January 2017

A Midwest School District's Implementation Process Of A New Teacher Evaluation Model

Kristopher G. Arason G. Arason

A MIDWEST SCHOOL DISTRICT'S IMPLEMENTATION PROCESS OF A NEW TEACHER EVALUATION MODEL

by

Kristopher G. Arason
Bachelor of Arts, University of North Dakota, 1992
Bachelor of Science, Mayville State University, 1996
Master of Education, University of North Dakota, 2001

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

May
2017

This dissertation, submitted by Kristopher G. Arason in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Dr. Steven D. LeMire

Dr. Pauline Stonehouse

Dr. Gail Ingwalson

Dr. Joshua Hunter

Dr. Jason Jensen

This dissertation is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Grant McGimpsey
Dean of the School of Graduate Studies

May 1, 2017
Date

PERMISSION

Title A Midwest School District's Implementation Process of a New Teacher Evaluation Model

Department Educational Leadership

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his/her absence, by the chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Kristopher G. Arason
March 29, 2017

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

ACKNOWLEDGMENTS

During the summer of 1994, I was working in the family business, Arason's Body Shop. Since college, it had always been in the plans for me to take over the family business, following in the footsteps of my grandfather and my dad. However, late one summer afternoon, I was sitting in the office and realized that I would not be happy doing this for the rest of my life. In the back of my mind, I had always wanted to be an educator. That summer, I told my dad I wanted to go back to college and be a teacher. I am forever grateful to my parents, Gerald and Dolores Arason, for their love, support, encouragement, and understanding as I left the family business to pursue my dream of being a teacher. My parents instilled in me the values of compassion, dedication, determination, and hard work, and for that, I am forever thankful and love them very much. Along the way, there have been so many others that have played an instrumental role in assisting me to complete my education.

First of all, I would like to thank my professors at the University of North Dakota who have helped me become more knowledgeable, challenged me to think critically, and provided me with the confidence to be a writer and researcher. I would also like to thank the members of my dissertation committee; I would not be at this point without a professional, knowledgeable, and dedicated committee. Thank you to Dr. Pauline Stonehouse, Dr. Joshua Hunter, Dr. Gail Ingwalson, and Dr. Jason Jensen. Also, a special thank you to Sharon Fields. I cannot count the number of tasks she helped me with during

my years in the Educational Leadership department to get me to the point I am today. She has been a true icon in the department, and I will be forever grateful for her friendly smile and assistance. Finally, a sincere thank you to Dr. Steven LeMire, my advisor, mentor, and one of my biggest cheerleaders. I have appreciated his assistance, encouragement, feedback, and life lessons he has taught me during our discussions throughout the dissertation process. I will be forever grateful for his inspiration and guidance as my advisor during my doctoral journey.

Secondly, I would like to thank my colleagues in the school district who have supported me throughout my career as an educational leader. Dr. Jeffrey Schatz provided me with my first leadership opportunity, hiring me as an associate principal. As a mentor, he challenged me, yet provided me with excellent guidance and support. Mr. Buck Kasowski continued to provide me with leadership opportunities under his guidance, but most importantly, he taught me how to be a caring and compassionate leader. Also, a sincere thank you to Mr. Jody Thompson and Dr. Larry Nybladh who believed in my educational leadership abilities and provided me with my first opportunity to be a principal. Dr. Terry Brenner allowed me to envision my project and supported me with resources, feedback, and encouragement along the way. Finally, thank you to my current administrative colleagues, Dr. Kelly Peters, Mr. Chris Douthit, and Mr. Tyler Nelson, who have been supportive and understanding as I pursued my degree.

Finally, I am extremely grateful for the love and support of my family. To my three children, Hannah, Jake, and Logan, thank you for understanding why Dad had to spend so much time going to class, reading, studying, and writing his dissertation. During this process, you have all grown up so much, and I am extremely proud of each you.

Lastly, but most importantly, to my wife, Terry, who has been by my side throughout this entire process, thank you for being patient, flexible, supportive, and most of all understanding throughout this journey. I do not know where I would be today without you. I love you, Terry!

To My Family, Terry, Hannah, Jake, and Logan

ABSTRACT

The purpose of this study was to test the effectiveness of the implementation of the Marzano Teacher Evaluation Framework (MTEF) in a Midwest school district. The study used quantitative methods to examine the process used to bring about change in the school district's teacher evaluation model. The researcher examined teacher and administrator perceptions regarding *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with ease of use of the MTEF model.

The researcher used data collected from one Midwest school district. The school district consisted of twelve elementary schools, four middle schools, and three high schools. A total of 682 teachers and 26 administrators were surveyed. Data was collected by way of an on-line survey. The survey included three sections. The first section contained three demographic questions. The second section consisted of five questions regarding the study's research constructs: (a) *change*, (b) *professional development*, (c) *instructional improvement*, (d) *reliability*, and (e) *overall satisfaction* with the MTEF model's ease of use. The final section consisted of 19 questions that aligned with the study's research constructs. For each research construct, there were three to five questions. Data gathered from participants' responses were analyzed and used to provide recommendations to other school districts and educators around the state and nation as they implement new teacher evaluation models.

Keywords: Teacher Evaluation, Danielson, Marzano, and McREL

CHAPTER I

INTRODUCTION

> *We will elevate the teaching profession to focus on recognizing, encouraging, and rewarding excellence. We are calling on states and districts to develop and implement systems of teacher and principal evaluation and support, and to identify effective and highly effective teachers and principals on the basis of student growth and other factors. These systems will inform professional development and help teachers and principals improve student learning.*
> (U.S. Department of Education, 2010, p. 4)

Since the passage of the No Child Left Behind (NCLB) Act of 2001, public schools have entered a new era of school reform. This has included high-stakes testing, data driven decision-making, school choice, and deregulation of schools. It has also included "highly-qualified" teachers, performance pay, and competition among schools (Ravitch, 2010). If "it" could not be measured, it was not important to the politicians that were trying to reform schools by implementing a new business model (Ravitch, 2010). So, what are schools doing to improve student performance and meet the state standards? According to a Common Core Organization (CCO) report from 2009, many schools across the country have narrowed their curriculum to focus on English language arts, mathematics, and test preparation, all at the expense of a well-rounded liberal arts education (Common Core, 2009). However, research has consistently shown that the teacher is the number one factor in determining whether or not students will increase their academic achievement (Greenstone, Looney, & Shevlin, 2011). What are states, school districts, and building leaders doing to ensure classroom teachers are effective and capable of a high level of student achievement?

Teacher evaluation models in K-12 schooling have evolved and are now at the forefront of education (Torff & Sessions, 2009). In the Fall of 2011, President Obama and Education Secretary, Arne Duncan, announced that State Education Associations (SEAs) could apply for Elementary and Secondary Education Act (ESEA) of 1965 Flexibility Waivers that would release states from some of the requirements of NCLB (U.S. Department of Education, 2011). The most notable requirement states would be released from was the 2013-2014 timeline requiring 100 percent proficiency in reading and math for students. In order for states to receive flexibility waivers, each state would have to develop a plan to address three critical areas for improving student achievement, and one of those areas would be evaluating and supporting teacher and principal effectiveness.

> Each State that receives the ESEA flexibility will set basic guidelines for teacher and principal evaluation support systems. The State and its districts will develop these systems with input from teachers and principals and will assess their performance based on multiple valid measures, including student progress over time and multiple measures of professional practice, and will use these systems to provide clear feedback to teachers on how to improve instruction. (U.S. Department of Education, 2011)

President Obama signed the latest change effecting K-12 education into law on December 10, 2015. The new Every Student Succeeds Act (ESSA) of 2015 has required states to continue to set high standards and maintain the accountability that has been put in place as a result of NCLB. However, the new ESSA requirements differ from NCLB by "empowering state and local decision-makers to develop their own strong systems for school

improvement based on evidence, rather than imposing cookie-cutter federal solutions" (U.S. Department of Education, 2015, para. 3).

In 1991, Ferguson concluded the best investment that can be made in education to improve student performance is to invest in the quality of the classroom teacher. According to Marzano's (2003) research, students with effective teachers will learn more than their peers in an academic year. If the trend continues over a 3-year period, the gap could be as much as 50 percentile points between students who have an effective teacher compared to students that do not (Marzano & Pickering, 2003). According to Marzano, Frontier, and Livingston (2011), "What occurs in the classroom [strategies and behaviors] has the most direct causal link to student achievement" (p. 5). Essential for student achievement is requiring every student to work with a high-quality teacher in every classroom (Stronge & Hindman, 2006).

Over the last two decades, educational research has witnessed a significant amount of growth in the area of effective pedagogy and has been able to transfer that research into improving instruction (Marzano & Pickering, 2003; Marzano et al., 2011). At one point in time, it was believed that expertise was considered something that could not be taught, in other words, "It was a gift from the Gods," either you had it or you didn't (Marzano et al., 2011), which is consistent with 1960s education research. Research led by Coleman and fellow colleagues concluded, "Schools bring little influence to bear on a child's achievement that is independent of his background and general social context" (Coleman et al., 1966, p. 325) which was disturbing news for teachers, parents, and society in general.

