perceptions of the James Stronge teacher evaluation system, one of five approved evaluation systems by the New Jersey Department of Education from the Teacher Effectiveness and Accountability for the Children of New Jersey Act (TEACHNJ) in 2012. Fourteen administrators from a suburban district were interviewed to examine their views of the Stronge model’s influence on teachers’ instructional practice, the accuracy of the ratings found from using this model, and its influence on their instructional leadership.

The results indicate the seven performance indicators at the heart of the Stronge evaluation system are valuable to administrators as they communicate with teachers regarding their instruction. The Instructional Delivery indicator within the observation process was overwhelmingly identified as a significant tool for administrators. Regarding accuracy of ratings, the administrators in this study found the evaluation process to be subjective despite the objective design of the system. Additionally, it was troubling to use one system as both a tool for growth and professional development as well as evaluation. Finally, as instructional leaders, the administrators interviewed for this study felt the Stronge system had not had a major impact on their approach to professional development and goal setting, although it was acknowledged that as part of an upper middle class suburban community with high student achievement results, they were already able to focus on professional development in a way that may be different from other communities.
Acknowledgments

Completing my dissertation was a long process during which I received a great deal of support, encouragement, and patience from many people. First and foremost, I need to thank my advisor, Barbara Strobert, for sticking by me for all this time. If not for her suggestion to change my topic after it was apparent that I would be forever stuck, I never would have completed my dissertation. She gave me the tough love I needed to finally just do it!

I want to thank my colleagues, some of whom completed this process before me and served as role models. They were always encouraging and optimistic that I could and would finish. Special thanks to Lorna and Bill for being part of my committee and offering me guidance and support.

My beloved Ridge School family has also given me encouragement and confidence along the way. Special thanks are owed to Peggy for her unfailing belief that I would get this done.

My family and close friends have walked the line of being encouraging without nagging, which I have appreciated. My brothers and sisters-in-law, Beth, Diane, Dana, and Leah have all known just what to say and, importantly, what not to say at the right times. Elissa has also walked that line with me all this time and finally resorted to bribery! The love and faith I felt from all of them was always present.

“Just get it done!” is now part of my past.
Dedication

This dissertation is dedicated to my parents. While they always believed (and told me often) I would succeed as an educator, I don’t think any of us ever imagined I would one day receive a doctoral degree. My father, Richard Schoenlank, was always my champion and believed I could do pretty much anything I set my mind to. My mother, Helen Schoenlank, was the single greatest influence on my choice of career and on my eventual development as a leader in education. Although I miss them both terribly, they have been on this journey with me.
Table of Contents

Chapter 1 INTRODUCTION
The Problem ................................................................. 1
Purpose of the Study .................................................. 4
Research Questions .................................................... 4
Theoretical Framework ................................................ 5
Design and Methodology ............................................ 5
Significance of the Study ............................................. 7
Limitations ................................................................. 8
Delimitations ............................................................. 8
Definition of Terms .................................................... 8
Summary ....................................................................... 9

Chapter 2 REVIEW OF THE LITERATURE
Introduction ................................................................. 11
Literature Search Procedures ....................................... 11
Criteria for Inclusion and Exclusion of Literature .......... 12
Theoretical Framework ............................................... 12
A Brief History of Teacher Evaluation ......................... 13
Teaching Quality ....................................................... 18
Common Characteristics of Teacher Evaluation Models ... 20
Perceptions of Evaluation by Administrators ................. 24
Teacher Evaluation in New Jersey ............................... 26
The Stronge Model ..................................................... 29
Influence of Teacher Evaluation on Performance .......... 31

Chapter 3 METHODOLOGY
Introduction ................................................................. 34
Background ............................................................... 34
Design ........................................................................ 37
Sampling ..................................................................... 37
Profile of the Site and Participants ............................... 38
Data Collection .......................................................... 40
Research/Interview Questions ..................................... 40
Data Analysis ............................................................. 43
Protection of Human Subjects and Ethical Considerations 44
Validity and Reliability ............................................... 44
Role of the Researcher and Researcher Bias .................. 45
Summary ..................................................................... 46

Chapter 4 FINDINGS
Introduction ................................................................. 47
Research Question 1 Themes ........................................ 48
Research Question 1 .................................................... 48
Findings ..................................................................... 48
CHAPTER 1
INTRODUCTION

The Problem

Teacher evaluation has become a focal point for education reform in the 21st century (Hull, 2013). Ideally, the process of evaluating teachers should create open, honest conversation and coaching between the teacher and supervisor, leading to professional growth, new perspectives, goals to improve the teacher’s craft, and ultimately student achievement; however, when teacher evaluation becomes part of national and state regulations, with funding attached, the “ideal” might be driven away by mandates and fear. In this study, administrators in a suburban New Jersey school district provide their perceptions of one state-approved evaluation system in order to gain an understanding of its value in furthering teachers’ performance and professional development.

Teacher evaluation has existed in some form since the beginning of public education in the United States. Like most issues in education, it has seen its share of trends, changes, and new initiatives. Perhaps the most stringent of these initiatives is taking place right now across the country with states utilizing money from the United States Department of Education’s Race to the Top Fund to overhaul teacher evaluation and tenure, among other things. “Race to the Top invited state leaders to put forward plans to improve not one or two isolated elements of their schools, but to develop and implement comprehensive statewide plans to improve entire systems” (USDOE, 2015, p. vi).

The state of New Jersey received Race to the Top funding and created the
Teacher Effectiveness and Accountability for the Children of New Jersey Act (TEACHNJ) in 2012. Included in the act are requirements for educator evaluation systems, other professional growth and development systems, and tenure decisions (NJ DOE, 2014). Since the 2013-2014 school year, teachers throughout the state have been evaluated based on a four-point rubric in order to further the development of all teachers while providing data-based evidence for eliminating ineffective teachers. Similar efforts are taking place in other states across the country.

Teacher evaluation has been part of public education in the United States as early as the 18th century, continually evolving and changing over the course of time (Burke & Krey, 1975). Researchers have been able to identify the philosophical underpinnings, historical flaws in the design of evaluation systems, and the ideal components for effective evaluation, although educators are still faced with uncertainty and disagreement when it comes to evaluating teachers in the early 21st century. With pressure mounting to hold teachers accountable for student achievement and gain competitive footing when comparing our educational system internationally, teacher evaluation is both mandated and under scrutiny due to funding grants offered by President Obama’s American Recovery and Reinvestment Act of 2009 and Race to the Top (RTTT) (Riordan, Lacireno-Paquet, Shakman, Bocala, & Chang, 2015). Despite all of this pressure and focus on evaluation, one best method for evaluating teachers has not yet been universally accepted.

In the state of New Jersey, the Teacher Effectiveness and Accountability for the Children of New Jersey Act was passed in August 2012. TEACHNJ mandated that school districts adopt one of several pre-approved evaluation systems in order to “raise student
achievement by improving instruction through the adoption of evaluations that provide
specific feedback to educators, inform the provision of aligned professional development,
and inform personnel decisions” (TEACHNJ Guide, 2014). By 2013, most New Jersey
school districts had adopted one of five evaluation systems or developed their own using
specific criteria from the Department of Education. The five most commonly used
instruments are Charlotte Danielson’s Framework for Teachers, James Stronge’s Teacher
and Leader Effectiveness Performance System, the Mid-Continent Research for
Education and Learning (McREL), Marzano’s Causal Teacher Evaluation Model, and the
Marshall Rubrics. Regardless of the system chosen, the goal for evaluation should be
universal; Linda Darling Hammond, preeminent researcher of teacher evaluation,
identifies an effective evaluation system as “part of a teaching and learning system that
supports continuous improvement, both for individual teachers and for the profession as a
whole” (2014, p. 5).

Not to be overlooked in the rush to overhaul teacher evaluation in New Jersey and
the 39 other states that applied for RTTT funding (Hull, 2013) are the administrators who
must conduct the observations and follow the often-intricate evaluation procedures
adopted by state departments of education. Since the turn of the 21st century,
administrators have experienced a dramatic increase in the amount of time and skill
required to implement new and more demanding teacher evaluation systems (Kersten &
Israel, 2005), even more so since the implementation of RTTT grants. Administrators
have their own unique perspective and front row view of evaluation, providing a largely
untapped resource for evaluation and education reform as a whole.
Purpose of the Study

The Stronge evaluation system is one of a handful of evaluation tools accepted by the state of New Jersey as part of TEACHNJ. The purpose of this study was to explore administrators’ attitudes and perceptions towards teacher evaluation specifically with the implementation of James Stronge’s Teacher Effectiveness Performance Evaluation System. Do administrators perceive it as a vehicle for professional development for teachers leading towards improved practice? Does it help all teachers become good and push good teachers to be great (Almy, 2011) as, ideally, an effective evaluation instrument should? A review of the literature revealed limited research regarding administrators’ perceptions of this specific model other than what James Stronge of Stronge & Associates has provided. Since approximately 11% of school districts in New Jersey are using the Stronge Evaluation Model (NJ Spotlight 2013), it is necessary to gain insight and feedback from people outside the Stronge organization. As a point of comparison, a great deal of research has been conducted on the Danielson Framework for Teaching model, used by approximately 85% of school districts in New Jersey (Mooney, 2013) and commonly used across the country as the default teacher evaluation framework (“Charlotte Danielson’s Framework,” 2011). The same cannot be said about the Stronge model. This study produces needed insight from the point of view of administrators into an evaluation tool that is used to support and evaluate teachers while also being used as a basis for personnel decisions.

Research Questions

The following questions were used to guide the qualitative research in this study:
1. To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System improves teachers’ instructional practices?

2. To what extent, if any, do administrators perceive that the Stronge Teacher Effectiveness Performance Evaluation System proficiency scale (Highly Effective, Effective, Partially Effective, Ineffective) is an accurate reflection of teacher quality?

3. To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System enhances their roles as instructional leaders?

**Theoretical Framework**

Reflective practice was the guiding theoretical framework for this study. Osterman and Kottkamp (1993) define reflective practice as “a means by which practitioners can develop a greater level of self-awareness about the nature and impact of their performance, an awareness that creates opportunities for professional growth and development” (p. 2). For the purpose of this study, administrators use reflective practice as a means to cultivate it within their teachers. With improved practice and achievement pushing the agenda for teacher evaluation, fostering an atmosphere that encourages and inspires reflection among teachers is an essential part of that process for administration.

**Design and Methodology**

After being granted permission from the superintendent of a suburban New Jersey school district, experienced administrators were invited to participate in this study in a letter of solicitation. The letter assured all participants of complete anonymity and
confidentiality. Those who responded with interest were sent a demographic questionnaire to ensure that a variety of levels and perspectives (elementary, secondary, district) were represented. A mix of levels and experience was tapped, identified from information gathered in the demographic questionnaire from the pool of volunteers. The researcher interviewed all qualified volunteers, a total of 14 from the pool of 24 potential administrators. The participants’ experience using the Stronge evaluation model ranged from two years (the minimum needed to participate in the study) to a maximum of five years, the number of years the Stronge model had been used in the district.

The study had a qualitative design, which lent itself best to examining administrators’ perceptions of an evaluation instrument. Data collection and analysis followed recommended procedures within the qualitative research format (Miles & Huberman, 1994). Interview questions were derived from each of the broad research questions in order to gain a thorough understanding of administrators’ perceptions. A jury of experts reviewed the questions and advised the researcher in order to ensure the quality of the questions and increase validity and reliability. The formal, scripted interview questions were part of each interview in an effort to identify common themes. The interviews followed a semi-structured format of approximately 20-30 minutes, allowing for follow-up questions and a thorough understanding of each participant’s responses.

Prior to the start of each interview, participants were provided with blank paper copies of the essential components of the Stronge model, such as the observation form, the summative form, and the seven standards used to evaluate teachers, to use for reference throughout the interview, raising the likelihood of common, accurate
vocabulary and terminology. The interviews were recorded with an audio device and then transcribed.

Following data collection, data analysis took place, examining the transcribed interviews for the purpose of establishing patterns or themes and reporting out a detailed description and interpretation of the research problem (Creswell, 2007).

**Significance of the Study**

Research confirms that teachers play a significant role in student achievement and even have the potential to overcome the influences of socioeconomic status on learning and achievement (Bembry, Jordan, Gomez, Anderson, & Mendro, 1998; Darling-Hammond, 2000). Meaningful evaluation of teachers is essential for identifying teachers’ strengths and weaknesses and providing ongoing professional development that promotes quality teaching and learning. Administrators, therefore, are under pressure to provide accurate evaluations of teachers to both guide teachers towards professional growth and guide administrators’ personnel decisions. While current research is available regarding teachers’ perceptions towards evaluation, research that focuses on administrators’ perspectives is less plentiful.

Although there is mounting research on Charlotte Danielson’s Framework for Teaching since it has become the default evaluation framework used by many states (“Charlotte Danielson’s Framework,” 2011), but scant research is available regarding the Stronge Teacher Effectiveness Performance Evaluation System. This study will provide needed perspectives of the Stronge evaluation system by administrators who have had experience using it.
Limitations

Limitations of the study were as follows:

1. Only one evaluation model was addressed in this study. Other models are used throughout the state of New Jersey.

2. The researcher has used the model in this study as an evaluative tool for teachers and therefore brings her own perceptions and biases to the topic.

3. Administrators were asked to provide their honest reflections and opinions regarding their use of the Stronge evaluation model.

Delimitations

Delimitations set the boundaries of a study, and the following were set:

1. The study was narrowed to a suburban New Jersey school district. The results of the study, therefore, are limited to the perceptions of the administrators in this district and may not be representative of all administrators using the Stronge evaluation model.

2. Administrators who used the Stronge evaluation model for at least two years were included in order to ensure they had experience evaluating teachers with this particular model and could share perceptions based on that experience.

3. Only the perceptions of administrators were used in this study; teachers’ perceptions were not sought.

Definitions of Terms

The following definitions provide clarification for terms specific to this study:

Administrator – An educator who is certified to supervise teachers in the state of New Jersey
Effective (3) – The standard for teachers to achieve within the Stronge teacher evaluation model. Other scores include Ineffective (1), Partially Effective (2), and Highly Effective (4).

Teacher Evaluation – The formal process a school uses to review and rate teachers’ performance and effectiveness in the classroom

TEPES – James Stronge’s Teacher Effectiveness Performance Evaluation System, the evaluation system on which this study is based

Value-added Measures – Used to estimate or quantify how much of a positive (or negative) effect individual teachers have on student learning during the course of a given school year

Summary

Chapter 1 of this study provided an overview of teacher evaluation from an historical perspective as well as the current status of teacher evaluation in the state of New Jersey. The purpose and significance of the study, as well as how this study can add to the literature, were described within the context of teacher evaluation in the state of New Jersey. Research questions were identified as well as the qualitative design and methodology used to gather data to inform the topic.

Chapter 2 provides a review of the relevant literature. The review starts with a brief history of teacher evaluation and then reviews the literature on teaching quality, common characteristics of teacher evaluation models, perceptions of evaluation by administrators, teacher evaluation in New Jersey, the Stronge evaluation model specifically, and finally the influence of teacher evaluation on performance.
Chapter 3 describes in detail the research design and methodology used to collect data related to the topic, including the researcher’s background, a profile of the site and participants, and a description of the data collection and data analysis. In Chapter 4 the findings for each of the three research questions and the interview questions are shared.

Finally, Chapter 5 offers conclusions of the study as well as recommendations. In addition to a summary of the major findings, the findings within each of the three research questions are analyzed. Based on those findings there are recommendations for school districts utilizing the Stronge evaluation system, recommendations for policy in the state of New Jersey, and recommendations for future research to extend the findings of this study.
CHAPTER 2

REVIEW OF THE LITERATURE

Introduction

This chapter provides a review of relevant research and literature related to teacher evaluation, both historical and current. Following a description of the theoretical framework that guides this study, the literature review begins with an examination of the history of teacher evaluation. A brief review of the influence of teacher quality on student achievement is examined for its relevance on evaluation. Next is a review of administrators’ perspectives of teacher evaluation systems, with special emphasis on those systems that have been implemented in the last several years. New Jersey’s evaluation system mandate under the TEACHNJ Act and the Stronge evaluation system specifically are included for review as they relate directly to this study. Finally, the researcher examined the influence of teacher evaluation on actual teacher performance.

Literature Search Procedures

The review of relevant research and literature related to this study was primarily conducted using the Seton Hall online library search engine. It was extended to include EBSCOhost, Educational Resource Information Center (ERIC), and Google with a focus on peer-reviewed educational journals, reports, dissertations, and books relevant to the topic. The New Jersey Department of Education website was also utilized. Given the nature of the topic, current research was given priority with a focus on the past decade.

The following key words were entered in various combinations to identify relevant literature: teacher evaluation, student achievement, James Stronge, Stronge
evaluation system, administrator perceptions and attitudes on teacher evaluation, New Jersey teacher evaluation, TEACHNJ Act, and Race to the Top.

**Criteria for Inclusion and Exclusion of Literature**

Literature that helped clarify historical and current perspectives on teacher evaluation was included in the literature review. Gaining a historical perspective regarding the evaluation of teachers was essential in order to frame the reform efforts taking place in the last decade. Sources were limited to peer-reviewed research, books, and reports by reputable researchers, education organizations, and departments of education.

**Theoretical Framework**

This study hinged on the reflective practice of administrators as they considered the questions regarding the Stronge teacher evaluation model and its impact on their teachers, as well as on their own work. Just as reflective practice is an essential part of a teacher’s growth and development (Ostorga, 2006), it must also guide administrators as they supervise and support teachers while addressing the many different aspects of their positions. Osterman and Kottkamp (1993) define reflective practice as “a means by which practitioners can develop a greater level of self-awareness about the nature and impact of their performance, an awareness that creates opportunities for professional growth and development” (p. 2). The process of reflection and its influence on meaningful practice by educators is not new. John Dewey (1933) identified three attitudes in the process of reflective thinking: open-mindedness, responsibility, and wholeheartedness. Of the three, open-mindedness was considered to be most significant, requiring a teacher to remain open to multiple, alternative possibilities and continuously
question routines and practices (Ostorga, 2006). Such reflective learning has become a dominant feature in professional development and adult education, particularly in formal teaching development programs to enable new teachers to learn from their experiences and transform their practice (Russell, 2008). While the research does not reflect a similar impact on professional development for administrators, it is fair to hypothesize that the effects of reflective practice would be similar. In the process of evaluating teachers, administrators are well served to continuously examine and question their practice.

**A Brief History of Teacher Evaluation**

Teacher supervision and evaluation have evolved as public education has evolved. Even in the beginning stages of formal education in the United States, dating as early as the 1600s, teacher evaluations took place in some form. Burke and Krey (1975) have identified specific touchstone characteristics of six different eras of evaluation, the first being the Period of Administrative Inspection from 1642-1875. Teachers were judged more on their personal qualities and appearance than on their instructional techniques and, in fact, met with “inspectors” rather than evaluators. Inspection was related to the observance of pupil control, managing the school’s physical plant, and meeting the requirements of the prescribed curriculum. Improvement of instructional procedure was not a priority.

In the middle of the 19th century, evidence suggests a change of focus to instruction. “The period from the beginning of formal education in the United States up to the mid-1800s saw the dawning of the awareness that pedagogical skills are a necessary component of effective teaching. Although there was little or no formal discussion about the specifics of these skills, the acknowledgment of their importance
might be considered the first step in the journey to a comprehensive approach to developing teacher expertise” (Marzano, Frontier, & Livingston, 2011, p. 2). The job of a teacher was beginning to be viewed as a more complex endeavor, where feedback and evaluation were essential in order to improve instruction.

In the mid to late 1800s scientific management began to take root in education based on Frederick Taylor’s principles measuring specific behaviors of factory workers to improve production. If it worked for factories, why not students? Burke and Krey (2005) call this the Period of Efficiency Orientation. Inspectors of schools transitioned into supervisors of teachers who worked to improve instruction by focusing on efficiency and measurement. In this phase, relationships between supervisors and teachers began to foster improved instruction. Notable names in education who applied Taylor’s factory worker principles were Edward Thorndike, who viewed measurement as the tool needed in order to have a more scientific approach towards education, and Ellwood Cubberley, who used the factory metaphor to identify a set of principles for school administrators who would then analyze data to ensure schools and teachers were productive (Marzano et al., 2011). Additionally, Charles Bobbit drew connections between what worked in industry and what should work in schools, trying to produce predictable and improved results (Shinkfield & Stufflebeam, 1995). Students were viewed as raw materials to be molded by efficient and regulated teachers to make sure learning goals were attained. To some degree, there has been a return to this philosophy today with an emphasis of tying standardized test results directly to teachers’ performance.

Overlapping with the era of scientific management and the Period of Efficiency in Education, Medley, Coker, and Soar (1984) identified their own distinct 20th century
periods of education, beginning with the Search for Great Teachers in the early 1900s. In this first phase, specific desirable teacher characteristics were judged in the hopes of relating positively to student achievement, but it was ultimately viewed as a failed approach. Rather than relying on teacher characteristics such as personality, gender, age, intelligence, and teacher’s education, actual teaching behaviors became favored targets for evaluation such as instructional techniques, professional attitude, cooperation, and maintenance of discipline records (McNergney & Imig, 2003; Shinkfield & Stufflebeam, 1995). Effective teaching behaviors were sought for a more direct link to student achievement, leading into the Period of Examining Teacher Performance as identified by Medley et al. Attempts were made to link teaching behaviors to standardized achievement test scores, though there were unclear results.

Following World War II the trend was away from scientific management. Supervision emphasized the development of individual teachers and focused less on students as raw materials and products (Marzano et al., 2011). The post-war baby boom brought enlarged schools and school systems along with a shortage of teachers. An increased need for supervision of instruction went hand in hand with the changing role of supervisors and the need for clarity in evaluation. There was a “growing belief that the entire system must gain from improved teacher performance arising from more widely acceptable evaluation processes” (Shinkfield & Stufflebeam, 1995, p. 23).

By the 1960s and 1970s, expectations of teacher accountability were commonplace. Teacher evaluation attained growing importance and by the 1980s it was viewed as a normal part of the educational process (Shinkfield & Stufflebeam, 1995). The public became more aware of the significance of quality teachers while at the same
time clinical supervision “spread like wildfire” (Marzano et al., 2011). Robert Goldhammer described the process of clinical supervision in his 1969 book, *Clinical Supervision: Special Methods for the Supervision of Teachers*, outlining five phases: pre-observation conference, classroom observation, analysis, supervision conference, and analysis of the analysis. The philosophy behind Goldhammer’s definition of clinical supervision was to continually provide professional education and opportunities for growth to teachers. He envisioned rich conversations between supervisors and teachers that led to improved student learning; but as with many ideas in education, rather than a means to improved instruction, clinical supervision became an end unto itself. The five phases became the widely used model for supervision, absent the inquiry-based, rich dialogue for which Goldhammer hoped (Marzano et al., 2011).

The 1980s and 1990s saw new trends emerge while largely maintaining the structure of clinical supervision. It was in the latter part of the 20th century that *A Nation at Risk* was published by the National Commission on Excellence in Education, raising alarms for reform at the elementary and secondary levels and focusing on the need to improve teacher performance (Shinkfield & Stufflebeam, 1995). The era of accountability had begun, and supervision and evaluation blended into the same model. Whether Goldhammer foresaw clinical supervision as an evaluative model or not, it became one. Madeline Hunter (1994) introduced her seven-step lesson structure, which also merged into an evaluative tool widely used by supervisors. Clinical supervision became the structure, while Hunter’s model became the content.

The Period of Emerging Patterns of Participation beginning in the 1990s saw teachers becoming more active participants in the supervision and evaluation process as
opposed to passive receivers of information (Burke & Krey, 2005). Certainly, the Hunter model required significant discourse between the teacher and supervisor as the supervisor scripted lessons and engaged in the process of analyzing each part of the lesson with the teacher.

The standards movement in the 2000s introduced individual state standards and national Common Core Standards, as well as standards for supervisors published by the Interstate School Leaders Licensure Consortium. The demand for accountability in education had never been so high. Criticism of evaluation practices brought sweeping changes via President Obama’s American Recovery and Reinvestment Act of 2009 and more than four billion dollars of Race to the Top (RTTT) funds offered to states willing to restructure and recreate their evaluation practices. In 2009 only 14 states required annual teacher evaluations, yet by 2012 forty-three states required that all new teachers be evaluated annually (Riordan, Lacireno-Paquet, Shakman, Bocala, & Chang, 2015).

In Rush to Judgment (2008), Toch and Rothman identified the federal No Child Left Behind Act (NCLB) of 2002 as a driving force in focusing yet again on evaluation components that do not necessarily impact teacher performance and student achievement. Their report was critical of the continued use of “drive-by” observations and evaluations that had little value towards directly addressing the quality of instruction. Similarly, a study called The Widget Effect cited flaws in evaluations that tend to assume classroom effectiveness is the same from teacher to teacher and fails to identify specific development needs in teachers by treating them all as if they were the same (Marzano et al., 2011; Partee, 2012). In an era when teacher accountability is taking root as never before, teacher evaluations were an easy target for states wanting to earn RTTT funding.
Focused feedback and teacher self-reflection resulting in clear goals for improvement became the newly sought after standard in 21st century teacher evaluation.

**Teaching Quality**

Although the reasoning behind evaluating teachers has shifted and evolved over time, research during the second half of the 20th century confirmed what early educators suspected: teachers matter when it comes to student achievement. “Two years of effective teachers could not remediate the achievement loss caused by one year with a poor teacher” (Bembry, Jordan, Gomez, Anderson, & Mendro, 1998). This is a powerful finding that should guide all educational reform efforts. While plenty of research points to the significance of socioeconomic status and social background as major influences on student achievement (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966) current research also identifies the significant role the teacher plays in student learning. In fact, the effects of well-prepared teachers on student achievement have the potential to overcome the influences of socioeconomic backgrounds (Darling-Hammond, 2000). It is important to note, however, that the positive association between high socioeconomic status and high achieving students has had an unintended consequence of higher teacher evaluation scores for those teachers. “Teachers who have students with higher incoming achievement levels tend to receive classroom observation scores that are higher on average than those received by teachers whose incoming students are at lower achievement levels” (Warring, 2015, p. 706). Policymakers and educators who rank schools by test scores are essentially ranking schools by socioeconomic status as well, which is not a fair or consistent measure of teacher success.
Methods to examine the process of teacher evaluation in different school environments need to be considered and explored.