With the Obama administration, reauthorization of the ESEA, and the importance placed on teacher evaluation models that identify highly effective teachers on the basis of

student growth and other factors, state education agencies have been reevaluating their teacher evaluation models and implementing new ones, in most instances (Ravitch, 2010). What is unknown is how an entire teacher evaluation system will overhaul and implement a new model that measures teacher growth. To address this need, education leaders need to examine a theory of change, how to provide professional development for teachers and administrators, and finally, monitor this change for effectiveness once implementation has taken place. This researcher will examine one Midwest school district's implementation of the Marzano Teacher Evaluation Framework (MTEF). Central to the research will be an (a) examination of strengths and weaknesses of implementation of the MTEF model, (b) determination of perceptions of administrators and teachers of the model's ability to improve instruction, and ultimately (c) recommendations for success of future implementations.

Statement of the Problem

Dating back to the 1980s, teacher evaluation has been a buzzword associated with educational reform. During the Reagan years and with the subsequent *Nation at Risk Report* published in 1983, the teacher accountability movement gained momentum and policy makers began to view teachers as the bottom line to improving education in U.S. schools (Ellett & Teddlie, 2003). More recently, Papay (2012) reported that there has been a consensus building in the U.S. among teachers, administrators, and policy makers that teacher evaluation models have needed fixing. According to Marshall (2005), "The theory of action behind supervision and evaluation is that it will improve teachers' effectiveness and therefore boost student achievement" (p. 728). However, the reality is that most teacher evaluation models used in school districts are ineffective, lack meaning, and have little to no

positive effect on teacher performance in the classroom (Weisberg, Sexton, Mulher, & Keeling, 2009).

Until recently, local school districts have been responsible for teacher evaluations. However, federal programs have been asking states to implement teacher evaluation models that measure teacher and student growth (Stumbo & McWalters, 2011). The U.S. Department of Education (2009) application for "Race to the Top" funding has required states to "design and implement rigorous, transparent, and fair evaluation systems for teachers . . . that take into account data on student growth" (p. 34). For too long, yearly teacher evaluations have been rituals lacking meaning and devoid of context (Ramirez, Lamphere, Smith, Brown, Pierceal-Herman, 2011; Weisberg et al., 2009). Prior research on the topic of effective teaching methodology that leads to an increase in student achievement has been well documented (Marzano, 2007; Danielson, 2007). Teacher evaluations must move past the point of simply being used to satisfy state requirements, and in some cases, terminate a teacher's contract. Instead, they should be used to "activate (or amplify) a supervisory voice inside the teachers' head that will guide them in their work with students" (Marshall, 2005, p. 730).

Purpose of the Study

The purpose of this study was to test the effectiveness of implementation of the Marzano Teacher Evaluation Framework (MTEF) in a Midwest school district. Teachers and administrators were surveyed regarding their perceptions of this implementation. Research constructs that were studied included: change, professional development, and perceptions of the model and the educators' ability to improve instruction, its reliability, and participants overall satisfaction with the model. To support the literature review, the researcher examined:

history of teacher evaluation, trends in teacher evaluation, current teacher evaluation models, theory of change, professional development, instructional improvement, and reliability.

In the field of teacher evaluation, there exists a body of research indicating that a highly qualified teacher is essential for student achievement (Danielson, 2007; Marzano et al., 2011; Dean, Hubbell, Pitler, & Stone, 2012). According to Hargreaves (2009), "High quality learning depends on high-quality and highly qualified teachers and teaching" (p. 28). In research conducted by Marzano and Waters (2009), results showed teachers improve their expertise from year to year with effective teacher supervision; and with modest improvements in teachers' skills, student achievements results were significant. However, missing from the research was how to effectively overhaul a school district's entire teacher evaluation system and implement research-based practice with fidelity.

Research Questions

1. What are the perceptions of one school district's administrators regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?
2. What are the perceptions of the school district's teachers regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?
3. Was there a difference between the school district's administrators and teachers regarding the five research constructs that frame this study, which include *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with the new Marzano Teacher Evaluation Framework model?

Importance of the Study

At the inception of NCLB, policy makers were concerned with "highly-qualified" teachers, which was measured by the courses teachers had taken in preparation for their teaching career (Stumbo & McWalters, 2011). At the time of this study, a shift moved the focus towards teacher effectiveness and how well they perform with their students (Stumbo & McWalters, 2011). With the recent shift (at the time of this report) of the U.S. Department of Education's stance on teacher effectiveness, Race to the Top funding, Flexibility Waivers, and ESSA, states and school districts around the country have been looking for ways to evaluate and improve the practice of individual teachers by using unbiased, reliable, and valid measurement instruments (Marzano, 2012; Maslow & Kelley, 2012; Papay, 2012). With the rapidly changing environment of teacher evaluation in the United States, the purpose of this study was to analyze implementation of the MTEF in a Midwest school district, evaluate its strengths, and provide other school leaders with a blueprint for success.

Theoretical Framework

The theoretical framework that guides this study is based on Fullan's theory of educational change. Fullan's (2011) work examined Drivers for Whole System Reform. Drivers are the levers that have the best chance of successfully implementing reform. Fullan (2011) also identified what he called "wrong drivers" that on the surface look like they would work to achieve a desired result, but when used have little chance of succeeding. There are four criteria Fullan (2011) used to judge the likelihood a driver would be effective in bringing about change; all of these criteria must be met concurrently in order for a driver or drivers to be effective in implementing change, they are:

1. foster intrinsic motivation of teachers;
2. engage educators . . . in continuous improvement regarding instruction and learning;
3. inspire [collaboration and] . . . team work; and
4. affect all teachers and students (p. 3)

Combining intrinsic motivation, instructional improvement, collaboration, and inclusion of all stakeholders are crucial elements to move whole system reform forward (Fullan, 2011). The right drivers are effective because they work directly on changing the culture of a school, which Sarason (1995) described as underestimated and underappreciated force that drives change. According to Fullan (2011), schools and districts that do not pay attention to these four criteria are destined for failure. Figure 1 displays criteria needed for effective drivers.

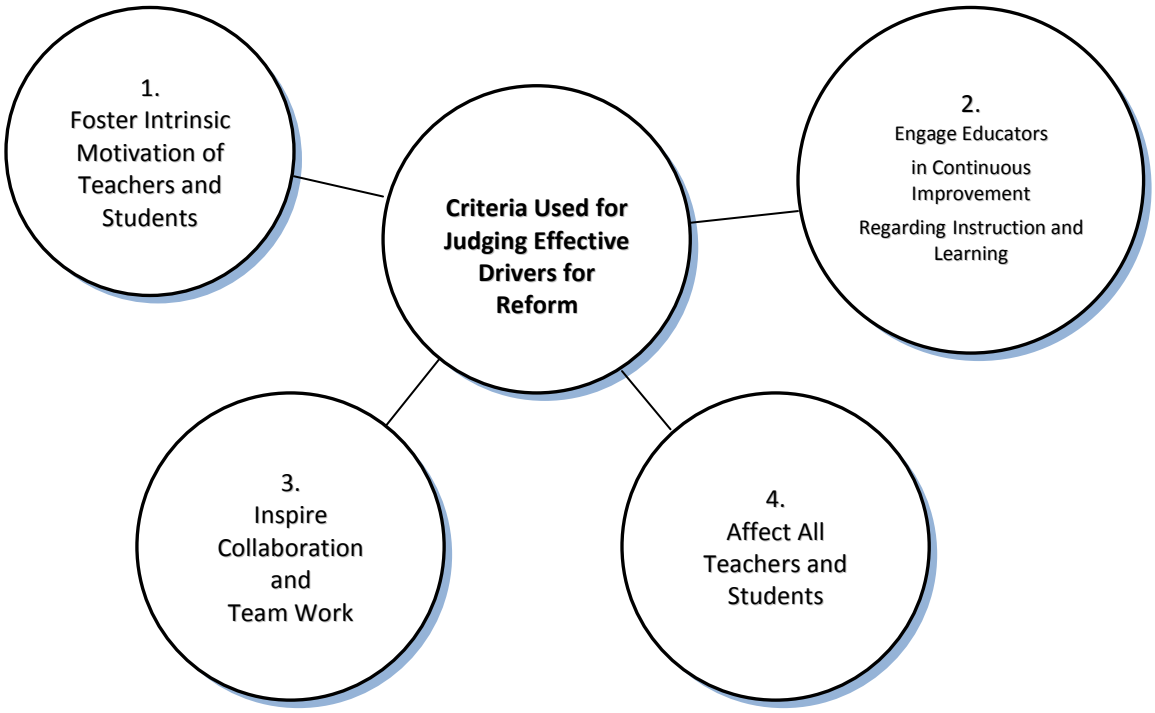

Figure 1. Criteria Used to Judge the Effectiveness of Drivers Used in System-Wide Reform (based on work by Fullan, 2011).

Effective drivers can be defined as strategies that are used to produce better results across an entire system whether it is school, district, state, or on a national level (Fullan, 2011). Simply stated, effective drivers increase student achievement results measurably. Fullan (2011) was careful to point out that effective drivers are not urgent, quick fixes to a problem. Instead, they take time and are effective because they work directly on changing culture. Leaders who rush forward and strive for quick fixes, typically select wrong drivers for change, which has been the case in the United States education system (Fullan, 2011). There are four criteria Fullan (2011) identified as characteristics of drivers that make situations worse, instead of better:

1. **accountability**: using test results, and teacher appraisal, to reward or punish teachers and schools vs capacity building;
2. **individual teacher and leadership quality**: promoting individual vs group solutions;
3. **technology**: investing in and assuming that the wonders of the digital world will carry the day vs instruction;
4. **fragmented strategies** vs integrated or systematic strategies. (p. 5)

Drivers composed of some or all of these four criteria are not always wrong, but they are badly placed lead drivers; and if used for whole system reform, failure is sure to follow, like it has in the United States (Fullan, 2011). Table 1 compares Fullan's criteria of right drivers to criteria of wrong drivers.

Table 1. Comparison of Fullan's Drivers for Change, Wrong Drivers Versus Right Drivers.

Fullan's Drivers for Change	
Characteristics of Wrong Drivers	Characteristics of Right Drivers
Accountability by Individual Appraisal	Capacity Building
Individual Solutions	Collaboration Promoting Group Solutions
Too Much Emphasis on Technology	Pedagogy Emphasized Over Technology
Fragmented Strategies	Systems or Integrated Strategies

The key to system-wide reform is to empower teachers and students to be the central driving force behind reform. When goals and reform initiatives are aligned with participants' needs, intrinsic energy is created and drives change (Darling-Hammond, 2013; Fullan, 2011; Sutton & Shouse, 2016). New policies and strategies must create an environment in which intrinsic motivation is able to flourish (Fullan, 2011). People are motivated by initiatives that are personally meaningful, make a difference, are part of a team effort, and build new skills (Chenoweth, 2016; Darling-Hammond, 2013; Fullan, 2011). According to Fullan (2011), both strong motivation and increased skills are necessary for change to occur.

The United States education system has seen numerous accountability and increased rigor initiatives over the past 30 years with little or no progress in student achievement scores (U.S. Department of Education, 2010). According to Fullan (2011), the United States has traditionally used the wrong drivers to enact change; leading with accountability, assessments, and rewards, which can only tighten up a loose system, but will not create a condition that will sustain system-wide reform. Unfortunately, for the United States, the wrong drivers have been used to move an education system forward. The current models do not build widespread capacity or increase intrinsic motivation. The wrong drivers are ineffective because they fail to address the culture of a school system (Chenoweth, 2016;

Firestone, 2014; Fullan, 2011; Minnici, 2014). Accountability, high standards, and assessments are tools to use along the way, but they cannot change a system (Fullan, 2011).

The theoretical framework for this study will be discussed in greater detail in Chapter II. In addition, it will be used in the discussion section of Chapter V when drawing conclusions and developing recommendations based on teachers' and administrators' perceptions of the MTEF.

Scope of the Study

This study will analyze the implementation of the MTEF in a Midwest school district. The process was implemented over a 5 year time period. The implementation process began at the start of the 2012-2013 school year and concluded in May of 2017. The researcher analyzed the study's data regarding implementation to date by surveying teachers and administrators. Data regarding teachers' perceptions of the implementation were gathered during the Fall of 2013, and the administrators' data was gathered in the Spring of 2014.

Definition of Terms and Acronyms

The paper includes the following terms, definitions, and acronyms:

Administrator: An administrator is defined as a school-level leader who completes evaluations on teachers. Administrators include principals, associate principals, and BRCs.

Building Resource Coordinators (BRCs): An elementary level employee for schools with an enrollment of over 400 students. A BRC serves as an assistant to the principal, works on a teacher's contract, and receives a stipend for their administrative duties. BRCs have administrative credentials, and they supervise and evaluate teachers.

Elementary Secondary Education Act (ESEA): A law passed in 1965 as part of the "War on Poverty." The goal was to equalize education access and establish high standards of accountability.

Every Student Succeeds Act (ESSA): An update to the No Child Left Behind Act of 2001 and a reauthorization of the Elementary and Secondary Education Act of 1965. ESSA was signed into law by President Obama on December 10, 2015.

Marzano Teacher Evaluation Framework (MTEF): The teacher evaluation process that is being implemented in the school district. The model is based on Robert Marzano's book *The Art and Science of Teaching* (2007).

Nation at Risk: A 1983 report conducted by the National Commission on Excellence in Education (NCEE) under the direction of President Ronald Regan. The report's findings that American schools were failing set off an education reform movement.

No Child Left Behind (NCLB): A federal law enacted in January of 2001 that set high standards and established measurable goals in order to improve public education. NCLB was part of a reauthorization of the ESEA.

Race to the Top: A federal education program that was part of the American Recovery and Reinvestment Act (ARRA) of 2009. The act's goal was to create education innovation and reform at the state and local level. States that were interested in the initiative and received funding were required to meet policies set forth by the federal government.

State Education Association (SEA): The governing body of each state that is responsible for implementing and monitoring state and federal education laws.

Teacher Evaluation: A process used to observe a teacher's knowledge of subject matter, teaching techniques, classroom management, and their ability to improve student learning.

Assumptions

1. All participants will answer questions honestly.
2. All administrators will implement the teacher evaluation process with fidelity.

Delimitations

This study was limited to perceptions of teachers and administrators of one Midwest school district that is in the early stages of implementing the MTEF.

Researcher's Background

The researcher has been an educator for the past 21 years. During that time, he has worked in two different school districts and has served each of those districts in numerous capacities. In addition to being a teacher and administrator, the researcher has served on numerous school and district level committees, coached and advised student groups, and served as summer school program director.

As a teacher for 10 years and an administrator for the past 11 years, the researcher has experienced both sides of the teacher evaluation process and has seen first-hand the need for an improved process. In his 10 years as a teacher, he had approximately 20 teacher evaluations conducted on him by principals. Of those 20 evaluations, the feedback and the processes used by principals varied. In the researcher's position as a principal for the past 6 years (at the time of this study) and an associate principal for 5 years previous to his position at the time of this research, the researcher has completed approximately 200 teacher evaluations using the school district's previous teacher evaluation model. With the school

district's previous evaluation model, the researcher struggled to find a consistent way to provide teachers with meaningful feedback to improve their instruction. The previous teacher evaluation model used by administrators had been subjective. It had no standards that teachers were evaluated against, which led to each teacher being evaluated differently depending on the administrator who was writing the evaluation.

Organization of the Study

This study has been organized into five chapters. Chapter II provides a brief history of teacher evaluation and supervision in K-12 education along with current trends. In addition, the literature review addresses change theory, professional development, instructional improvement, and reliability. Chapter III presents the design and methodology of the study. Chapter IV provides the data and results. Chapter V provides a summary, conclusion, discussion, recommendations, and reflections on the study.

CHAPTER II

LITERATURE REVIEW

Introduction

Teacher evaluation has been a topic that has generated a great deal of interest over the past decade. With the federal government's "Race to the Top," a grant initiative funded by the American Recovery and Reinvestment Act (ARRA) of 2009 (U.S. Department of Education, 2009), and the Every Student Succeeds Act (ESSA) of 2015, the improvement of teachers' classroom instructional skills has become, and will continue to be, a focus of federal, state, and local education policy initiatives. While it was the federal government pushing for accountability and growth through the use of teacher evaluation, states and local school districts were left with the challenging task of successfully implementing a new teacher evaluation model that meets federal guidelines. According to Donaldson (2009), previous teacher evaluations have failed to improve teachers' instructional performance and student learning. Instead, teacher evaluations have suffered from what is known as the "Lake Wobegon Effect" where most, if not all, teachers in schools were rated as satisfactory (Donaldson, 2009). As a result, school districts must not only make structural changes to improve teachers' skills, but more importantly, change the culture of teacher evaluation from a fixed mindset to a growth mindset.

This chapter includes a general synthesis of the literature regarding teacher evaluation from 1643 to 2016, so the reader may gain an understanding of the challenges and models

previously used in schools to evaluate performance of teachers. Following this historical review, an explanation of value added and standards based teacher evaluation models are presented to provide the reader with an understanding of the types of data that can be collected and used to evaluate teachers in order to satisfy federal requirements. Next is an overview of three different teacher evaluation models that are used in school districts around the state and country in order to provide the reader with an understanding of current trends at the time of this report. Additionally, Chapter II examines research from which the five constructs that frame this study are derived, and factors to consider when implementing a new teacher evaluation model in a school district.

History of Teacher Evaluation—1643 to 2016

The supervision and evaluation of teachers dates back to colonial times when public schools were first established (Ayer, 1954; Marks, Stoops, & King-Stoops, 1978). On January 2, 1643, the first free public school supported by taxpayers was established in Dedhem, Massachusetts. The teacher was a man by the name of Ralph Wheelock, an ordained priest (Cremin, 1976). When schools were first established, education was not considered a professional discipline; there was no existing structure in place for the education, hiring, or supervision of teachers (Rury, 1991). Decisions regarding teachers were the responsibility of local governments and clergy (Burrup, 1960; Grieder & Romine, 1965). Oftentimes, responsibility for overseeing a town's lone teacher rested in the hands of the clergy because of their level of formal education and the emphasis that was placed on religious instruction in schools (Tracy, 1995).

Throughout colonial times and into the early 1800s, teachers were considered servants to a community and were expected to respond to their community's directives

(Grieder & Romine, 1965). With an emphasis on local control of education at this time in American History, community leaders composed of merchants, clergy, and representatives from various professions formed committees to establish a school's schedule, curriculum, discipline guidelines, and hiring of teachers (Tracy, 1995). These committees were given power not only to establish school guidelines, but to ensure they were being implemented (Marks, Stoops, & King-Stoops, 1985).

Supervision of teachers during the late 1600s-1700s varied significantly depending on a community and their supervisory committee. Frequency of teacher performance observations ranged from monthly to once a year (Lucio & McNeil, 1968). During classroom visits, supervisory committees monitored quality of instruction, assessed students' progress, monitored the curriculum being taught, cleanliness of the classroom, and judged the appropriateness of instructional methods employed by a teacher (Lucio & McNeil, 1968; Marks et al., 1985). Communities vested significant power in committees that supervised teachers. Committees were given the power to immediately dismiss teachers for what they judged to be ineffective instruction **(**Burke & Krey, 2005; Lucio & McNeil, 1968).