According to a report by the National Commission on Teaching & America’s Future called *What Matters Most: Teaching for America’s Future* (1996), there are three basic premises that should be at the core of our educational reform: (1) What teachers know and can do is the most important influence over what students learn, (2) Recruiting, preparing, and retaining good teachers is the central strategy for improving our schools, and (3) School reform cannot succeed unless it focuses on creating the conditions in which teachers can teach, and teach well. In accordance with these findings and recommendations, recent federal policy efforts like the Race to the Top competition and the No Child Left Behind waivers highlighted teacher evaluation systems as a hopeful means toward ensuring effective instruction (Garrett & Steinberg, 2015; McGuinn, 2015).

With this knowledge in hand, evaluation of teachers in the 21st century has become the foundation of reform efforts. The RTTT grant program alone initiated a rush to reform teacher evaluation programs in the 40 states that applied. “For decades, teacher evaluations were little more than a bureaucratic exercise that failed to recognize either excellence or mediocrity in teaching. Increasingly, this is no longer the case. Since 2009, the vast majority of states have made significant changes to how teachers are evaluated for the main purpose of improving instruction” (Hull, 2013, p. 1).

Linda Darling-Hammond, a preeminent researcher in the areas of teaching and evaluation, distinguishes between teacher and teaching quality (2014). When examining the factors that contribute to student achievement, teaching quality is high on the list.
While teacher quality might be defined as personal traits and skills, the focus of supervision and evaluation in the late 1800s and early 1900s, teaching quality refers to strong instruction that enables students to learn. According to Darling-Hammond, traits of teacher quality found to be important are strong content knowledge, content pedagogy, understanding of learning development, abilities to organize and explain ideas, and adaptive expertise. On the other hand, teaching quality comes with instructional skills that enable a wide range of students to learn; quality instruction meets the instructional goals, the demands of the discipline, and the particular needs of the students. Teaching quality is, in fact, a function of teacher quality (2014), hence the movement towards holding teachers more accountable throughout the country.

**Common Characteristics of Teacher Evaluation Models**

Since we know teachers are vital to student achievement, then the primary purpose of teacher evaluation should be promoting and improving instruction while helping teachers grow professionally. The best evaluation systems will be those that help all teachers become good and push good teachers to be great (Almy, 2011). With so many states creating and recreating evaluation systems for teachers based on RTTT funding criteria, there is now a wealth of information and recommendations available regarding what these systems should look like and how they can best influence instruction. We know from decades of poor evaluation systems what does not work, but change is difficult and identifying and implementing effective systems is complex. Darling Hammond states, “What this country really needs is a conception of teacher evaluation as part of a teaching and learning system that supports continuous improvement, both for individual teachers and for the profession as a whole. Such a
system should enhance teacher learning and skill, while at the same time ensuring that teachers who are retained and tenured can effectively support student learning throughout their careers” (2014, p. 5). This kind of vision requires a great deal more from both teachers and their evaluators than evaluation systems used in the 20th century. “Drive-by” observations historically comprised the entirety of a teacher’s year-end evaluation. In order to base teachers’ effectiveness on more evidence, teacher evaluation systems following the RTTT mandate now commonly include value-added student achievement scores and student learning objectives (SLO’s) in addition to classroom-based observations (Garrett & Steinberg, 2015).

In 2012, the Center for American Progress compiled a report titled Using Multiple Evaluation Measures to Improve Teacher Effectiveness (Partee, 2012), reviewing the plans of 23 second-round NCLB waiver applicants. In the states’ plans for new systems of evaluation there were overarching themes. Waivers winners recognized that one or two indicators cannot capture a teacher’s effectiveness; a range of measures and methods are required such as classroom observations, self-assessments, surveys, student-learning measures, and teaching artifacts. Additionally, successful systems need an infrastructure of support. The commitment to ongoing learning and the investment of time for training and implementation of these components are significant. Teachers and principals need appropriate training in using the rubrics and protocols and analyzing the data used for linking student achievement to teacher effectiveness. Under the NCLB waiver process, states agreed to the following criteria when establishing teacher and principal evaluation and support systems. These criteria represent the insights gained since the implementation of NCLB in 2001. The evaluation system must function as follows:
• Be used for continual improvement for instruction
• Differentiate performance using at least three performance levels
• Use multiple valid measures in determining performance levels, including significant factor data on student growth for all students and other measures of professional practice
• Evaluate teachers and principals on a regular basis
• Provide clear, timely, and useful feedback, including feedback that identifies needs and guides professional development
• Be used to inform personnel decisions

With these criteria in mind, multiple components of teacher performance within an evaluation tool have become commonplace across the states. In fact, the Center for Public Education reported that 41 states now require or recommend teachers be evaluated using multiple measures of performance (Hull, 2013). Most states continue to include classroom-based observations (what used to be the sole component of teacher evaluation) but now also include some kind of value-added modeling and student learning objectives in addition (Garrett & Steinberg, 2015). “Value-added measures, or growth measures, are used to estimate or quantify how much of a positive (or negative) effect individual teachers have on student learning during the course of a given school year” (The Glossary of Education Reform). In other words, it is a way of measuring a student against himself and his/her peers from year to year.

Value-added models (VAMs) are arguably the most controversial component in current teacher evaluation systems. The appeal of using student achievement data is adding objective information to a process that has been historically subjective. Teachers
are not penalized for students who are below grade level as long as they make progress similar to other students at the same achievement level (Almy, 2011). Students do not need to grow at the same rate; however, VAMs have been found to be highly unstable (Darling-Hammond, 2012). Teachers’ ratings vary substantially from class to class and year to year and are also significantly affected by the differences in the students assigned to them. End-of-year standardized assessments are limiting in the amount of information that can be directly attributed to an individual teacher, not to mention that the tests themselves change from year to year. Factors such as students’ experiences, socioeconomic status, and parental involvement have a significant influence on student learning gains as well (McNerney & Imig, 2003). Because these limitations are well known, there is resistance by many stakeholders to include them in the evaluation of teachers (Hull, 2013). Darling-Hammond is not suggesting that districts cannot include specific evidence of student learning, but the assessments should be appropriate for the curriculum and the students being taught.

Overall, there is universal agreement that an effective teacher evaluation system should promote improvement of teaching and student learning. How do we get there? Darling-Hammond (2012) summarized her research into seven recommendations for an effective evaluation system:

1. Teacher evaluation should be based on professional learning standards utilized by novice and expert teachers alike.

2. Evaluations should include multi-faceted evidence of teacher practice, student learning, and professional contributions.
3. Evaluators should be knowledgeable about instruction and well trained in the evaluation system.

4. Evaluation should be accompanied by useful feedback and connected to professional development.

5. The evaluation system should value and encourage teacher collaboration;

6. Expert teachers should be part of the assistance and review process for new teachers and teachers who need assistance.

7. Panels of teachers and administrators should oversee the evaluation process to ensure quality and fairness.

These recommendations are widely accepted and utilized by states across the country as they continue to refine their teacher evaluation systems. As an example, in California they were directly referenced in a recent report for a recommended evaluation system (Adams et al., 2015).

Another relevant aspect of these recommendations worth noting is that they help to create evaluations that are both formative and summative. An effective system should serve both purposes. Ongoing, consistent formative assessments foster professional growth and improved practice, while periodic summative evaluations of performance support employment decisions (NEA, 2010).

**Perceptions of Evaluation by Administrators**

As the primary evaluators of teachers, principals and supervisors will be the main determinants of the success of new evaluation tools. While there is plentiful research available regarding teachers’ perceptions and attitudes towards evaluation systems, research into the perceptions and attitudes of the evaluators who administer them is
limited. Since the turn of the 21st century, administrators have experienced a dramatic increase in the amount of time and skill required to implement new and more demanding teacher evaluation systems (Kersten & Israel, 2005), even more so since the implementation of RTTT grants. Multiple-component teacher evaluation systems require principals and supervisors to be trained in each of the components, such as observation instruments, VAMs, student learning objectives, student surveys, and artifacts of teaching performance. They must also provide meaningful feedback and coaching if the goal is to move all teachers forward, regardless of their starting point. Additionally, most evaluation systems require an online feature where data are collected for reporting purposes of individual teachers.

Some studies have emerged within the last few years reflecting principals’ perspectives on new evaluation systems implemented with RTTT grants in their respective states (Derrington, 2014; Dodson, 2015; Tennessee DOE, 2012; Riordan et al., 2015). In Kentucky and Illinois, principals themselves are subject to a high-stakes assessment to determine their proficiency in implementing the teachers’ evaluation instrument. The reasoning behind the assessment is for observers to demonstrate they can be accurate and consistent while applying the rubric and making personnel decisions based on the results (Dodson, 2015). Regardless of the reasoning it has created “angst” among principals. Anecdotal evidence suggests it may have caused early retirements despite the fact that the majority of principals and evaluators passed it on their first try.

There were some positive early findings of new evaluation systems across several states studied. Results showed that 66% of principals surveyed in Kentucky and 83% of those surveyed in New Hampshire agreed that the new evaluation systems improved their
schools’ instructional programs (Dodson, 2015; Riordan et al., 2015). Also in New Hampshire, 67% of administrators believed that the new evaluation system would result in accurate ratings of teachers. On the other hand, respondents in Tennessee and Kentucky negatively cited the time demand. Specifically, in addition to the time needed for thorough training using the instruments, the increased number of observations, walk-throughs, and conferences while maintaining the online collection of data were all time-intensive for school leaders.

**Teacher Evaluation in New Jersey**

On August 6, 2012 Governor Chris Christie signed into law the Teacher Effectiveness and Accountability for the Children of New Jersey Act (TEACHNJ) in order to “raise student achievement by improving instruction through the adoption of evaluations that provide specific feedback to educators, inform the provision of aligned professional development, and inform personnel decisions” (TEACHNJ Guide, 2014, p.1). At its core the Act was created to reform the processes for earning and maintaining tenure. In preparation for the start date of TEACHNJ in the fall of 2012, the New Jersey Department of Education (NJDOE) notified all school districts in the state on September 1, 2011, that there would be a pilot program of new teacher evaluation systems during the 2011-2012 and 2012-2013 school years in preparation for full implementation by all school districts in the state for the 2013-14 school year. Commissioner Christopher Cerf stated in his memo that they were “taking an important step toward developing a fair, consistent, and learning-centered evaluation system by providing 10 districts across the state with $1.1 million to collaboratively design and implement state-of-the-art educator evaluation systems.” He highlighted the fact that the evaluation systems would include
multiple measures to identify both learning outcomes and effective practice. Another 20 school districts joined the pilot for the 2012-2013 school year (McGuinn, 2012). The NJDOE created a State Evaluation Advisory Committee and District Evaluation Advisory Committees to solicit feedback from the pilot districts.

The pilot years provided the NJDOE with four key lessons: (1) Stakeholder engagement and communication are essential, (2) High quantity and high quality training are also essential. How an instrument is implemented is more important than which instrument is selected, (3) Non-testing grades and subjects present a major challenge if a value-added measure is to be included for every teacher, and (4) Capacity issues for principals present a challenge as well since they spend significantly more time conducting evaluations, providing feedback, and offering guidance for teachers (McGuinn, 2012).

After the pilot years, the TEACHNJ Act identified the changes that would be in place regarding evaluation rubrics: there would be four-tiered measurement (Highly Effective, Effective, Partially Effective, and Ineffective) as opposed to the typical binary measurement system of Satisfactory or Unsatisfactory used previously; multiple measures of student achievement and educator practice would be used; multiple observations of all teaching staff members would be required rather than just non-tenured staff; ongoing training, calibration, and monitoring of the system implementation would be required.

The New Jersey Department of Education pre-approved several evaluation systems to be fully implemented during the 2013-2014 school year, considered to be major frameworks recognized both nationally and regionally, and gave school districts
the opportunity to choose one of those or develop one of their own using specific criteria (Mooney, 2013). By early 2013, 85% of New Jersey’s school districts had selected one of five different systems:

- Charlotte Danielson Framework for Teachers – 291 districts or 60%
- Stronge Teacher and Leader Effectiveness Performance System – 53 districts or 11%
- Mid-Continent Research for Education and Learning (McREL) – 45 districts or 9%
- Marzano’s Causal Teacher Evaluation Model – 44 districts or 9%
- The Marshall Rubrics – 32 districts or 6.5%

Along with the adoption of a specific teacher evaluation system, the state of New Jersey allowed individual districts to train their administrators and teachers (McGuinn, 2012).

The New Jersey Department of Education enlisted the support of the Rutgers University Graduate School of Education to conduct an independent assessment of 10 districts throughout the state following the 2011-2012 pilot year of the teacher evaluation program (Firestone et al., 2013). Both administrators and teachers were surveyed with mixed results. In general, the study found that administrators viewed the new teacher evaluation process more positively than teachers. Administrators saw more positive effects on professional development and growth than teachers did. Additionally, 74% of the administrators surveyed found the evaluation rubrics to be accurate assessments compared to 32% of the teachers. In fact, “many teachers described the evaluation rubrics as subjective” (p. 6). Similar to studies in other states, it is significant to note that 90% of administrators reported spending more time conducting observations and entering
observation data than they had previously. In fact, the study noted that the greatest time
demands of the teacher evaluation program were placed on administrators.