During this period (1600s-1700s), teacher supervision was known as the inspection stage – keeping a school clean, organized, and properly maintained were essential tasks (Barr, Burton, & Brueckner, 1947; Lucio & McNeil, 1968). Supervisors had minimal skill and understanding of effective instructional practices and the quality of feedback teachers received varied significantly among supervisors (Marks et al., 1985). According to Barr et al. (1947), a supervisor's most important role was to ensure a community's values and mores were being taught to students. Teachers' effectiveness was judged by students' ability to read scriptures and conform to their community's mores (Tracy, 1995).

As the political landscape of the country began to change during the early to mid 1800s, so did education. Communities began to expand in rural areas, and towns began to grow into large urban areas, which made it impossible to meet the educational needs of a community in a one-room school (Grieder & Romine, 1965). Marks et al. (1985) referred to this time in education as the professionalization phase, a time when schools and their teachers shifted away from community accountability and leadership and moved towards a status of professional educators. New professions in education were created in large urban areas that included superintendent, principal, and head teacher, which started a hierarchical system of education (Rury, 1991). As public schools began to grow larger and more complex, it became evident that clergy and community inspectors did not have the necessary backgrounds and knowledge bases to make informed decisions regarding teacher effectiveness (Lucio & McNeil, 1968). Tracy (1995) stated, "Rather than simply understanding the mores of the community, the supervisor now needed to have subject area knowledge and teaching skills" (p. 323).

As schooling systems evolved during the first half of the 1800s, so too did the role of a teacher. No longer was a teacher viewed as simply a servant to their community. Instead, the teaching profession now required a teacher to have expertise in their discipline, and supervisors needed abilities to provide effective feedback (Grieder & Romine, 1965). As clergy began to be phased out of public education, the need for an administrator who could handle complex roles increased (Marks et al., 1985). States employed superintendents of education and relied on county superintendents to supervise schools (Tracy, 1995). Eventually, in the late 1800s, local communities began forming their own structures to manage schools, which gave way to the role of principal (Barr et al., 1947). The role of

school principal evolved as a lead teacher that was selected to handle administrative responsibilities. This ultimately became the building principal responsible for supervising teachers (Barr et al., 1947; Marks et al., 1985).

As a result of a rise in industrialism, increasing educational needs of the mid 1800s gave rise to an increasing awareness regarding the importance of effective instructional skills (Tanner & Tanner, 1987). Teacher institutes started evolving to help teachers acquire necessary skills and better prepare teachers to effectively educate students (Tracy, 1995). The focus of teacher supervision changed from a system of compliance to a system that focused on teachers' pedagogical skills. Supervisors' reports still included cleanliness of a classroom and whether or not rules were being followed, but the most important trait being observed were teachers' pedagogical skills (Marks et al., 1985; Tanner & Tanner, 1987). According to Blumberg (1985), evidence dates back to the mid 1800s that references supervisors working with struggling teachers to improve their instructional methods. Although there was very little discussion or consensus about desired instructional skills at this time, it was understood that good teaching was essential in public schools and was considered a first step in developing effective teachers (Blumberg, 1985; Myers, Kifer, Merry, & Foley, 1938).

The Early 1900s—A Rise in Teacher Accountability

During the early 1900s, scientific management principles began to enter into the field of education (Wiles & Bondi, 1980). An education professor at the University of Chicago, Franklin Babbot, began to connect industrial management principles of control, accountability, and efficiency to the teacher evaluation process (Tanner & Tanner, 1987). Bobbitt, Parker, and Monahan (1913) believed that "scientific management finds the methods of procedure which are most efficient for actual service under actual conditions, and secures

their use on the part of the workers" (p. 51). The scientific method of evaluation centered on the idea that some teaching methods were more effective than others in relation to student achievement (Tanner & Tanner, 1987).

The scientific method of teacher evaluation continued to evolve into the early 1930s with an idea that teachers and students were similar to factory workers and raw materials (Wiles & Bondi, 1980). Cubberley (1922) compared schools to factories where students were molded and shaped to meet very specific demands in life. As disturbing as this view of education may sound, the idea helped solve problems education was facing at that time. The education system changed drastically as a result of the Industrial Revolution and support was needed to help organize growth, overcrowding, and increasing curricular demands that were taking place in schools (Cremin, 1976). The goal of using a scientific model to evaluate teachers was to help supervisors understand quality educational practices in order to assist teachers in improving their pedagogical skills (Ayer, 1954; Tanner & Tanner, 1987). At the time, education leaders believed the scientific model that had worked to solve business problems could do the same for education (Lucio & McNeil, 1968).

The scientific model of teacher evaluation still measured some of the same characteristics that were measured at the turn of the century (1900), including grooming, cleanliness of a classroom, integrity, and enthusiasm (Bobbit et al., 1913). In addition to personal traits of teachers, evaluations started focusing on instructional techniques, classroom management, and student assessment results (Bobbit et al., 1913; Marks et al., 1985). Under such a system, teachers stressed memorization of facts that students had to remember until an exam was complete, but the information never become part of their long-term knowledge base (Cubberley, 1922). Starting with the idea that students were "raw products," Cubberley

set out to find the best process to educate students, through collection of data and observations of teachers. Cubberley created a checklist supervisors could use to give teachers specific feedback regarding their instructional performance (Marzano et al., 2011).

The 1930s-1940s—Teachers Treated as Individuals

In the 1930s, the Hawthorne studies were published and education began to see a shift away from the scientific method of teacher evaluation to a more humanistic approach (Shinkfield & Stufflebeam, 1995). Instead of teachers and students being viewed as raw materials, they were viewed as individuals. The humanistic approach allowed education to transition from a production model focused on standards and outputs, to a field that focused on individuals within an organization (Ayer, 1954).

The Great Depression brought about an increase in awareness of societal problems, which cultivated a shift in the approach of administrators working with teachers in a supervisory role (Myers et al., 1938; Barr et al., 1947; Cohen & Manion, 1985). The Hawthorne studies examined motivation of employees in factories and concluded work effort, production, and morale improved when supervisors paid attention to employees as individuals and were supportive of their emotional needs (Marzano et al., 2011). It was believed that teachers would flourish in supportive environments in which supervisors acted as resources for teachers and focused on their personal satisfaction (Marks et al., 1985; Myers et al., 1938). Teachers began to be involved in the decision-making process regarding curriculum and instruction, and in creating a cooperative work environment with their supervisor, an environment that was substantially different than the previous authoritarian style of the early 1920s (Ayer, 1954; Barr et al., 1947).

With teacher supervision emphasizing social and emotional needs of teachers, it became a challenge for supervisors to guide and influence teachers' instructional practice. Supervisors feared that direct classroom supervision could place a strain on teacher-supervisor relationships they had worked so hard to develop (Tanner & Tanner, 1987). As a result, teacher supervision oftentimes resulted in a hands-off supervision model with teachers receiving very little guidance (Lucio & McNeil, 1968).

While this era of supervision focused on the social and emotional needs of a teacher, the role of a supervisor and their responsibilities continued to evolve (Marks et al., 1985). The list of responsibilities for supervisors during this time was extensive. According to Swearingen (1962), supervisors were responsible for teaching personnel, curriculum, the emotional quality of a classroom, resources and materials, school lunch service, attendance, public relations, and working cooperatively with groups and agencies in a community. Other job descriptions for supervisors during this time included numerous additional responsibilities, including teaching, faculty supervision, business, and social meetings (Marks et al., 1985). Thompson (1958) noted additional principal tasks such as: working with parents, placing students within a grade level in a school, completing paperwork, committee work, attending student conferences, recruiting new staff, modeling appropriate instructional practices, and acting as resource for a variety of stakeholders within the school community.

The increase in responsibilities for supervisors made a challenging task, but out of this era developed a consensus that teacher evaluation was critical to improving teachers' performance (Goldman, 1966). Whitehead's (1952) article, "Teachers Look at Supervision," surveyed teachers on six areas of teacher supervision and concluded advancements must be made in the area of supervisors' classroom observation practices.

> Improvements were still needed in following up the visitation with a conference, and in having the principal see the importance of remaining the entire period. It is not fair to teachers to visit them and not hold a conference following the visitation nor is it just to visit in a "piecemeal" fashion. (Whitehead, 1952, p. 102)

Whitehead (1952) believed that supervisors should pay more attention to the what he believed was the essence of education, effective teaching. With the importance of classroom observations justified, the foundation had been laid for teacher evaluation to move forward and evolve (Marks et al., 1985).

The 1950s—Focus on Instructional Skills

The focus of teacher evaluation began to shift once again in the 1950s and focus on teachers' competency and quality of their instruction. For the first time, evaluations were being conducted not only by supervisors, but also by peers, and through a teacher's own self-evaluation (Wiles, 1967). Evaluations of the time were primarily practiced as a formative evaluation tool to provide teachers with an opportunity to grow, or in some cases, to provide a means to transition an ineffective teacher out of the profession (Millman & Darling-Hammond, 1990; Shrinkfield & Stufflebeam, 1995; Wiles, 1967).

According to Bridges (1979), evaluation practices during the 1950s, 1960s, and 1970s were fundamentally weak and ineffective. Training for supervisors was inadequate, which led to ineffective evaluations of teachers that lacked necessary feedback and monitoring of teachers (Bridges, 1979). As a result, ineffective teachers were not provided necessary support and were eventually granted tenure.

According to Ryan and Kuhlen (1958), there were three different patterns during this time period that identified the quality and characteristics of teachers. In his study of 5000

elementary and secondary teachers, Ryan and Kuhlen identified a pattern of teacher characteristics they labeled X, Y, and Z. Pattern X was comprised of teachers that were understanding, sympathetic, and friendly versus those that were aloof, restricted, and egocentric. Pattern Y teachers included those who were businesslike, responsible, and systematic versus those who were haphazard, unplanned, and slipshod. The final pattern, Z, was comprised of teachers that were stimulating and creative versus those that were boring and believed in rote learning (Ryan & Kuhlen, 1958).

In the late 1960s and early 1970s, education professionals were introduced to a clinical supervision model (Wiles & Bondi, 1980). Developed at Harvard by professors Morris Cogan and Robert Anderson along with their graduate students, the clinical supervision model caught on quickly in education (Cogan, 1972; Shinkfield & Stufflebeam, 1995). The model was developed as a systematic approach to work with student teachers in the Master of Art teaching program at Harvard (Cogan, 1972). Comprised of a blended scientific and objective system of classroom observation that focused on collegial relationships, planning, flexibility, and inquiry based emphasis on student learning, the clinical supervision model began to formally develop (Cogan, 1972). In the second edition of *Clinical Supervision: Special Methods of the Supervision of Teachers*, Goldhammer, Anderson, and Krajewski (1980) outlined five phases of teacher evaluation including: (a) pre-observation conference, a teacher and supervisor discuss elements of a lesson and agree on what is to be observed; (b) classroom observation, supervisor observes what has been agreed upon in the pre-conference; (c) analysis, supervisor organizes data from the observation to help teacher analyze their own teaching; (d) supervision conference, supervisor and teacher discuss data from classroom observation, and teacher is asked to

reflect on their teaching; and (e) analysis of the analysis, supervisor examines the process and reflects on their own practice.

The clinical supervision process was developed as a method to observe the holistic approach to teaching, which included interactions between teachers and students (Shinkfield & Stufflebeam, 1995). According to Cogan (1972), supervisors should have no preconceived notion of what effective teaching should entail. The clinical supervision model did not describe effective instructional practices that should be used and evaluated in a classroom. Instead, the model and its five phases were designed to discover effective practices through collegial and inquiry-based discussion between teachers and supervisors (Goldhammer et al., 1980). Cogan (1972) went as far as cautioning supervisors that their own personal experiences may interfere with their ability to provide teachers with effective feedback.

At the heart of the clinical supervision model was a collegial relationship between a supervisor and a teacher (Tanner & Tanner, 1987). Through a nonjudgmental relationship based on mutual trust, student learning was expected to improve through the inquiry-based evaluation process (Cogan, 1972). However, the rich dialogue that Goldhammer et al. (1980) and Cogan envisioned never fully transpired, and the model began to fall out of favor with educators in the early 1980s (Marzano et al., 2011).

The 1980s—An Increase in Accountability

In 1983, the National Commission on Excellence in Education (NCEE) published a report called *A Nation at Risk.* The report called for numerous changes in the nation's education system because T. H. Bell, then Secretary of Education was concerned about a "widespread public perception that something is seriously remiss in our educational system" (T. H. Bell as cited in National Commission on Excellence in Education [NCEE], 1983, p. 1).

Recommended changes to the education system included longer schools days, increasing the rigor of all classes, and improvement in teacher quality (Danielson, 2001). *A Nation at Risk* stated, "Persons preparing to teach should be required to meet high educational standards, to demonstrate an aptitude for teaching, and to demonstrate competence in an academic discipline" (NCEE, 1983, p. 30).

Reform and accountability in education quickly became an agenda item at local, state, and national levels. School districts around the country worked to design and implement teacher evaluation models that were educationally meaningful holding teachers to high standards (Ellett & Teddlie, 2003; Shinkfield & Stufflebeam, 1995). The NCEE's (1983) recommendations at the time stated that "salary, promotion, tenure, and retention decisions should be tied to an effective evaluation system that includes peer review so that superior teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated" (NCEE, 1983, p. 30). The idea that teacher evaluation models should be growth orientated was a new concept for the decade; until this time, growth was not part of the discussion surrounding teacher evaluation (Darling-Hammond, 1998).

With an increased emphasis on teacher evaluation being tied to merit pay, job retention, and master-teacher status, Wise, Darling-Hammond, McLaughlin, and Bernstein (1985) called for an evaluation system that was "standardized, objective, and externally defensible" (p. v). Wise et al. studied 32 school districts' teacher evaluation models and found a varied approach among them. Wise et al. concluded teacher evaluation models they studied lacked clear purpose and goals, principals were ineffective evaluators, and a majority of teachers were apathetic and resistant to change (Wise et al., 1985). Positive results of

evaluations studied included improved communication between teachers and administrators, improved instructional skills, and reduced feelings of isolation (Wise et al., 1985).

As a result of their study, Wise et al. (1985) developed a number of characteristics needed for effective teacher evaluation to take place. First, an evaluation should align with a school district's mission, vision, and goals. Second, there must be commitment by a school district to make teacher evaluation a top priority by committing necessary time, energy, and resources to evaluations. Third, principals conducting evaluations must be properly trained. Finally, researchers recommend evaluations be conducted with "master teachers" to assist principals with content expertise that principals may lack.

Another concern with teacher evaluations that emerged during the 1980s was the lack of training principals received on how to effectively conduct evaluations (Stiggins & Duke, 1988). In addition to training, Stiggins and Duke also recommended time for principals and teachers to have conversations about effective instructional practices; encouraged teachers to become more involved in their evaluation process; and included data from a variety of source including students and peers in order to develop student outcomes.

While the goal of an effective teacher evaluation system was to improve teachers' performance in their classroom and ultimately student learning, Bridges (1979) called for a model that would also support dismissing a teacher if their performance did not improve. The National Commission on Teaching and America's Future (1996) also advocated removing ineffective teachers from a classroom. The Commission's approach was to develop a peer review model that worked closely with a principal to provide necessary support to struggling teachers. It was believed that over time a combined model would better support teachers in

need of improvement, and if necessary, recommend the dismissal of a teacher (National Commission on Teaching and America's Future, 1996).

The National Commission on Teaching and America's Future (1996) based their recommendations on changes that occurred in Rochester, New York; Toledo, Columbus, and Cincinnati, Ohio; and Seattle, Washington. In each of these cities, "more teachers have been given help and have made major improvements in their teaching, and more teachers have been dismissed than ever occurred under the old systems of evaluation and administrative review" (National Commission on Teaching and America's Future, 1996, p. 99). School districts in these cities established peer review panels consisting of teachers and administrators that were responsible for providing support to low performing teachers, creating an environment of professional accountability while improving instruction (National Commission on Teaching and America's Future, 1996).

The 1980s saw a significant shift in the way principals conducted teacher evaluations. Reports stating schools and teachers were failing students set off the accountability movement, the public and lawmakers wanted to see change in schools (NCEE, 1983). With mounting evidence that teachers were the most important variable in a student's ability to learn, development of teacher evaluation models that could identify effective teachers, help improve struggling teachers, and dismiss low performing teachers became crucial (Ellett & Teddlie, 2003; NCEE, 1983). Holding teachers to high standards was a consistent message in the literature written during the decade and paved the way for a focus on standards.

The 1990s—Start of the Teacher Evaluation Framework

During the early 1990s and into the 21st Century, teacher evaluation continued to be at the forefront of the school reform movement with the focus on accountability, school

improvement, and professional development (Ellett & Teddlie, 2003). A variety of new developments in teacher evaluation methods, including a shift in classroom-based evaluations from teaching to student learning and the development of the *National Board of Professional Teaching Standards*, highlight two of the changes during the 1990s (Danielson & McGreal, 2000; Ellett & Teddlie, 2003).

Another significant development in teacher evaluation during the 1990s was the development of Charlotte Danielson's (1996) *Enhancing Professional Practice: A Framework for Teaching*. According to Marzano et al. (2011), the Danielson model is the standard by which teacher supervision and evaluation models are to be judged. The model is broken into four domains: (a) Planning and Preparation, (b) the Classroom Environment, (c) Instruction, and (d) Professional Responsibilities (Danielson, 1996). According to Danielson, the framework was designed to accomplish three things. First, honor educators by recognizing the complexity of teaching. Second, create a common language to be used by educators in professional conversations. Third, create a structure for teachers to self-assess and reflect on their professional practice.

During the early part of the 21st century, the focus of teacher evaluation began to shift towards standards-based and value-added models (Papay, 2012). The No Child Left Behind (NCLB) Act of 2001 called for "scientifically based research" and "evidence-based practices." Legislation at the time asked for a quantitative approach that could measure the cause-and-effect relationship between an educational condition and outcomes produced (Feuer, Towne, & Shavelson, 2002). Through the use of quantitative data, policy makers hoped that findings would be able to be generalized to help inform the decision-making process regarding education policy (Lin, Wang, Klecka, Odell, & Spalding 2010).

Value-Added Teacher Evaluation Models

With an increase in standardized testing as a result of the NCLB Act of 2001 and the availability of student testing data, K-12 educators began to realize the potential of value-added teacher evaluation models (Papay, 2012). Value-added models were designed to measure teachers' effectiveness on students' achievement (Donaldson & Donaldson, 2012). By using past testing results, students' background information, and characteristics of peers at specific schools, principals would be able to predict students' performances and compare it to actual test results as a way to evaluate teachers' instructional performance (Papay, 2012). Linking a teacher's evaluation score to their students' test scores aligned with reasoning of politicians, economists, and education theorists that believed teachers' employment and compensation should be tied to student performance (Hastings, Kane, & Staiger, 2006).