The past five years have presented significant change and challenge in the area of
teacher evaluation in New Jersey. With full implementation in place for the past three
years, an opportunity for reflection on the current systems being used is relevant and
useful moving forward.

**The Stronge Model**

Approximately 11% of New Jersey’s school districts are using James Stronge’s
Teacher Effectiveness Performance Evaluation System (TEPES) to evaluate their
teachers under the TEACHNJ mandate enacted in 2012 (Mooney, 2013). The Stronge
model comprises seven performance standards that define the major duties performed by
teachers. The highly used Danielson Framework, on the other hand, consists of four
domains broken down into 22 components, which then contain from one to five elements
equaling over 70 indicators. Stronge’s comparatively minimal seven standards are as
follows: Professional Knowledge, Instructional Planning, Instructional Delivery,
Assessment of/for Learning, Learning Environment, Professionalism and
Communication, and Student Progress (Stronge, 2012). Teachers and their administrators
collect and present data within each of the seven standards to document performance
based on well-defined job expectations.

The Stronge system as presented to a school district includes a variety of
components from which the district can choose to evaluate its teachers. The observation
and summative performance reports utilizing the seven standards are mandatory. Also
included are forms for teacher self-assessment, professional development plans, student
growth objectives, student surveys, and parent communication logs. School districts can choose from these forms and add their own in order to best customize the evaluation system to suit their needs and goals.

In his book, *Qualities of Effective Principals* (2008), Stronge makes his case for teacher evaluation as a catalyst for school improvement. He notes that professional growth and performance accountability were typically viewed as mutually exclusive; but, in fact, for teacher evaluation to be most beneficial, both of these factors should be combined. He identifies four key research findings that led to the development of his evaluation system:

- **Teacher supervision and evaluation are fundamental responsibilities of the principal,** yet principals and teachers find their supervisory interactions to be difficult and unsatisfying.
- **The number of incompetent teachers is well above the number of teachers who were actually documented as incompetent.**
- **Over the past two decades, teacher evaluation systems have been recognized as integral to teacher improvement and to overall school improvement.**
- **Teacher evaluation systems of the past have failed to fulfill both purposes of accountability and improvement because of poor implementation and a negative atmosphere in which the evaluation takes place.**

When it came to promoting his evaluation system as an effective tool for New Jersey school districts, Stronge highlighted that it had been endorsed by the New Jersey Supervisors and Principals Association and that, in addition to the teacher evaluation model, it also includes a model for evaluating principals, part of the state’s mandate.
under TEACHNJ. Additionally, it includes the student performance piece required by the state. Stronge acknowledges that including a numerical summative rating for each teacher “is a bit artificial,” but there is “a margin of error, and that is a best estimate” (Mooney, 2012, p. 2).

Influence of Teacher Evaluation on Performance

With so much attention placed on teacher evaluation and creating the best vehicles to measure teacher and student performance in the last decade, one hopes research reveals that evaluation tools actually influence teacher behaviors and skills. The research is in fact mixed, dependent on the many factors that influence the creation and implementation of an evaluation tool. Teachers are generally motivated to teach since many enter the profession for idealistic reasons or because they enjoy working with children. They tend to have long careers “during which their skills develop and are revealed. Therefore, career concerns can be more important than incentives based on short-run performance, even if the latter can be measured with sufficient accuracy” (Dixit, 2002, p. 719). Given these parameters, it is essential that teachers be recognized as significant stakeholders in the evaluation process and included in the design and rollout of new evaluation systems, particularly those that include value-added models (McGuinn, 2015).

Positive results were identified in a study of mid-career math teachers in Cincinnati (Taylor, Tyler, Bettinger, Chay, Figlio, Hoxby, & Staiger, 2012), strongly suggesting that as a result of Cincinnati’s Teacher Evaluation System (TES), teachers developed skill and/or changed their behavior in a lasting manner. Feedback during evaluation provided teachers with new information about their performance that led to the
development of new skills; however, the researchers were not able to identify exactly the behaviors or practice that led to student achievement gains, just that gains were made as a result of evaluation feedback.

Toch and Rothman in *Rush to Judgment: Teacher Evaluation in Public Education* (2008) reported on evaluation as a tool for improving teacher performance by examining a number of national, state, and local evaluation systems. Even before the evaluation reform movement started, they identified the need for comprehensive evaluations. Holistically, they stated that when evaluations contain multiple components such as classroom observations by multiple evaluators, student work, and teacher reflections, then evaluations “are valuable regardless of the degree to which they predict student achievement, and regardless of whether they’re used to weed out a few bad teachers or a lot of them. They contribute much more to the improvement of teaching than today’s drive-by evaluations or test scores alone. And they contribute to a much more professional atmosphere in schools” (p. 13). In addition, Toch and Rothman made several recommendations that would help schools judge teachers’ strengths and weaknesses fairly and effectively while helping to improve their teaching. Evaluation systems should take into consideration the following recommendations:

- A hybrid model that combines observations and test scores
- Significant training for evaluators and consideration for creating evaluation teams
- Evaluation of the evaluation systems themselves needs to be ongoing
- Incentives for schools and school leaders such as performance-based rewards, staffing authority, and professional development linked to outcomes
• Performance pay

Most of these recommendations have become commonplace, although performance pay and utilizing student test scores remain controversial.

In the state of New Jersey, teacher evaluation models have been revamped and in place since 2013. Teachers are evaluated annually based on multiple observations and student growth objectives. Despite the goals behind these changes, a study by Callahan and Sadeghi (2015) found that while observations are conducted more frequently, their value has not improved. The observations are more rigid and tend to follow a script. In addition, professional development opportunities have changed little even though the intent is to tie professional development to the outcome of evaluations. They speculate that perhaps evaluations are designed to measure teachers as opposed to assessing professional development needs. Callahan and Sadeghi conclude their study with these words: “We need to move beyond checklists and rubrics that fail to acknowledge teaching excellence and we need to identify and offer professional development strategies that are most effective to improving teaching pedagogy and ultimately improving student achievement” (p. 57).

The literature encompasses the history and current state of teacher evaluation with a great deal of research identifying the theory behind evaluation components and implementation. Turning theory into effective practice continues to be the challenge, as it is in most aspects of the United States educational system.
CHAPTER 3

METHODOLOGY

Introduction

The purpose of this study is to explore and analyze school administrators’ views of the Stronge teacher evaluation tool and its influence on teaching practices. A phenomenological qualitative research design was chosen, as it is most suitable to the posed problem. The research from this phenomenological study will describe the lived experiences of school administrators as they use a common teacher evaluation tool. As defined by Creswell (2007), qualitative research is most appropriate when there is a problem or issue to be explored, when we want to empower individuals to share their stories, and when quantitative measures and statistical analyses simply do not fit the problem. The research in this study can help fill a gap in the literature and improve existing practice as opposed to measuring outcomes more appropriate to a quantitative study (Maxwell, 1993).

In this chapter, I explain how I became interested in seeking out administrators’ views on a teacher evaluation tool. I provide details on the design and methods used to address my research questions. I address how participants were selected and how I collected and analyzed data with validated research procedures.

Background

I began teaching in 1985 in a northern New Jersey suburban district and have remained with that district throughout my career. I was a fifth grade teacher for 18 years before serving as an elementary school principal for the past 11 years. Teacher evaluation
practices remained relatively unchanged when I was a teacher; but during my tenure as a principal, we have encountered significant changes.

For all but the past five years, tenured teachers were evaluated using a district-created evaluation tool. There were four areas of evaluation: Instructional Performance, The Employee as a Professional Staff Member, Progress on Approved Professional Development Plan for Current Year, and Evaluator’s Assessment of Overall Performance and Recommendation. Within each area there were subcategories in which the teacher was rated either Satisfactory or Unsatisfactory. Non-tenured teachers were evaluated using a similar tool; although instead of the binary system, they were rated in each subcategory using a five-point scale (Superior, Good, Average, Below Average, Poor.) The goal for non-tenured teachers was to be perceived as “Clearly Superior” by the end of the third year in order to earn tenure. This system was the one used to evaluate me as a teacher and the same one I used to evaluate teachers once I became a principal.

Beginning in 2012-2013, however, our district adopted a new evaluation tool due to the new state mandates under TEACHNJ. The state-approved evaluation system chosen by our school district (and 11% of all New Jersey school districts) was James Stronge’s Teacher Effectiveness Performance Evaluation. The school year 2012-2013 was a pilot year for us so that we could get used to the tool and collaborate with the local education association to work out negotiable details. All administrators and teachers participated in mandatory training in the various components of the evaluation system and its online management tool. By 2013-2014, the evaluation system was in place and used to evaluate all tenured and non-tenured teachers using the same four-point scale (Highly Effective, Effective, Partially Effective, and Ineffective.)
An effective teacher evaluation system should promote improvement of teaching and student learning. As one of the administrators charged with implementing the new system, I do my best to use it as a tool for talking to teachers about their instructional practices while helping them reflect and grow. I knew my teachers were somewhat concerned about the changes and the perceived heavy-handedness of the state and its newfound involvement in their evaluations. I talked them through the changes every time I observed a teacher in that first year and again during their year-end summative evaluations. While there were some parallels between the district-created evaluation tool and the Stronge tool, I found the new standards within the Stronge tool somewhat refreshing and a good opportunity to recharge our instructional focus. I could not help but wonder if the teachers felt similarly and how my colleagues were feeling about the change. This turned out to be the perfect topic choice for my dissertation research. I knew others were researching teachers’ perceptions of New Jersey’s new evaluation tools, but what were administrators’ thoughts? I was interested to hear the perceptions from this largely untapped resource. The following research questions shape the study:

1. To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System improves teachers’ instructional practices?

2. To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System proficiency scale (Highly Effective, Effective, Partially Effective, Ineffective) is an accurate reflection of teacher quality?
3. To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System enhances their roles as instructional leaders?

**Design**

A qualitative design was the appropriate method for this study. As defined by Creswell (2007), a qualitative study includes the voices of its participants in a “natural setting” and the researcher gathers information by “talking directly to people and seeing them behave and act within their context” (p. 37). Furthermore, it had an emergent design, which allowed the researcher to learn about the issue from participants and address the research to obtain information about that issue.

School administrators were interviewed in order to gather their perceptions on a relatively new state-mandated teacher evaluation system. Since it is the administrators who are solely responsible for carrying out the mandate, it was essential to hear directly from them their thoughts on the effectiveness of the teacher evaluation system. Statistical procedures or other means of quantification were not appropriate in this case since a more naturalistic approach would aid the researcher in understanding the context-specific phenomena (Golafshani, 2003).

**Sampling**

There is a total of 27 administrators in the northern New Jersey suburb chosen for this study. Those who have used the Stronge evaluation system to evaluate teachers for at least two years were invited to participate in the research. Permission from the superintendent was sought and received. Of the 27 administrators, 24 had the required two-years’ experience and were invited to participate through a letter of solicitation sent
via email. Those who agreed to participate were given a Demographic Profile Questionnaire. The researcher interviewed all qualified volunteers, a total of 14 from the pool of 24 administrators. “A small sample that has been systematically selected for typicality and relative homogeneity provides far more confidence that the conclusions adequately represent the average members of the population than does a sample of the same size that incorporates random or accidental variation” (Maxwell, 1993, p. 235).

Following are the criteria used for selection in this study:

- Certificated New Jersey administrator
- Minimum of two years’ experience using the Stronge evaluation tool at the selected school district
- Representation from various levels: elementary, secondary, district

The participants’ experience using the Stronge evaluation model ranged from two years (the minimum needed to participate in the study) to a maximum of five years, the number of years that the Stronge model had been used in the district. It was important to invite administrators from various levels and disciplines in order to get as broad a view as possible within the relatively narrow sampling. Administrators not selected for this study were thanked for their time and willingness. Those who participated included the following: five elementary principals, one middle school principal, and two middle school assistant principals, one high school principal, two secondary subject-area supervisors, and three K-12 administrators.

Profiles of the Site and Participants

For the purpose of this study and to protect the confidentiality of its participants, names were not included in the findings. A pseudonym, Placidville, was used in place of
the actual name of the northern New Jersey suburban K-12 school district. Placidville is an upper middle class community consisting of a number of schools across the K-12 continuum. Information from the 2014-2015 New Jersey School Performance Report indicates a 97% overall graduation rate.

Regarding teacher evaluation, the Placidville’s current Board of Education policy states that the evaluation of teaching staff members is consistent with the Teacher Effectiveness and Accountability for the Children of New Jersey Act (TEACHNJ) and the AchieveNJ administrative codes. The minimum requirements for the evaluation procedures for teachers as outlined in New Jersey Administrative Code 6A:10-2.4 are followed. Furthermore, the Board annually adopts evaluation rubrics for all teaching staff members with four defined annual ratings: Highly Effective, Effective, Partially Effective, and Ineffective.