Value added teacher evaluations have received support among policy makers for a variety of reasons. First, evaluators are using testing data to measure student growth that was implemented as a result of federal legislation (Papay, 2012). An evaluation is based on external assessment with the potential to eliminate the bias of the evaluator (Donaldson, 2009; Papay, 2012). Most importantly, a value-added model focuses on teachers and their ability to increase student achievement, and eliminates the status quo of all teachers being scored "satisfactory" (Donaldson, 2009).

Despite the quantitative data that value added teacher evaluation models produce, there are compelling arguments that say a value added model does not accurately capture teachers' effects on student learning (Donaldson, 2009). Even when student assessments are reliable, valid, and standardized, quantifying teachers' impact on student test scores poses serious problems (Callister Everson, Feinauer, & Sudweeks, 2013; Hanushek & Rivkin,

2010). At best, standardized test scores reflect how teachers' affect a limited range of students' abilities in content areas that the assessment attempts to measure (Callister Everson et al., 2013). What assessments are not able to measure are outside influences beyond a teacher's control, including students' prior education experiences, bias in distribution of students in classrooms, home environment, learning disabilities, poverty, homelessness, and hunger (Callister Everson et al., 2013; Garrett, 2011; Greene, 2002; Hanushek & Rivkin, 2010; Hill, Kapitula & Umland, 2011).

In addition to what assessments are able to accurately measure being problematic, Harris (2010) pointed out systematic and random errors that can occur during value-added accountability. In his research studying low-performing schools, Harris found school-based performance measurements were systematically biased against schools that were serving economically disadvantaged students. Random errors in value-added measures point to instability over time. According to Harrris, value-added measures for each individual teacher change over time, teachers that are high performing one year can get worse the following year. Taken together, many questions and critiques surrounding value-added scores for teacher evaluation suggest a value-added model does not accurately reflect the competence and effectiveness of teachers (Hill et al., 2011).

In a study conducted by McCaffrey, Lockwood, Koretz, Louis, and Hamilton (2004), it was concluded that value-added measures are more stable "over time," and McCaffrey et al. recommended using multiple years' data when evaluating teachers' testing data. Hill et al. (2011) recommended that value-added scores are not sufficient to stand alone and identify teachers for tenure, promotions, reward, or termination. Furthermore, in their study of math teachers, Goldhaber, Goldschmidt, and Tseng (2013) concluded value-added models pose a

risk of misclassifying teachers. However, when comparing value-added models against less rigorous models that were in place at the time of this study, value-added models allow for comparison between teachers (Golhaber et al., 2013). While there is agreement among researchers that value-added models alone may not accurately portray teachers' effectiveness, there are others who believe it is far better than available alternatives (Greene, 2002; Harris 2010).

Standards-Based Teacher Evaluation Models

Prior to standards-based teacher evaluation models, evaluation of teachers was a topic of significant debate (Heneman & Milanowski, 2003). Problem areas identified in the literature include a lack of shared values on what effective teaching looks like (Danielson & McGreal, 2000), validity concerns (Medley & Coker, 1987), lack of effective feedback, an emphasis on following rules and procedures instead of improving instructional performance, and apathy towards the evaluation process among both administrators and teachers (Johnson, 1990). Milanowski and Heneman (2001) claimed old evaluation models were cumbersome, were outdated, placed little emphasis on instruction, and were primarily used to ensure a minimal level of acceptable performance.

Problems with common language, validity, and poor of feedback made teacher evaluation, for the most part, an ineffective process that lacked the ability to improve teachers' performances. As a result, standards-based teacher evaluation models began to emerge and were based on detailed teaching standards designed to capture quality instruction (Heneman & Milanowski, 2003). Standards-based teacher evaluations are based on the following characteristics:

a. A comprehensive description (competency model) of teacher performance reflecting the current consensus of good teaching.

b. Explicit standards and multiple levels of performance (rather than simply pass/fail), defined by detailed behavioral rating scales (usually called rubrics), that provide guidance: (a) to evaluators on how to rate, and (b) to teachers on what behaviors are expected of high performers.

c. More frequent observations of actual classroom practice and use of multiple lines of evidence, such as lesson plans and samples of student work, to provide a richer picture of teacher performance.

d. Trained evaluators (Danielson & McGreal, 2000, p. 20).

Standards-based teacher evaluation models are based on the notion that there is a common set of effective teaching behaviors that have been established through empirical and theoretical research (Danielson, 1996; Danielson & McGreal, 2000). Through work of Danielson (1996) and Ellett (1997), standards-based teacher evaluations began to be used more frequently in schools. The new evaluation models reduced subjectivity and provided administrators with a valid method to judge teachers' effectiveness based on a set of common criteria (Kimball & Milanowski, 2009). Standards-based models assessed teachers' instructional skills using a comprehensive set of standards and rubrics that were designed to improve instruction and strengthen accountability in education (Borman & Kimball, 2005). Standards-based evaluation models take into consideration a variety of evidences regarding a teacher's instructional practice (Kimball & Milanowski, 2009). According to Henemen, Milanowski, Kimball, and Odden (2006), standards-based systems provide evaluators with standards that allow them to judge a teacher's performance.

Standards-based teacher evaluation models were developed based on research that there was a connection between quality teaching practices and student achievement (Feeney, 2007). Standards-based teacher evaluation models developed a foundation and a common language between teachers and evaluators of what effective instruction should look like (Henemen et al., 2006). A common language provided teachers with an opportunity to measure their performance by examining standards and reflecting on their level of proficiency regarding their performance (O'Pry & Schumacher, 2012). According to Aseltine, Faryniarz, and Rigazio-DiGilio (2006), when teachers have opportunities to receive meaningful, constructive feedback that promotes reflection, instruction will ultimately improve. An administrators' ability to work collaboratively with teachers is essential to developing a growth mindset (Danielson, 2007; Feeney, 2007; Glickman, 2002). When using a standards-based evaluation model, improvement only happens when teachers are given an opportunity to become self-directed, self-monitoring, and self-managing (Glickman, 2002).

While a standards-based evaluation model was a significant improvement over previous models, there were still problems with the model's implementation and composition (Hazi & Rucinski, 2009; Milanowski & Heneman, 2001). Peterson (2006) argued standards-based models constricted teaching and reduced its complexity to a simplistic level. In addition, Peterson (2006) noted there was no empirical evidence that connected standards-based models to having a positive effect on instructional improvement. Schumacher (2011) found that standards-based models can be time intensive and complex for both teachers and administrators, both are reasons that can lead to a model's downfall.

Trends in Teacher Evaluations

At the time of this study, numerous models for teacher evaluations were available and being implemented and used in school districts across the United States. Some of the models were developed by individual school districts, others by state agencies, and some were developed by individual researchers or research organizations. Table 2 displays five nationally recognized teacher evaluation models.

Table 2. Comparison of Teacher Evaluation Models.

Standards / **Domains**	**InTASC** 4 Domains 10 Standards 174 Indicators	**Danielson** 4 Domains 22 Components 76 Elements	**Marzano** 4 Domains 60 Elements	**Marshall** 6 Domains 60 Elements	**McREL** 5 Domains 25 Items
1	Learner and Learning	Planning and Preparation	Classroom Strategies and Behaviors	Planning and Preparation for Learning	Teachers Demonstrate Leadership
2	Content Knowledge	The Classroom Environment	Planning and Preparing	Classroom Management	Teachers Establish a Respectful Environment for a Diverse Population of Students
3	Instructional Practice	Instruction	Reflecting on Teaching	Delivery of Instruction	Teachers Know Their Content
4	Professional Responsibility	Professional Responsibility	Collegiality and Professionalism	Monitoring Assessments and Follow-up	Teachers Facilitate Learning
5				Family and Community Outreach	Teachers Reflect on Their Practice
6				Professional Responsibilities	

The researcher selected three models for comparison purposes; two models were developed by nationally recognized leaders in the field of education and one was developed by a respected research organization. Selection of models was based on the researcher's review of literature regarding most common models being used throughout the country. The first model was based on the work of Marzano (2007); the second model was developed by Danielson (1996); and the third model was developed by the McREL group consisting of Dean, Hubbell, Pitler, and Stone (2012).

The Marzano Teacher Evaluation Framework

In an effort to develop a teacher evaluation model that was complex and robust enough to handle the numerous demands of the teaching profession, Robert Marzano (2007), created the Marzano Teacher Evaluation Framework (MTEF) model based on his book, *The Art and Science of Teaching*. The model was based on behaviors and strategies that had been subjected to over 300 experiments and control studies taking place across 14 school districts, involving 38 schools, 300 teachers, and 14,000 students (Marzano, 2007). The MTEF was not developed as a one size fits all model. Instead, it was designed for teachers to use and make decisions regarding their instructional practices and thus improve student learning. Based on his research of effective teaching practices, Marzano then developed an observation protocol to be used for supervising and evaluating educators that was aligned with *the Art and Science of Teaching* (Marzano et al., 2011). Recognizing that teaching was both an art and a science, the MTEF aligned with previous work of Reagan, Case, and Brubacher (2000), who stated, "Teaching entails elements of both artistic sensitivity and technical skill and that good teaching is impossible with both elements" (p. 18). Throughout Marzano's Evaluation Framework, he incorporated both scientific and artistic qualities of teaching.