All non-tenured teachers have three observations per year in Placidville while all tenured teachers have two observations. All teachers are assigned a primary evaluator who conducts at least one of the observations as well as the summative evaluation at the end of the year. Administrators utilize the Stronge teacher evaluation system, using the My Learning Plan online tool provided by Frontline Education, Frontline Technologies Group. All observations are conducted by the superintendent, assistant superintendent, business administrator, director of special programs, building administrators, and instructional supervisors. Administrators are required to conduct two observations per year with another administrator to satisfy inter-rater reliability standards. In addition, informal discussion around a shared observation report takes place during monthly administrative meetings in order to create consistency across the district.
Data Collection

This qualitative study derived its data from interviews with administrators regarding their perceptions of the Stronge teacher evaluation instrument. Permission from the superintendent of Placidville to interview the administrators was sought and received. The researcher reviewed each of the evaluation components provided by Stronge that administrators use in the annual evaluation process of teachers. Using these components, information gleaned from the literature review, and the researcher’s own experience with the instrument, three overarching research questions were developed. “Research questions should have a clear relationship to the goals of your study and should be informed by what is already known about the phenomena you are studying and the theoretical concepts and models that can be applied to these phenomena” (Maxwell, 1993, p. 217). Interview questions were derived from each of the broad research questions in order to gain a thorough understanding of administrators’ perceptions. Prior to the interviews taking place, the interview questions were shared with a jury of experts who were consulted to assist with clarity and focus, ensuring the questions addressed the targeted research problem. Table 1 lists the research questions and the subsequent interview questions.

Research/Interview Questions

Research Question 1

To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System Improves teachers’ instructional practices?
Sub-questions

1. How many times per year do you meet with a teacher using the TEPES components and for what purpose?

2. What specific parts of the TEPES, if any, have the most impact on teachers’ instructional practices and how?

3. Of the seven TEPES standards, which ones, if any, effectively improve teachers’ instructional practices?

Research Question 2

To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System proficiency scale (Highly Effective, Effective, Partially Effective, Ineffective) is an accurate reflection of teacher quality?

Sub-questions

1. How does your pre-existing knowledge of teachers affect your ratings of them using the TEPES?

2. To what extent is the TEPES an objective instrument that supports the ratings produced using the proficiency scale?

Research Question 3

To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System enhances their roles as instructional leaders?

Sub-questions

1. In what ways, if any, has the TEPES impacted your role as an instructional leader?
2. What are the strengths and weaknesses of the TEPES regarding your role as an instructional leader?

Additional Question

Is there anything about the TEPES that has not been discussed with you that you would like to add to this interview?

The interviews provided the data needed to examine the administrators’ perceptions of Stronge’s teacher evaluation instrument. Each of the study participants was interviewed for approximately 20-30 minutes, face to face. The interviews took place in private settings throughout the school district, typically in the administrators’ offices. The formal interview questions provided structure, although the interviews were semi-structured, allowing the researcher to utilize follow-up questions ensuring thorough understanding and complete responses. It is appropriate for the questions to change during the process of research to reflect an increased understanding of the subject (Creswell, 2007). Initial questions help frame the data collection, but other specific questions arise as result of an interactive design process (Maxwell, 2008). For reference during the interview process, administrators were provided blank copies of the TEPES documents they regularly use to evaluate teachers such as the observation form, the summative form, and the seven standards used to evaluate teachers. The interviews were recorded using a recording device for later transcription. The individual transcripts were then sent to each participant to ensure accuracy, completeness, and final approval. During the interviews, the researcher took field notes indicating information that may or may not have been recorded for transcription. All documents obtained from the research including transcripts, digital recordings, questionnaires, and other printed materials are kept in a
locked filing cabinet to protect the anonymity of the participants. All data will be kept for a period of three years.

**Data Analysis**

The research questions and subsequent interview questions provided a framework for data collection that allowed the researcher to identify themes across the participants’ responses. The data analysis was conducted simultaneously with the data collection in order to progressively focus the interviews and test emerging conclusions. Interview transcripts were reviewed and coded, identifying categories that arose addressing administrators’ perceptions. “Such categorizing makes it much easier for you to develop a general understanding of what is going on, to generate themes and theoretical concepts, and to organize and retrieve your data to test and support these general ideas (Maxwell, 1993, p. 237).

Matthew Miles and Michael Huberman (1994) condense data analysis into what they consider to be a classic set of moves that can be used across most qualitative research:

- Affixing codes to a set of field notes drawn from observations or interviews
- Noting reflections or other remarks in the margins
- Sorting and sifting through these materials to identify similar phrases, relationships between variables, patterns, themes, distinct differences between subgroups, and common sequences
- Isolating these patterns and processes, commonalities and differences, and taking them out to the field in the next wave of data collection
• Gradually elaborating a small set of generalizations that cover the consistencies discerned in the database
• Confronting those generalizations with a formalized body of knowledge in the form of constructs or theories (p. 9)

Upon completion of all interviews, the researcher created a chart for each interview question to record data gleaned from the transcripts related to that question as a means to isolate patterns, identify emerging themes, and elaborate on a small set of generalizations.

**Protection of Human Subjects and Ethical Considerations**

Ethical concerns should be involved in every aspect of the design of a qualitative study (Maxwell, 1993). While conducting this study, care was taken to ensure anonymity of the participants and the study site, maintain confidentiality throughout, and secure informed consent; these are noted ethical concerns to be addressed in qualitative research. Informed consent includes notifying the participants of the nature of the study, their potential role, the identity of the researcher and the associated institution, the objective of the research, and how the results will be published and used (Ali Cheraghi et al., 2014). Researchers have the responsibility to protect all participants from potentially harmful consequences; therefore, anonymity, confidentiality, and informed consent are of utmost priority.

**Validity and Reliability**

When quantitative researchers refer to validity and reliability, they usually point to research that is credible, while qualitative research depends on the ability and effort of the researcher her/himself (Golafshani, 2003). The researcher carries the responsibility of
accurately recording and interpreting the data. Reliability is a construct of generating understanding in a qualitative study more so than generating facts or results. There are ways to ensure greater validity to qualitative research, which is inherently more variable than quantitative research. Respondent validation includes systemically soliciting feedback about data and conclusions from the participants in the study in order to rule out possible misinterpretation of what they said and minimizing bias and misunderstanding (Bryman, 1988). In this study, participants were invited to review the interview transcripts for accuracy and completeness prior to data analysis.

**Role of the Researcher and Researcher Bias**

In a qualitative study the researcher is the key instrument, collecting and analyzing data for the purpose of addressing the posed research problem. “We represent our data, partly based on participants’ perspectives and partly based on our own interpretation, never clearly escaping our own personal stamp on a study” (Creswell, 2007, p. 43). To this study, I bring years of experience as both a teacher and an administrator; I have been an evaluator using the very evaluation instrument at the center of this study. As a result, it is difficult to escape my own personal biases as the primary researcher.

Bias is defined as ways in which the data collection or analysis are distorted by the researcher’s theory, values, or preconceptions (Maxwell, 1993). In a qualitative study as opposed to a quantitative study, one does not control for the effect of the researcher but rather tries to understand it and use it productively. It is important to understand how the researcher influences the interviewee and how to productively and ethically use this influence to address the research questions.
Summary

Chapter 3 provided an overview of the methodology used in this qualitative study, including descriptions of the researcher’s background and the site studied. The processes of data collection and analysis were summarized. The validity and reliability of the research as well as the researcher’s biases were also examined in order to accurately frame the research and the methodology used to address the research questions.
CHAPTER 4

FINDINGS

Introduction

The purpose of this study was to examine the perceptions of school administrators using James Stronge’s Teacher Effectiveness Performance Evaluation System to evaluate teachers. As one of a select number of evaluation systems used by districts throughout the state of New Jersey, understanding its use from an administrator’s point of view will contribute to the research on teacher evaluation and particularly help fill a void concerning administrators’ perceptions.

This qualitative study was conducted during the spring and summer of 2017 to ascertain the perceptions of administrators using the Stronge teacher evaluation system specifically in the Placidville school district, a northern New Jersey suburban district. Fourteen administrators were interviewed using the same interview questions in a semi-structured interview format. The interview questions were derived from three research questions that guided the study: (1) To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System improves teachers’ instructional practices? (2) To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System proficiency scale (Highly Effective, Effective, Partially Effective, Ineffective) is an accurate reflection of teacher quality? (3) To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System enhances their roles as instructional leaders?
The first year Placidville implemented the Stronge evaluation model was 2012-2013, a pilot year during which the administration and teachers were trained to use each of its components. Beginning in 2013-2014, the results were submitted to the New Jersey Department of Education following the mandates of the Teacher Effectiveness and Accountability for the Children of New Jersey Act (TEACHNJ) of 2012. The Placidville administrative team is given some autonomy regarding implementation of the Stronge model. While all forms must be filled out and submitted, how that occurs from administrator to administrator varies.

**Research Question 1 Themes**

**Research Question 1**

To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System (TEPES) improves teachers’ instructional practices?

**Findings**

The Stronge evaluation model leads to an average of 3-5 meetings per year between Placidville administrators and teachers regarding teachers’ instructional practices. The administrators have been given some autonomy by the district regarding those meetings. Overall, they find that the observation report is the component that has the most potential to impact teachers’ instruction. Of the seven performance standards that constitute the observation report, Instructional Delivery was named most frequently by administrators as the standard they focused on to support instruction. Professional Knowledge, Professionalism, and Student Progress were seen to be least supportive when discussing instruction with teachers. In general, Placidville administrators find the seven
standards are a benefit in this model since they provide opportunities for constructive discussion about instruction, although some found the standards to be too confining.

**Interview Results for Research Question 1**

Four interview questions were asked of each administrator to address the topic of the Stronge model’s impact on teachers’ instructional practices. As the interview process developed and progressed, the researcher added two sub-questions regarding the administrators’ views of pre-observation and summative evaluation meetings. It became relevant to the discussion of supporting teachers’ practice; therefore, the questions were added. It is appropriate for the questions to change during the process of research to reflect an increased understanding of the subject (Creswell, 2007).

**Sub-question 1-1**

How many times per year do you meet with a teacher using the TEPES components and for what purposes?

The administrators in Placidville meet with each teacher an average of 3-5 times per year specifically related to components of the Stronge model. Those can include pre- and post-observation meetings, summative evaluation meetings, and others such as discussion about Student Growth Objective goals and Professional Development Plan goals. Observations for individual teachers are shared among administrators so that the number of times an administrator observes an individual teacher varies. More meetings took place with non-tenured teachers. Face to face pre-observation meetings varied greatly, with five administrators stating they always meet with teachers to discuss the pre-observation form, four who meet only with non-tenured teachers beforehand, and the remaining five who leave it up to the teachers to meet or not. Administrator 3 said,
“When you’ve been [observing teachers] as many years as we have, sometimes it just gets redundant and especially with the unannounced [observations], you can’t.” On the other hand, Administrator 6 stated, “Personally I think a pre-conference is essential. I really like the clinical approach to observation that requires a lot of self-reflection from the teachers themselves, so I think a pre-conference is essential no matter who it is and I try.” Administrator 8 does not always meet with teachers before an observation but said, “I always ask people what they want me to look for, what they want me to focus on when I come, through email or informal conversation.” Administrator 13 always holds pre-observation meetings, saying, “I’ve often found when a teacher comes to you in a pre-conference and you make a suggestion and they actually implement it, that tends to stay with the teacher.” Nine administrators always meet face to face with teachers to review the summative document, while the remaining five make it the teacher’s choice.

Sub-question 1-2

What specific parts of the TEPES, if any, have the most impact on teachers’ instructional practices and how?

Administrator 4 cited the self-assessment form since “it forces [teachers] to go back and think about their year; it forces them to set goals for the next year.” The remaining 13 of 14 administrators in the study named the observation report as the component that has the most potential impact on teachers’ instruction. A common reason was the conversation that occurs about instruction when discussing an observation. Administrator 2 said, “We’re looking at all the different aspects of instruction and I think it lends itself to great conversation about planning and delivery and assessment and all those different components.” Similarly, Administrator 5 said, “I think that’s probably
where I have the most good conversations about teaching and learning,’” and Administrator 12 felt “the classroom observation feels like the most supportive, like I’m giving you real feedback on a real thing which hopefully feels like something you can grab onto.” Administrator 1 liked that the form offers specific feedback that teachers value. “Do I think it’s the form that’s giving that? Probably not. It’s the exercise of having an observation, but the form does give clarity to my feedback in a clear way to them to make it more useful.”

The summative form was mentioned as an impactful document by five of the administrators, both positively impactful and negatively. As a tool for supporting instruction, Administrator 9 said, “A summative is just a little too vague and gets into a lot of other stuff about their duties, their coaching responsibility, all the other things they do.” Administrator 5 said, “I think unfortunately they get hyped up about the summative and the score report because they want to see the number.” However, Administrators 2 and 10 felt that it coupled well with the observation report as support to teachers’ instruction.

**Sub-question 1-3**

Of the seven TEPES standards, which ones, if any, effectively foster teachers’ instructional practices?

The Stronge model centers on seven performance standards: Professional Knowledge, Instructional Planning, Instructional Delivery, Assessment of/for Learning, Learning Environment, Professionalism, and Student Progress. These seven standards are components of every observation report and summative report.
Instructional Delivery was named by 13 of 14 administrators as the performance standard that most fostered teachers’ instructional practices. Nine of those administrators named it first. It is in that area of the observation report in which they spend most of their time writing during an observation. Since switching to the Stronge model, some administrators continue to “script” the lessons, doing their best to record as much of the dialogue and action as possible in the Instructional Delivery section. Administrator 3 described its importance this way: “That’s where I spend most of my time since that’s where I describe the lesson, and in my description of the lesson, which is always narrative, it’s when I just throw out the most feedback in regard to how they’re doing both positively and negatively.”