The MTEF was designed to provide teachers with specific and timely feedback using three different types of evaluations including walkthroughs, informal observations, and formal observations. Walkthrough evaluations last 5-10 minutes, are unannounced, and provide administrators with a pattern of the instructional practices being used in a school; they help lower teacher apprehension to supervision making formal observations more productive. Informal observations can be announced or unannounced and last between 10-20 minutes, allowing an administrator enough time to view a segment of a lesson. The last type of evaluation is a formal evaluation that consists of a pre-conference, a classroom visit lasting approximately 50 minutes, and a post-conference (Marzano et al., 2011). Of the data collected from the three types of observation, only the informal and formal observations count towards the teacher's final evaluation score, while walkthroughs should are used as anecdotal feedback for teachers (Marzano et al., 2011). According to Marzano (2010/2011), teachers need "a robust model of teaching that is used as the basis of feedback for teachers that does not simply assume all researched-based instructional strategies should be present in every lesson" (p. 25). Providing teachers with quality feedback ensures they are able to make informed decisions regarding their instructional practices. Marzano (2010/2011) was careful to point out that not all feedback will have positive effects on teachers' performances. If evaluators are not trained properly, their feedback may lack specificity, evaluating situations out of context, lack timeliness, and contain poor communication (Ericsson, Charness, Hoffman, & Feltovich, 2006).

Marzano's teacher evaluation model is based on three distinct phases and segments; effective instructional strategies, effective classroom management, and effective curriculum design (Marzano, 2007). The MTEF model was designed to recognize the complexity of

teaching and honor work of classroom teachers. The MTEF is not a simple checklist. Instead, it is a layered approach to evaluating teaching and learning based on educational research that has a highest probability for student success (Marzano, 2007). In 2008, Brown made the following statements regarding the MTEF:

1. Teachers have the ability to make an enormous difference in the academic success of all students.
2. There are three necessary components of effective instruction: (a) consistent use of research based instructional strategies, (b) creating a community of learners through the use of effective classroom management strategies, and (c) effective curriculum design and development.
3. Students should know what they are learning and the reason for learning it.
4. Students should know their progress on class learning goals and be able to assess their proficiency.
5. Effective teaching and learning requires that students move toward conceptual understanding and transferable application of course material.
6. Effective classrooms are collaborative environments that create a community of learners (Brown, 2008).

In 2009, the MTEF was released as a tool to help school leaders evaluate teachers and provide them with feedback. The model was developed to provide specific feedback based on the book, *The Art and Science of Teaching.* The MTEF consists of four domains:

Domain 1: Classroom Strategies and Behaviors

Domain 2: Planning and Preparing

Domain 3: Reflecting on Teaching

Domain 4: Collegiality and Professionalism (Marzano et al., 2011, p. 4)

Within each of these domains are subcategories that consist of 60 separate elements. Domain 1 consists of 41 elements, divided into nine specific design questions. Domain 2 consists of eight elements and one design question. Domain 3 consists of five elements and Domain 4 consists of six elements.

The MTEF is divided into segments that categorize teaching lessons in the following way:

Domain 1: Classroom Strategies and Behaviors

- Routine Segment: Teaching strategies that an evaluator would expect to see during all lessons. Routines include communicating a lesson's learning goal, reviewing rules and procedures when appropriate, tracking students' progress, and celebrating their success.
- Content Specific Segment: This segment is the actual teaching of a lesson and is divided into three categories:
    - Helping students interact with new knowledge: If a teachers' lesson involves new content, evaluators would expect to observe identification of critical information, organizing students in a way that is conducive to interacting with the new information, previewing of new content, chunking information, and time to interact, process, elaborate, and reflect on new content.
    - Helping students practice and deepen new knowledge: If required, a teacher would review content using homework assignments, activities that

include similarities and differences, and practice skills, strategies, and processes that reinforce content.

    - Helping students generate and test hypotheses: If a lesson required students to generate and test a hypothesis, students would be organized in a way that would allow them to complete complex tasks while a teacher guides students and provides necessary support.

- Enacted on the Spot Segment: Items included in this section would be seen on an as needed basis. Teachers using this segment would notice when students are not engaged, or when a rule is not being followed. Depending on how a lesson was progressing, teachers may change the pace of the lesson, increase intensity and enthusiasm, use physical movement, provide opportunities for students to talk about themselves, and present unusual or intriguing information about a topic. Establishing and maintaining appropriate relationships, using verbal and nonverbal cues to show interest in students, being objective and in control, and communicating high standards for all students are also characteristics that teachers would display when necessary.

Domain 2: Planning and Preparing

- The planning and preparing domain emphasizes effective lesson and unit design. Effective planning consists of proper scaffolding of information within each lesson and each lesson that comprises a unit along with attention to established content standards. Other areas that need to be considered for proper planning include the use of traditional resources and technology, how the needs

of English Language Learners will be a addressed (if applicable), special education students, and students that lack support for schooling.

Domain 3: Reflecting on Teaching

- The reflecting on teaching domains require teachers to evaluate areas of pedagogical strength and areas for growth, effectiveness of lessons and units, and pedagogical strategies and behaviors. Teachers then develop a written growth and development plan and monitor their progress.

Domain 4: Collegiality and Professionalism

- The collegiality and professionalism domain emphasizes the promotion of a positive environment that encourages interaction among colleagues, students, and parents. Positive environments promote a culture of sharing where teachers seek out assistance from colleagues in areas of interest or need and fellow teachers readily share their expertise and mentor their colleagues. In addition to a collaborative environment at schools, teachers also participate in school and district initiatives along with following rules and procedures.

The goal of the MTEF is to develop expert teachers by identifying critical teaching behaviors and providing high quality feedback that, if used, could lead to improved student achievement. Researchers published in Ericsson et al. (2006) identified quality feedback as the critical component for development and improvement of experts; without it, efficient learning is nearly impossible.

The Danielson Framework for Teaching

The Danielson Framework for Teaching is a research-based set of instructional components that are grounded in a constructivist view of learning and teaching and are aligned with 10 principles of the Interstate New Teacher Assessment and Support Consortium (Danielson, 2010). The model was first published in 1996. During the development of the Praxis III program, it became evident there was a need to create a comprehensive teaching framework that could provide teachers with a common language and format that would allow them to reflect and have discussions with colleagues about their teaching (Danielson, 2007).

The Danielson teacher evaluation model was designed to address areas of deficiency that existed in previous teacher evaluation models, including:

1. Outdated evaluative criteria (posting objectives on a board instead of teaching for understanding, scored using a checklist),
2. Ineffective evaluative comments ("outstanding," "meets criteria," and "needs improvement," instead of comments on what teachers can do to improve),
3. Lack of shared values regarding teaching (no common language),
4. Lack of precision in evaluating performance (simple rating scales that consist of "1 to 4" or "needs improvement" or "satisfactory"),
5. One-way communication (evaluators document a teacher's performance and give feedback with no teacher input),
6. No differentiation between experienced and novice teachers, and
7. Consistency problems among evaluators (Danielson, 2007).

The Danielson model was designed to ensure validity and was meant to be applicable to all teachers in various educational settings (Danielson, 2007). According to Danielson (2007), the model is:

1. "Comprehensive": The framework attempts to "describe *all* of teaching, in all its complexity. It is comprehensive, referring not only to what occurs in the classroom but also to what happens behind the scenes and beyond the classroom walls." (p. 19)
2. "Grounded in Research": According to Danielson, the framework was designed based on research that "seeks to identify principles of effective practice and classroom organization" (p. 20).
3. "Public": The framework is known by all teachers ahead of time in order to prevent a "gotcha" mentality (p. 21). Teachers are encouraged to discuss the model with colleagues and develop a common language of what effective instruction looks like.
4. "Generic": Because no one teacher, student, or classroom is alike, Danielson (2007) designed a model that attempts to identify "powerful commonalities" instead of behaviors, because "behaviors themselves depend on the context" (p. 22).
5. "Coherent in Structure": The framework takes into consideration the complexity of teaching by dividing into four domains:

   Domain 1 – "Planning and Preparation" (p. 22),

   Domain 2 – "The Classroom Environment" (p. 22),

Domain 3 – “Instruction” (p. 23), and

Domain 4 – “Professional Responsibilities” (pp. 22-23).

Each of the domains consists of five or six smaller components for a total of 22 components in all four domains. Within each component there are individual elements (76 total) that describe each of the components in greater detail. Rubrics ranging from unsatisfactory to distinguished are used to score teachers on each of the components.

According to Danielson (1996; 2007), her framework was designed independent of any specific teaching methodology because no single teaching style can work in every situation. Instead, individual teachers need to have the ability to select an appropriate strategy for a given situation they feel will give them a desired outcome. However, Danielson (2007) pointed out that, “The framework for teaching does not endorse any particular teaching style for all teachers; it does, however, enable educators to engage in conversations about the appropriateness of choices” (p. 25).

The Danielson framework set the bar for all teacher evaluation models. Since the Danielson framework was first introduced in 1996, it has been adopted by educators around the world (Danielson, 2007) and is considered the standard by which all models are measured. According to Marzano et al. (2011), the Danielson model must be used as a reference point for any new teacher evaluation proposals. “The level of specificity supplied in the Danielson model provided the foundation for the most detailed and comprehensive approach to evaluation to that time” (Marzano et al., 2011, p. 24). Research by Ericsson, Krampe, and Tesch-Römer (1993) and others in Ericsson et al. (2006) emphasized that specific feedback focused on behaviors and skills was essential in promoting and developing teachers’ expertise.