Three other performance standards were mentioned as having a positive impact on teachers’ instruction: eight administrators mentioned Assessment of/for Learning, seven mentioned Learning Environment, and five found Instructional Planning to be impactful. Least mentioned and, in fact, found to be difficult to use as supports for instruction by most of the administrators were Professional Knowledge, Professionalism, and Student Progress. Student Progress relates specifically to teachers’ Student Growth Objectives, which aren’t always the focus of a specific observed lesson. Similarly, Professional Knowledge and Professionalism were seen as pertaining more to the summative document than as tools for fostering teachers’ instruction. Administrator 2 captured other administrators’ views by saying, “I think Professionalism, Student Progress, and Professional Knowledge are the ones that I don’t seem to put as much effort into, or they seem just to have some standard responses from when I’m looking deeper into the lesson of the teacher.”
Sub-question 1-4
What are the strengths and weaknesses of the TEPES regarding teachers’ instructional practices?

Eight administrators stated that having seven explicit performance standards was a strength of the Stronge model. It was a shared perspective of these administrators that the common language of the standards allowed for good opportunities to talk with teachers about instruction. On the other hand, seven administrators felt that some of the standards and the observation form itself got in the way of meaningful conversations with teachers. Administrators wished for a “blank piece of paper” or “one big box” that would allow for more “free form narrative of a lesson.” Two administrators mentioned the tension of using one instrument to both evaluate a teacher and help him/her grow. Administrator 12 said, “It feels very traditional, it looks very traditional, it feels very top down. It doesn’t feel like a human growth model.”

Research Question 2 Themes

Research Question 2
To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System proficiency scale (Highly Effective, Effective, Partially Effective, Ineffective) is an accurate reflection of teacher quality?

Findings
As a teacher evaluation ratings instrument, all of the administrators interviewed felt the Stronge model was not completely objective, using language from “not at all” to “somewhat.” They all stated that pre-existing knowledge of teachers impacted their ratings of them, despite the fact that during training they were told to rate teachers based
on the “preponderance of evidence” as a method for retaining objectivity. Although not completely objective, some administrators felt the seven performance standards helped bring definition to the process and supported some degree of objectivity; however, another common theme was frustration with the lack of clear definition between Effective and Highly Effective within the rubrics of each standard that can lead to contentious conversations with teachers.

**Interview Results for Research Question 2**

Three interview questions were asked of the administrators related to their perspectives of the Stronge model as a ratings tool. The TEACHNJ Act of 2012 mandated the use of a four-point scale to evaluate all teachers regardless of the evaluation instrument. In the summer of 2012, all Placidville administrators, as well as some teacher representatives, participated in mandatory training by James Stronge and his team before using his evaluation system for the 2012-2013 pilot year. Stronge’s training materials include the following statement: “In making judgments for the summative assessment on each of the seven performance standards and based on evidence from the multiple data sources, the evaluator should determine where the ‘preponderance of evidence’ exists, based on evidence from the multiple data sources” (Stronge, 2012, p. 19). “In addition to receiving a diagnostic rating for each of the seven performance ratings, the teacher will receive a single summative evaluation rating at the conclusion of the evaluation cycle. This summative rating will reflect an overall evaluation rating for the employee” (p. 20). Each new school year is to be viewed as separate from the previous year, with the teacher and the administrator accumulating new evidence in each performance standard area.
Sub-question 2-1

How does your pre-existing knowledge of teachers affect your ratings of them using the TEPES?

Of the 14 administrators interviewed, all believed that their previous knowledge of a teacher impacted the eventual ratings. In two cases, using that knowledge was purposeful. Administrators 13 and 14 (who observe in multiple schools) purposefully read through previous observations as a method of identifying an area of focus that may be needed. Most administrators, however, mentioned the subjective nature of evaluating teachers despite the objective aims of the instrument. For example, Administrator 2 reported that pre-existing knowledge plays “a large part in knowing the teacher as a whole and knowing them over time. I think knowing the teacher well and doing all those informal walk-throughs and collecting lesson plans, and talking to them, I have a sense of where they are before rating them effective or highly effective.” Administrator 3 said, “We bring our own perspectives into every thought and decision so inherently we know each teacher in our building who is, without looking at the scale, who we would believe to be, highly effective.” This contradicts James Stronge’s requirement to depend on the “preponderance of evidence.” Administrators expressed that they tended to know more about their teachers than the components of the Stronge system allowed. Administrator 7 said, “If I’m familiar with a teacher’s practice over the last seven years, I haven’t found that I’ve gotten to the summative report and reviewed the comments and come to a new realization about how effective or not they are.” Speaking directly to Stronge’s training, Administrator 8 said, “A lot of it you see just from knowing the teacher in faculty meetings and departmental meetings and what’s going on in the building, and what you
hear over here and see over there. It’s not stuff that you can quantify in the way that I know that he likes to say- don’t trust your eyes, trust the evidence. That’s what [Stronge] said to me. But if I did that, I think the teachers would be really at a disadvantage because I would rate them way more poorly than I know them to be, just because I had never seen that evidence in other arenas, not within this form.”

**Sub-question 2-2**

To what extent is the TEPES an objective instrument that supports the ratings produced using the proficiency scale?

This question served as an appropriate follow-up to the previous question, often segueing without even needing to be asked. Since all administrators stated that their pre-existing knowledge of teachers impacted their ratings of them to varying degrees, it is not surprising that all 14 administrators also stated that the Stronge model was not completely objective. Although two administrators felt that it is not objective at all, the other 12 believe that it is objective to some degree. Administrator 1 felt that the presence of a rubric for each performance standard helped to provide more definition and clarity, thereby making it mostly objective, although there is “still a degree of what you think versus what I think.” Administrator 8 thinks it is “a pretty objective instrument, but rubrics are extremely difficult to make objective and not subjective.” Administrator 11 tries to establish a fresh start each year and says, “The slate is clean as far as I’m concerned. There’s nothing they’ve done, positive or otherwise, that should really affect what I’m seeing in the here and now. I think it’s hard because human nature is such where you want to rely on your past experiences with people.” Administrator 12 also tries to “compartmentalize things” and take a “real analytical approach,” although he said, “I
think plugging anybody into a box is hard and I think it’s also sort of weird to take somebody who’s been teaching 25 years and look at them a year at a time; you got Highly Effective in Professional Knowledge because you went to two workshops last year, but you didn’t go to any workshops this year, so now your Professional Knowledge is effective.” Two administrators talked about the challenge of using an “objective” instrument that can potentially be harmful to the evaluator/teacher relationship.

Administrator 9 said, “It’s really a delicate balance because these people are people we rely on every day. We rely on them being happy because happy teachers are better teachers.” Administrator 8 felt similarly that “human relations” are at the crux of the evaluator/teacher relationship and in most cases the teachers “aren’t going anywhere.”

Sub-question 2-3

What are the strengths and weaknesses of the TEPES regarding teacher ratings?

In summary, the most commonly identified weaknesses of rating teachers using the Stronge model were the partial subjectivity and the lack of clear definition between Effective and Highly Effective. Each performance standard does define Effective versus Highly Effective behaviors, but five administrators felt the definitions were not clear enough. Administrator 13 spoke of having conversations with teachers and explaining the difference between Effective and Highly Effective. “The most contentious conversations I’ve had haven’t been about being Partially Effective; they’ve been about Highly Effective.”

Regarding strengths of the Stronge model as an instrument to rate teachers, four administrators felt that the performance standards were helpful in steering them more towards objectivity. Administrator 1 saw the value in having inter-rater reliability
conversations at district administrative meetings. “The conversation has helped us in clarifying shared expectations and it gives us something to hand to teachers to really back up our decisions that isn’t just our words – it’s an objective tool.” The collaborative aspect was also seen as a benefit by Administrator 7 who feels it is “invaluable” seeing other supervisors’ comments and discussing those evaluations with each other.

**Research Question 3 Themes**

**Research Question 3**

To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System enhances their roles as instructional leaders?

**Findings**

Overall, the Placidville administrators did not find that the Stronge teacher evaluation model enhanced their roles as instructional leaders. While there are some supportive components, such as the self-assessment, most stated that their practice as instructional leaders had not changed significantly since the district’s implementation of the Stronge model.

**Interview Results for Research Question 3**

The final two interview questions focused on each administrator’s role as an instructional leader supporting teachers.

**Sub-question 3-1**

In what ways, if any, has the TEPES impacted your role as an instructional leader?

Six of the administrators did not feel that it played a significant role in their instructional leadership. Common themes were that their instructional goals were
unrelated to the Stronge components and that, in fact, the increased paperwork took time away from their instructional leadership. On a more positive note, the self-assessment was mentioned by four administrators as a helpful component since, according to Administrator 10, “I learn so much about what they do outside that they may not have told me. Sometimes it gives you a different perspective of what you see in the room because you didn’t realize it was from PD they took . . . they did it on their own time.”

Regarding professional development, Administrator 11 said, “It does frame some of the PD I’m steering people in the direction of because it’s either areas that I see as weaknesses or areas that they can improve on or general things I think they’ll enjoy.” Administrator 6 felt that although the mandated aspects of observations and meetings take a lot of time, “this model requires that you get into those classrooms and helps keep us talking to teachers about important things.” Similarly, Administrator 3 noted that the TEACHNJ Act, which mandated the structure of the Stronge evaluation model, has forced more formal classroom visits, which has had positive benefits.

Sub-question 3-2
What are the strengths and weaknesses of the TEPES regarding your role as an instructional leader?

At this point in the interview, most administrators had touched on the already identified common themes, both strengths and weaknesses. Repeated was the notion that using the Stronge model was time consuming and could be cumbersome. Also repeated was appreciation for the defined performance standards and accompanying rubrics. Having “a common language” was mentioned as helping to advance conversations on teaching and learning. Three administrators stated they preferred the narrative style of the
Stronge model as opposed to the checklist style of the Danielson model, which they saw as more limiting in its opportunities for strong conversations. Two administrators, who both have K-12 responsibilities, stated they did not care for the one-size-fits-all nature of the Stronge model, though arguably that can be said about each of the state-approved teacher evaluation systems.

To close the interviews, each administrator was asked if there was anything else he or she wanted to add regarding the Stronge model that had not been discussed. Five administrators mentioned the Student Growth Objective as a problematic component of the evaluation system. SGOs are required by TEACHNJ; thus, Stronge and other evaluation designers built them in. Administrator 11 felt strongly that a teacher’s SGO has the ability to inflate the overall rating into a category that may not be appropriate for that teacher. Administrator 13 said, “The SGO process needs a lot of improvement,” while Administrator 4 stated, “I don’t know if it’s a good judge of their instruction.”

**Summary**

Chapter 4 highlighted the overall findings of each of the three research questions from the 14 administrator interviews. Within each broad research question, more detailed information was provided regarding the interview questions and included some specific statements made by the administrators.
CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Introduction

The primary purpose of this study was to examine the perceptions of school administrators using a specific teacher evaluation tool, James Stronge’s Teacher Effectiveness Performance Evaluation System. Stronge’s TEPES was one of five teacher evaluation systems approved by the New Jersey Department of Education to meet the mandates issued in 2012 under the TEACHNJ Act. This study addresses one of the recommended evaluation systems from an administrator’s point of view. In general, research on teacher evaluation systems has sought the perspectives of teachers more than the administrators charged with evaluating teachers, although administrators are an essential part of the system. This study provides insight towards understanding administrators’ perspectives regarding the process and outcomes, using one specific evaluation tool.

This was a qualitative study, an appropriate methodology to explore the issue of teacher evaluation and empower the research participants to share their opinions and perceptions (Creswell, 2007). The study sample consisted of 14 administrators from Placidville, a northern New Jersey suburban district: 8 females and 6 males. All had used the Stronge evaluation system for a minimum of two years; 11 of the participants had used the Stronge system for all five years it had been in place. The study participants also reflected an average of 10 years of administrative experience among them, an average age of 45, and an overall even sampling of elementary, secondary, and K-12 district
responsibility. The research was conducted in the spring and summer of 2017 through individual interviews with each of the participants.

The three research questions were as follows: (1) To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System improves teachers’ instructional practices? (2) To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System proficiency scale (Highly Effective, Effective, Partially Effective, Ineffective) is an accurate reflection of teacher quality? (3) To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System enhances their roles as instructional leaders?

**Summary of Major Findings**

Linda Darling-Hammond (2014) states that an effective evaluation system should be a teaching and learning system that supports continuous improvement for individual teachers and the profession as a whole. There are various elements of the Stronge evaluation tool available to administrators designed to help teachers improve instruction. First and foremost are the performance standards: Professional Knowledge, Instructional Planning, Instructional Delivery, Assessment of/for Learning, Learning Environment, Professionalism and Communication, and Student Progress. Together they are the linchpin of the entire system and refer to the major duties performed by a teacher (Stronge, 2012). Both teachers and administrators are guided to collect evidence within each of the seven standards throughout the year so that administrators can then assign objective ratings on the year-end summative report.
Administrators in Placidville overwhelmingly mentioned the observation process, which can include both pre- and post-meetings as well as the formal observation report, to be the component that has the potential to most impact teachers’ instruction. It is from the observation that Placidville administrators bring attention to strengths and weaknesses in teachers’ Instructional Delivery, followed by a focus on Assessment, Learning Environment, and Instructional Planning – the performance standards that lend themselves most to constructive conversations about instruction as reported by the administrators. Whereas observations have been a standard part of teacher evaluation for decades and have been criticized for their potential to be “drive-by” visits with little value towards directly addressing the quality of instruction (Toch & Rothman, 2008), administrators in Placidville found observations to be the best part of the Stronge evaluation system for promoting instructional growth. In particular, the opportunity to focus on the specific performance standards and common language to describe instruction was viewed positively by most of the administrators in the study.

The 21st century has brought about significant changes and increased focus on teacher evaluation as a result of federal policy efforts such as No Child Left Behind in 2002 and the American Recovery and Reinvestment Act of 2009 (Riordan, Lacireno-Paquet, Shakman, Bocala, & Chang, 2015). Due to reform efforts that have sprung from these policies, teacher evaluations are now multi-dimensional and include student achievement data and student learning objectives when possible in addition to more traditional observation reports and walk-throughs. Teachers and administrators, such as those in Placidville, participate in training in using rubrics and protocols in order to collect and analyze data and evidence of student achievement (Partee, 2012). These
reform efforts are intended to lead towards more objective observations and personnel decisions.

Despite these efforts, all of the administrators in this study felt the Stronge evaluation system did not lead toward completely objective ratings. All indicated there were subjective elements to the evaluation process despite “evidence” on paper. Having developed relationships with teachers and, in many cases, knowing them for years made it difficult to disregard previous knowledge and a more global understanding of the teacher despite what was available as evidence through the evaluation components in the Stronge system in a single year. These findings are supported by the Rutgers University study completed after the 2011-2012 pilot year for new teacher evaluation in New Jersey that found many teachers believing their evaluation results were subjective (Firestone et al., 2013).

As instructional leaders, the administrators in Placidville overall felt the Stronge system had not had a major impact on their approach to professional development and their practice as instructional leaders had not changed significantly in the five years of its implementation. This contradicts some early results from studies performed in Kentucky and New Hampshire that showed a majority of administrators felt the new evaluation systems, as a result of the American Recovery and Reinvestment Act, helped them improve their schools’ instructional programs (Dodson, 2015; Riordan et al., 2015). The Rutgers study also found a majority of administrators surveyed reporting positive effects on professional development (Firestone et al., 2013), although it is notable that the Rutgers study lacked the participation of districts comparable to Placidville in socio-economic status. It is possible that the Placidville administrators, in an upper middle class
suburban community with strong student achievement results, have different instructional and professional development goals as a result.

**Summary of Findings for Research Question 1**

Research Question 1 asked to what extent the administrators perceived Stronge’s TEPES as a tool to improve teachers’ instructional practices. Administrators were asked about the frequency of meetings they had with individual teachers based on the Stronge system, as well as what parts of the TEPES and which performance standards had the most impact on teachers’ instruction. These questions enabled the researcher to focus on the administrators’ views of whether or not their use of the TEPES had supported teachers’ instructional growth.

The administrators recounted the number and purpose of meetings with individual teachers based on various components of the Stronge system. Since they are given some autonomy in Placidville regarding the number of face-to-face meetings they need to have using the evaluation components, the perceptions of value and time varied. For example, only five of the participants always hold pre-observation meetings with teachers face to face. Others have them only with non-tenured teachers, upon request, or rely solely on the pre-observation form filled out by the teacher. The administrators who were proponents of the pre-observation meetings felt strongly that a pre-observation conference led to productive discussion and goal setting towards instructional improvement; however, as with the changes to teacher evaluation practices across the country, it is documented that administrators have experienced a dramatic increase in the amount of time and skill required on their part (Kersten & Israel, 2005). Those
administrators who did not have pre-observation meetings all stated that time was the prohibitive factor.

Conversely, for the year-end summative evaluation, all but two of the administrators hold face-to-face meetings with teachers. While the summative report is a critical component of a teacher’s evaluation, the administrators reported that it did not lend itself to instructional growth as much as the observation report; all but one administrator named the observation report first as the component that most impacted teachers’ instruction with the Instructional Delivery performance standard as their primary focus. Despite the perceived value of the observation report, the majority of administrators felt they lacked the time to meet with teachers both before and after an observation, thereby potentially compromising the value of the observation and its impact on teachers’ instruction. Lack of time to dig meaningfully into the teaching craft could be why the literature is generally critical of observations as the primary tool used to evaluate teachers (Toch & Rothman, 2008) if administrators are not able to maximize the benefit.

**Summary of Findings for Research Question 2**

Research question 2 delved into administrators’ perceptions of Stronge’s four-point proficiency scale (Highly Effective, Effective, Partially Effective, Ineffective) as an accurate reflection of teacher quality. Raising and recognizing the importance of teacher quality is at the core of reform efforts related to evaluation (Hull, 2013). The best evaluation systems are those that help all teachers grow professionally and push good teachers to be great (Almy, 2011). To address the research question, administrators were asked about the Stronge evaluation system’s objectivity and its accuracy as a ratings tool. Moving beyond a simple two-point Satisfactory or Unsatisfactory rating was a significant
change for the Placidville administrators when they adopted the Stronge system in 2012. Although in their initial training of the Stronge system they were encouraged to rely on the rubrics and the preponderance of evidence that was accrued within each standard, all 14 research participants stated that subjectivity entered the process to varying degrees. In other words, the summative evaluation in the Stronge system has not been the objective measuring tool it was designed to be.

The presence of the seven performance standards has helped add objective information and guidance to conversations with teachers as reported by the majority of the administrators, but the sticking point often has been the subtle differences in rating teachers either Effective or Highly Effective. There are not a high number of Partially Effective or Ineffective teachers in Placidville at this time. As a matter of fact, five of the administrators specifically noted contentious conversations they have had with teachers regarding the distinction between Highly Effective and Effective within the performance standard areas and noted their frustration that many teachers had a difficult time seeing beyond the ratings. One administrator stated that his conversations regarding instruction and professional growth opportunities were often diminished with the presence of the ratings system. This finding is corroborated by a recent study of State Teachers of the Year (STOYs) conducted by the Educational Testing Service (2017). Upon surveying 266 STOYs regarding current teacher evaluation systems across the country, one of the recommendations that emerged was to focus more on targeted feedback for professional growth and improving instruction and place less emphasis on an evaluation “score.”

Another essential part of the Stronge system as mandated by NJACHIEVE is value-added measures. New Jersey’s Commissioner of Education in 2012, Christopher
Cerf, outlined a plan to include multiple measures in teacher evaluations to identify both learning outcomes and effective practice. As a result, summative evaluations for teachers in New Jersey are also based partially on Student Growth Objectives (SGOs) for all teachers as well as Student Growth Percentiles (SCPs) for teachers in standardized testing grades and subject areas. While the intent, as stated by Commissioner Cerf, was to “take an important step toward developing a fair, consistent, and learning-centered evaluation system,” some of the Placidville administrators took issue with the fairness of the SGO process specifically and the unintended outcomes the results can have on overall teacher ratings. According to at least six of the administrators in the study, they did not consider the SGOs to be weighted fairly. SGOs account for 15% of a teacher’s overall rating, which can be enough to shift a typically Effective teacher, in the administrator’s eyes, into Partially or Highly Effective ranges. Administrators are then left with decisions regarding leaving it as is or reexamining the performance standards ratings to allow for a more “accurate” rating based on their overall knowledge and experience with a teacher.

The literature supports the fact that value-added measures such as SGOs and SGPs are controversial components of teacher evaluations, have been found to be highly unstable (Darling-Hammond, 2012), and often discount other significant influences on student achievement such as socioeconomic status and parental involvement (McNergney & Imig, 2003). Rather than eliminate value-added measures altogether, Darling-Hammond advocates utilizing assessments that are appropriate for the curriculum and the students being taught. From the STOYs study (ETS, 2017) emerged two significant recommendations regarding value-added measures:
• Consider ways to measure teachers’ contributions to student growth that more accurately reflect students’ progress on important learning goals throughout the year, rather than focusing on results from a single standardized test.

• If standardized test scores are included in teacher evaluation, consider how to ensure fairness and accuracy as well as how much impact test scores should have in a teacher’s overall evaluation score.

**Summary of Findings for Research Question 3**

Research Question 3 asked the study participants if and how Stronge’s Teacher Effectiveness Performance Evaluation System enhanced their roles as instructional leaders and how its use may have impacted that particular aspect of their positions as administrators. Most of the administrators in the study stated that the implementation of the Stronge system did not have a major impact on their roles as instructional leaders. In their views, they had already been instructional leaders with a focus on professional development for their teachers. Other than time needed for implementation, which many stated imposed on their opportunities to be effective leaders, the study participants gave no specific negative feedback regarding the components of the Stronge system related to instructional leadership.

However, as was mentioned in the findings for Research Question 1, the majority of administrators did find the language and detail in the performance standards to be helpful in conversations and dialogue with their teachers. Although the performance standards were not directly identified as a source of support by the administrators in their roles as instructional leaders, the researcher noted that the specific language provided by the standards is most likely a supportive tool.
Without exception, the administrators in this study felt they already had a strong focus supporting their teachers’ growth, and they hadn’t seen a significant change in that area of responsibility with the implementation of the Stronge system. As stated earlier, Placidville is an upper middle class, high-performing suburban school district. As a result, the administrators have been able to craft professional development opportunities for teachers in areas other than success on standardized assessments, a choice more common to districts like Placidville. Perhaps there are hidden opportunities within the data from the Stronge evaluation system that are being overlooked due to a comfort level not afforded to lower socioeconomic districts.

**Findings Related to Theoretical Framework**

Reflective practice encourages educators to develop a greater level of self-awareness regarding the nature and impact of their performance, creating opportunities for professional development and growth (Osterman & Kottcamp, 1993). In this study, administrators utilized reflective practice throughout the research process as they addressed the interview questions thoughtfully and openly. They were direct and honest with the researcher, reflecting on the nature of the Stronge evaluation system and their own use of it to support teachers while also making challenging personnel decisions.

Dewey (1933) identified open-mindedness as a significant attitude in the process of reflective thinking. Overall, the Placidville administrators demonstrated a willingness to be critical of not only the model but of their own practice, thinking of alternative possibilities within the evaluation system. The interviews often evolved into a conversation about evaluation in general as well as aspects of the Stronge system they would like to change or implement more effectively. Overall, the administrators of
Placidville proved themselves to be responsible, ethical professionals and practitioners on the front lines of teacher evaluation.

**Recommendations for Practice**

The study participants were asked to reflect on the strengths and weaknesses of the Stronge evaluation model throughout the interview process. From their responses came insightful recommendations for practice for districts using the Stronge model or one that is similar. Since the model is used in conjunction with an online tool, there are opportunities for using longitudinal data as a resource for guiding professional development goals, a mostly untapped resource to this point. Some administrators surveyed pointed to looking back at the previous year’s information prior to completing the summative reports, but there are broader opportunities for the longitudinal evaluation data. The data could be mined for trends regarding individual teachers, individual administrators, schools, departments, and districts as a whole. From those data could emerge unseen or unrecognized opportunities for professional development and growth.

Second, based on differences of implementation of various evaluation components in the Stronge system as discovered in this study, districts who utilize this evaluation tool would benefit from regular conversation about best practices of the individual components. For example, should there be common expectations for pre- and post-observation meetings? Should summative meetings with teachers follow a common procedure? Should teachers in all schools and departments be asked to complete the self-assessment? Which components could support the identification of professional development goals? Discussion of questions such as these could lead to a common understanding of the purposes behind the different components of the evaluation system,
thereby supporting a common understanding regarding the purpose of evaluation overall. The potential for administrators to grow in their roles as instructional leaders could be enhanced by working together and engaging in more dialogue on teacher evaluation, regardless of the specific evaluation tool used.

One of the benefits of using the Stronge evaluation system is that districts have some autonomy to shape it to best fit their needs. For example, there are optional components such as the self-assessment, student surveys, and a parent communication log. In addition, the language used in the rubrics for the seven performance indicators can be revised and customized by each district. With five years of experience using the evaluation tool since the adoption of TEACHNJ in 2012, districts can take advantage of the opportunity to revise and revisit in order to reexamine their evaluation choices and practices. The possibility even exists for districts to create their own evaluation tool following DOE guidelines now that time and experience have given them perspective on what works and what does not.

Finally, based on administrators’ views in this study of the overall impact of Student Growth Objectives on the evaluation process, districts who find the impact problematic should consider further training for both administrators and teachers to ensure quality and accuracy. Since the inclusion of SGOs is a state mandate and the fact that all SGOs must be approved by administrators in advance, crafting SGOs to best reflect both the curriculum and student growth requires constant vigilance. The STOYs study recommended assurance that “evaluators are trained in the processes involved so they can provide guidance during the process and accurately assess outcomes” (ETS,
Continued training beyond the initial implementation of SGOs would be beneficial since their impact can be a game changer.

**Recommendations for Policy**

In its role in teacher evaluation reform, the National Education Association stated that ongoing, consistent formative assessments foster professional growth and improved practice while periodic summative evaluations of performance support employment decisions (2010). In reality, however, districts in New Jersey are using one evaluation system to accomplish both goals. Administrators are asked to use the Stronge system, or one of the other state-approved evaluation systems, as both a growth model and a ratings tool. As pointed out by several of the administrators in this study and in the literature, it is difficult to promote growth when teachers are focused on a bottom line rating within a 4-point number scale. If the evaluation systems are not contributing to improved dismissal practices and processes, then perhaps the state should reconsider using number ratings at all. With or without numbers attached, the terms *Highly Effective, Effective, Partially Effective,* and *Ineffective* could be fleshed out to a greater extent for equal application across districts. The controversial value-added measures such as Student Growth Objectives and Student Growth Percentiles are added to the mix and further complicate an already complicated tool. When one evaluation system is asked to accomplish so many goals, unfortunately all of those goals seem to be compromised.

Policymakers should reconsider the one-size-fits-all-districts and one-size-fits-all-purposes nature of the current evaluation tools. This study reflects a need to utilize a formative tool to promote growth and a separate and distinct summative tool to inform personnel decisions in order for teacher evaluation to accomplish its goals.
In addition, policymakers should reexamine the use of value added measures in the summative assessments for teachers. Limitations of VAMs are well known (Hull, 2013) and should be considered carefully now that we have several years of evidence to examine in the state of New Jersey.

**Recommendations for Future Research**

This study addressed one teacher evaluation system in one school district. It could be replicated in other school districts in New Jersey that use James Stronge’s Teacher Effectiveness Performance Evaluation System. In addition, the study could be replicated using any of the other state-approved evaluation systems. The uniqueness of this study is approaching it from an administrator point of view as opposed to that of a teacher, but it would also be fascinating and insightful to conduct research in one district gathering perspectives from both administrators and teachers.

A quantitative study examining the ratings of teachers in one district, several districts, or across the state since 2012 would be beneficial. As defined by the NJ DOE, TEACHNJ was enacted to “raise student achievement by improving instruction through the adoption of evaluations that provide specific feedback to educators, inform the provision of aligned professional development, and inform personnel decisions” (TEACHNJ Guide, 2014). Have those goals been achieved? Comparing teacher evaluation data prior to 2012 to data accrued since the implementation of the new evaluation tools, such as the TEPES, would help shine a light on the successes and/or failures of the new law.
Finally, this study and others like it have focused on specific evaluation systems in the United States. Further study of current teacher evaluation in other countries could shed light and add insight on our own practices.

Summary

Chapter 5 offered conclusions of the study as well as recommendations based on the findings. In addition to a summary of the major findings, the findings within each of the three research questions were analyzed. Based on those findings, there were recommendations for school districts utilizing the Stronge evaluation system, recommendations for policy in the state of New Jersey, and recommendations for future research to extend the findings of this study.
References


Ali Cheraghi, M., Sanjari, M., Bahramnezhad, F., Fomani, F. K., & Mahnaz, S. (2014). Ethical challenges of researchers in qualitative studies: The necessity to develop a specific guideline. *Journal of Medical Ethics and History of Medicine, 7*(14).


Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a


Appendix A

Letter of Solicitation
Dear Administrators:

You are invited to take part in a research study. As a student at Seton Hall University in the College of Education and Human Services, Jean Schoenlank is conducting research in partial fulfillment for a Doctor of Education degree in the Educational Leadership, Management, and Policy Department.

She has received permission from Superintendent Dan Fishbein to conduct her research in Ridgewood. The purpose of her study is to explore the perceptions of administrators regarding James Stronge’s Teacher Effectiveness Performance Evaluation System. As educators on the front lines using this specific system for teacher evaluation, your perceptions can provide valuable insight into the evaluation process and its impact on both teachers and administrators’ practice.

Administrators who participate in the study will be interviewed for approximately 30 minutes using questions that have been reviewed by a panel of experts. Interviews will be conducted in a mutually agreeable location, most likely your office. The interview will be recorded and transcribed. You will be provided a transcript of the interview to review for accuracy. All participants will be required to sign an informed consent form.

Approximately 14 participants are being sought for the study. Each volunteer will receive a Demographic Profile Questionnaire to be completed. From those profiles, administrators will be selected who meet the criteria needed to participate: certified administrators who have used the Stronge teacher evaluation system for at least two years.

Participation in this study is voluntary. You are free to withdraw from the study at any time and there is no penalty for doing so. You also have the right to refuse to answer any question in the interview process.

In addition, your participation in this study will remain confidential. All data collected from you will be coded so that your identity will be known only to the researcher. Your name will not be used in either the research or reporting phases of this study.

All research data connected with this study will be stored on a USB memory drive and stored in a locked cabinet. Once the recording files of interviews have been transcribed, they will be erased. Research data will be kept for three years and then it will be destroyed.

If you are interested in participating in this research study or want to discuss your participation in further detail, please contact me by email or reach out to me by phone: Pquinlan52@gmail.com or 201-669-1397.

Thank you for your consideration.

Peggy Quinlan
Assistant to the Principal
Ridge School
Appendix B

Demographic Profile Questionnaire
Demographic Profile Questionnaire

Date: _____________________

First name: ________________  Last Name: __________________________

<table>
<thead>
<tr>
<th>Years of experience as a certified administrator:</th>
<th>Years of experience as an administrator in this district:</th>
<th>Current Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>___ 2-5 years</td>
<td>___ 2-5 years</td>
<td>___ 20-29</td>
</tr>
<tr>
<td>___ 6-10 years</td>
<td>___ 6-10 years</td>
<td>___ 30-39</td>
</tr>
<tr>
<td>___ 11-15 years</td>
<td>___ 11-15 years</td>
<td>___ 40-49</td>
</tr>
<tr>
<td>___ 16-20 years</td>
<td>___ 16-20 years</td>
<td>___ 50-59</td>
</tr>
<tr>
<td>___ 21-25 years</td>
<td>___ 21-25 years</td>
<td>___ 60+</td>
</tr>
<tr>
<td>___ 26+ years</td>
<td>___ 26+ years</td>
<td></td>
</tr>
</tbody>
</table>

Years using Stronge evaluation model (check all that apply):

___ 2012-13 (pilot year)
___ 2013-14
___ 2014-15
___ 2015-16

Gender:

___ Female  ___ Male

What level(s) do you supervise?

___ Elementary  ___ Middle School  ___ High School

Your position: _____________________
Appendix C

Informed Consent Form
Informed Consent Form
Dissertation: Exploring School Administrators' Perceptions of the James Stronge Teacher Evaluation System

Researcher’s Affiliation
The researcher for this study is Jean Schoenlank, a doctoral student at Seton Hall University, College of Education and Human Services.

Purpose of the Research
The purpose of this study is to identify school administrators' perceptions of the Stronge Teacher Effectiveness Performance Evaluation System (TEPES) and whether the evaluation system fosters teachers' instructional practices, enhances administrators' roles as instructional leaders, and whether it is an accurate reflection of teacher quality. Participation in this research will require approximately one hour of each participant's time for the interview and for review of the interview transcript.

Description of the Procedures
Participants who choose to take part in the study will be asked to be part of a half hour face-to-face interview with the researcher in a location that is convenient for each participant (i.e., their school offices). Approximately 18 administrators will be interviewed. Interviews will be audio recorded and transcribed. Specific questions will be asked with appropriate follow-up questions based on the participants' responses. The following are samples of questions that will be asked during the interviews:

- How many times per year do you meet with a teacher using the TEPES components and for what purposes?
- To what extent is the TEPES an objective instrument that supports the ratings produced using the proficiency scale?

Instruments
Interview questions will be the sole source of research in this study.

Voluntary Participation
Participation in this study is voluntary; an administrator may decline to participate without penalty and may withdraw from the study at any time without penalty. If the participant feels uncomfortable during the interview process, he/she has the right to decline to answer any question or end the interview without penalty.

Anonymity
Participants' identities will remain confidential. Each participant will be given a code number known only to the researcher. The researcher will not disclose who participated in the study or disclose the identity of individual responses.

Confidentiality
All data associated with the study will remain strictly confidential. All documents obtained from the research including transcripts, digital recordings, and other printed materials will be kept in a locked filing cabinet to protect the anonymity of the participants. All data will be kept for a period of three years.

Records
Only the researcher and members of the dissertation committee will have access to the data. The dissertation committee is obligated to protect the data from disclosure.
Informed Consent Form

Dissertation: Exploring School Administrators’ Perceptions of the James Stronge Teacher Evaluation System

Researcher’s Affiliation
The researcher for this study is Jean Schoenlank, a doctoral student at Seton Hall University, College of Education and Human Services.

Purpose of the Research
The purpose of this study is to identify school administrators’ perceptions of the Stronge Teacher Effectiveness Performance Evaluation System (TEPES) and whether the evaluation system fosters teachers’ instructional practices, enhances administrators’ roles as instructional leaders, and whether it is an accurate reflection of teacher quality. Participation in this research will require approximately one hour of each participant’s time for the interview and for review of the interview transcript.

Description of the Procedures
Participants who choose to take part in the study will be asked to be part of a half-hour face-to-face interview with the researcher in a location that is convenient for each participant (i.e., their school offices.) Approximately 18 administrators will be interviewed. Interviews will be audio recorded and transcribed. Specific questions will be asked with appropriate follow-up questions based on the participants’ responses. The following are samples of questions that will be asked during the interviews:

- How many times per year do you meet with a teacher using the TEPES components and for what purposes?
- To what extent is the TEPES an objective instrument that supports the ratings produced using the proficiency scale?

Instruments
Interview questions will be the sole source of research in this study.

Voluntary Participation
Participation in this study is voluntary; an administrator may decline to participate without penalty and may withdraw from the study at any time without penalty. If the participant feels uncomfortable during the interview process, he/she has the right to decline to answer any question or end the interview without penalty.

Anonymity
Participants’ identities will remain confidential. Each participant will be given a code number known only to the researcher. The researcher will not disclose who participated in the study or disclose the identity of individual responses.

Confidentiality
All data associated with the study will remain strictly confidential. All documents obtained from the research including transcripts, digital recordings, and other printed materials will be kept in a locked filing cabinet to protect the anonymity of the participants. All data will be kept for a period of three years.

Records
Only the researcher and members of the dissertation committee will have access to the data. The dissertation committee is obligated to protect the data from disclosure.
### Research Question 1
To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System fosters teachers’ instructional practices?

**Sub-Questions**
1. How many times per year do you meet with a teacher using the TEPES components and for what purposes?
2. What specific parts of the TEPES, if any, have the most impact on teachers’ instructional practices and how?
3. Of the seven TEPES standards, which ones, if any, effectively foster teachers’ instructional practices?
4. What are the strengths and weaknesses of the TEPES regarding teachers’ instructional practices?

### Research Question 2
To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System proficiency scale (Highly Effective, Effective, Partially Effective, Ineffective) is an accurate reflection of teacher quality?

**Sub-Questions**
1. How does your pre-existing knowledge of teachers affect your ratings of them using the TEPES?
2. To what extent is the TEPES an objective instrument that supports the ratings produced using the proficiency scale?
3. What are the strengths and weaknesses of the TEPES regarding teacher ratings?

### Research Question 3
To what extent, if any, do administrators perceive that the use of the Stronge Teacher Effectiveness Performance Evaluation System enhances their roles as instructional leaders?

**Sub-Questions**
1. In what ways, if any, has the TEPES impacted your role as an instructional leader?
2. What are the strengths and weaknesses of the TEPES regarding your role as an instructional leader?

### Additional Question
Is there anything about the TEPES that has not been discussed with you that you would like to add to this interview?
Appendix E

School’s Permission Granted
Ms. Jean Schoenlank, Principal
Ridge School
Dear Jean:

Your doctoral research request was reviewed and permission has been granted for you to conduct research in the Ridgewood school district entitled “Administrators’ Perceptions of the Strong Teacher Evaluation Model.”

As stated, your research will be conducted with complete anonymity and confidentiality, and you will practice the highest level of responsibility in keeping with the New Jersey State Statutes, as well as the policies of the Ridgewood Public Schools.

I wish you success with this endeavor.

Sincerely yours,

Daniel Fishbein, Ed.D.
Superintendent of Schools
Appendix F

IRB Approval
June 14, 2017

Jean Schoenlank

Dear Ms. Schoenlank,

The Seton Hall University Institutional Review Board has reviewed the information you have submitted addressing the concerns for your proposal entitled “Exploring School Administrators’ Perceptions of the James Stronge Teacher Evaluation System”. Your research protocol is hereby approved as revised through expedited review. The IRB reserves the right to recall the proposal at any time for full review.

Enclosed for your records are the signed Request for Approval form and the stamped original Consent Form. Make copies only of this stamped form.

The Institutional Review Board approval of your research is valid for a one-year period from the date of this letter. During this time, any changes to the research protocol must be reviewed and approved by the IRB prior to their implementation.

According to federal regulations, continuing review of already approved research is mandated to take place at least 12 months after this initial approval. You will receive communication from the IRB Office for this several months before the anniversary date of your initial approval.

Thank you for your cooperation.

In harmony with federal regulations, none of the investigators or research staff involved in the study took part in the final decision.

Sincerely,

Mary F. Ruzicka, Ph.D.
Professor
Director, Institutional Review Board

cc: Dr. Barbara Strobert

Office of Institutional Review Board

Presidents Hall · 400 South Orange Avenue · South Orange, New Jersey 07079 · Tel: 973.313.6314 · Fax: 973.275.2361 · www.shu.edu

A HOME FOR THE MIND, THE HEART AND THE SPIRIT
Appendix G

Certificate of Completion NIH
Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that Jean Schoenlank successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 02/20/2017.

Certification Number: 2331505.