The McREL Teacher Evaluation System

The McREL Teacher Evaluation system was designed using 21st Century Learning and research-based instructional strategies that have been proven to increase student achievement (Williams, 2009; Dean et al., 2012). The McREL evaluation instrument was designed to promote leadership, quality teaching, and student learning. According to Williams (2009), the McREL evaluation system was designed to encourage professional growth, be adaptive to each individual teacher being evaluated, and serve as a reflective process for the establishment of professional goals and for identifying areas of professional improvement.

The purpose of the McREL Teacher Evaluation model is to measure a teacher's performance compared to a set of professional teaching standards that will help guide the teacher in developing a professional growth plan (Dean et al., 2012). The McREL model is a three-prong cooperative process between a principal and a teacher that includes self-assessment, presentation of artifacts and discussion, and classroom instructional feedback (Williams, 2009). Through discussions, evaluations, and self-reflection, the McREL model is designed to accomplish the following for the teacher, school, and district:

- Measure performance of teachers,
- Guide teachers in their reflection and improvement,
- Provide a framework for instructional improvement,
- Help schools and districts achieve their goals and objectives by monitoring, evaluating and supporting teachers,
- Provide a guide for teachers' professional development,

- Serve as a resource in developing mentoring and coaching programs,
- Assist in the implementation of curriculum,
- Provide a bridge for working with higher education as they develop and train new teachers (Williams, 2009).

The McREL model is based on the belief that an effective evaluation system has the ability to improve a teacher's and a school's performance through a cycle of continuous improvement and feedback loops that move along a growth continuum (Dean et al., 2012).

Included in the McREL model are five standards with a total of 25 elements that are based heavily on 21st century education practices. Teachers are scored as either developing, proficient, accomplished, or distinguished.

Standard I: Teachers Demonstrate Leadership

- Included in Standard I are five elements. Element I includes teachers being leaders in their classrooms by taking responsibility for the progress of their students, ensuring that they are ready for the next grade level, and making progress towards being competitive in a global society.
- Element II encourages teachers to work collaboratively with colleagues and school leaders to improve the learning experience for all stakeholders.
- Element III encourages teachers to be leaders in the teaching profession by contributing to a positive culture and being active in their school and district decision-making process.

- Element IV asks teachers to be advocates for their schools and students by working to effect change in policies and practices that will improve student learning.
- Lastly, Element V promotes high ethical standards for teachers in the areas of honesty, integrity, fair treatment, and respect for colleagues and students.

Standard II: Teachers Establish a Respectful Environment for a Diverse Population of Students

- Included in Standard II are five elements. Element I asks teachers to create an environment where each child has a positive, supportive, caring adult that they can trust.
- Element II encourages teachers to embrace diversity in their school, community, and world by teaching students to be open to different cultures by selecting materials that counteract stereotypes.
- Element III asks teachers to treat students as individuals and maintain high expectations for all students regardless of their background by building positive, appropriate relationships.
- Element IV requires teachers to be flexible and adapt their teaching appropriately for students with special needs.
- Lastly, Element V expects teachers to work in a collaborative manner with families and other significant adults that are involved in their students' lives.

Standard III: Teachers Know the Content They Teach

- Standard III includes four elements. Element I requires that teachers align instruction with their state standards and their district's curriculum.
- Element II expects teachers to know their content and have the ability to bring depth to understanding, and create curiosity in their students to enrich the learning experience.
- Element III looks for teachers to recognize interconnectedness of content, either vertically and/or horizontally in order to deepen the understanding for students.
- Lastly, Element IV asks teachers to make learning relevant for their students by deliberately including 21st century skills into their curriculum, skills such as leadership, accountability, adaptability, collaboration, social responsibility, and self-direction.

Standard IV: Teachers Facilitate Learning for Their Students

- Standard IV contains eight elements. Element I expects teachers to know the ways that learning takes place in their classroom, and understand the levels of physical, intellectual, and social emotional development of their students.
- Element II stresses the importance of teachers planning appropriate instruction to meet the needs of their students.
- Element III monitors the extent to which teachers vary their instructional strategies to meet the needs of diverse learners in their classroom.
- Element IV encourages the use of technology to improve students' learning.

- Element V encourages teachers to assist students development of problem-solving and critical thinking skills.
- Element VI highlights the importance of developing leadership qualities in students through collaboration and cooperation.
- Element VII stresses the importance of communicating clearly with students so that the teacher is easily understood.
- Element VIII requires teachers to use a variety of assessments, both formative and summative, to determine what students have learned.

Standard V: Teachers Reflect on Their Practice

- Standard V includes three elements. Element I requires that teachers analyze student learning and think critically about what they can do to improve school and classroom achievement.
- Element II expects teachers to link their professional growth with their professional goals.
- Element III reminds teachers that change is constant, and they must continuously strive for ways to improve their teaching and learning.

The McREL model attempts to incorporate a broad base of skills and knowledge into teachers' professional teaching standards (Williams, 2009). Teachers' instructional practices must reflect appropriately a knowledge of the needs of students so students are provided with the necessary skills to be successful as they enter a global society (Burkhardt et al., 2003). According to Williams (2009), students will require a K-12 education that supports a deeper understanding of content knowledge, competencies, skills, and outcomes. While there is

agreement that students need a different set of skills to be successful in the 21st century compared to previous times, there is no set standard universally accepted that outlines what that set of skills might be.

Concerns Regarding Standards-Based Teacher Evaluation Models

Prior to the Race to the Top initiative, there were numerous strategies employed by federal, state and local governing agencies to improve public education including performance pay, standardized testing, alternative certification, and licensing exams; neglected was the potentially powerful tool of teacher evaluations (Toch, 2008). As discussed in previous sections, the Danielson, Marzano, and McREL teacher evaluation models all claimed to align with standards that improve teachers' instructional skills and in return increase student achievement. However, according to Strong, Gargani, and Hacifazliouglu (2011), and Cohen (2015), we still do not have a quality evaluation tool that identifies teachers' instructional quality and student achievement.

Hanushek (1992) found there to be a difference of one academic year of achievement on standardized tests between students that had a "good" teacher opposed to those that had a "bad" teacher. While the gains in student achievement were clear, the characteristics of good teachers were not as clear. Hanushek was not able to capture the elusive characteristic that set good teachers apart from bad teachers in his study. In a more recent study based on classroom observations, Strong et al. (2011) discovered that "There is not much evidence to suggest a strong relationship between observation-based teacher evaluations ratings and student achievement outcomes" (p. 368).

Another concern in the literature surrounding standards-based teacher evaluation models centers on inter-rater reliability (Derrington & Campbell, 2015; Odden, 2004; Strong

et al., 2011). According to Strong et al., administrators in their study were able to identify highly effective and ineffective teachers, but struggled to correctly identify teachers in the middle. In addition, Kimball and Milanowski (2009) identified three factors that can potentially affect reliability; they are known as *will*, *skill*, and *evaluation context*. *Will* refers to an administrator's motivation in the context of performing an evaluation (Kimball & Milanowski, 2009). In the case of *will*, the relationship between an administrator and a teacher has the potential to positively or negatively impact a teacher's evaluation. The second factor that Kimball and Milanowski (2009) identified is *skill*. *Skill* is the ability of an administrator to make a correct judgment regarding a teacher's performance. The third and final factor identified by Kimball and Milanowksi is the *evaluator context*. *Evaluator context* refers to a school environment where an evaluator is observing. Environments identified as high-performing or low-performing tend to receive evaluation scores that are aligned with the performance level of a school, creating inflated or deflated scores (Kimball & Milanowski, 2009). According Kimball and Milanowski each of these factors are related to a subconscious process and have the ability to influence inter-rater reliability.

The size of standards-based teacher evaluation models can also be a concern. For the purpose of this study, a comparison of standards-based teacher evaluation models showed number of elements in standards-based evaluation models range between 25 and 76 elements that evaluators need to understand and be able to identify for evaluation purposes. According the Hill, Charalambous, and Kraft (2012), evaluation models that are large and complex have the ability to overwhelm administrators working memory and interfere with their ability to accurately score a teacher's lesson. As previously outlined in *Table 2. Comparison of Teacher Evaluation Models*, each model has large numbers of indicators that administrators

must navigate when conducting evaluations. For example, Marzano's model has 4 domains with 60 elements (Marzano, 2007). The Danielson model has 4 domains that include 22 components and 76 elements (Danielson, 2007).

Finally, the level of commitment required to properly implement and use a standards-based teacher evaluation model is significant. Cash, Hamre, Pianta, and Myers (2012) indicated standards-based models require intensive resources including time, training, and monetary support, which in today's educational climate can be a challenge. At a time when school district budgets are consistently being cut, and there are increasing demands on an administrator's time, using standards-based teacher evaluation models with fidelity has the potential to be problematic (Cash et al., 2012; Cohen & Goldhaber, 2016; Schumacher, 2011; Strong et al., 2011).

Implementation of a New Teacher Evaluation Model

Unfortunately, in the world of education, inertia is a powerful force. Stigler and Hiebert (2009) pointed out that while school reform initiatives come and go, "The substantive nature of what happens in the classroom stays pretty much the same" (p. 32). The following section examines five constructs that frame this study including *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with the ease of use of a new teacher evaluation model. According to the literature, all five constructs working concurrently are essential pieces to ensuring successful implementation of a new teacher evaluation model. Figure 2 displays a diagram of the research study, constructs, and survey questions aligned with each construct.

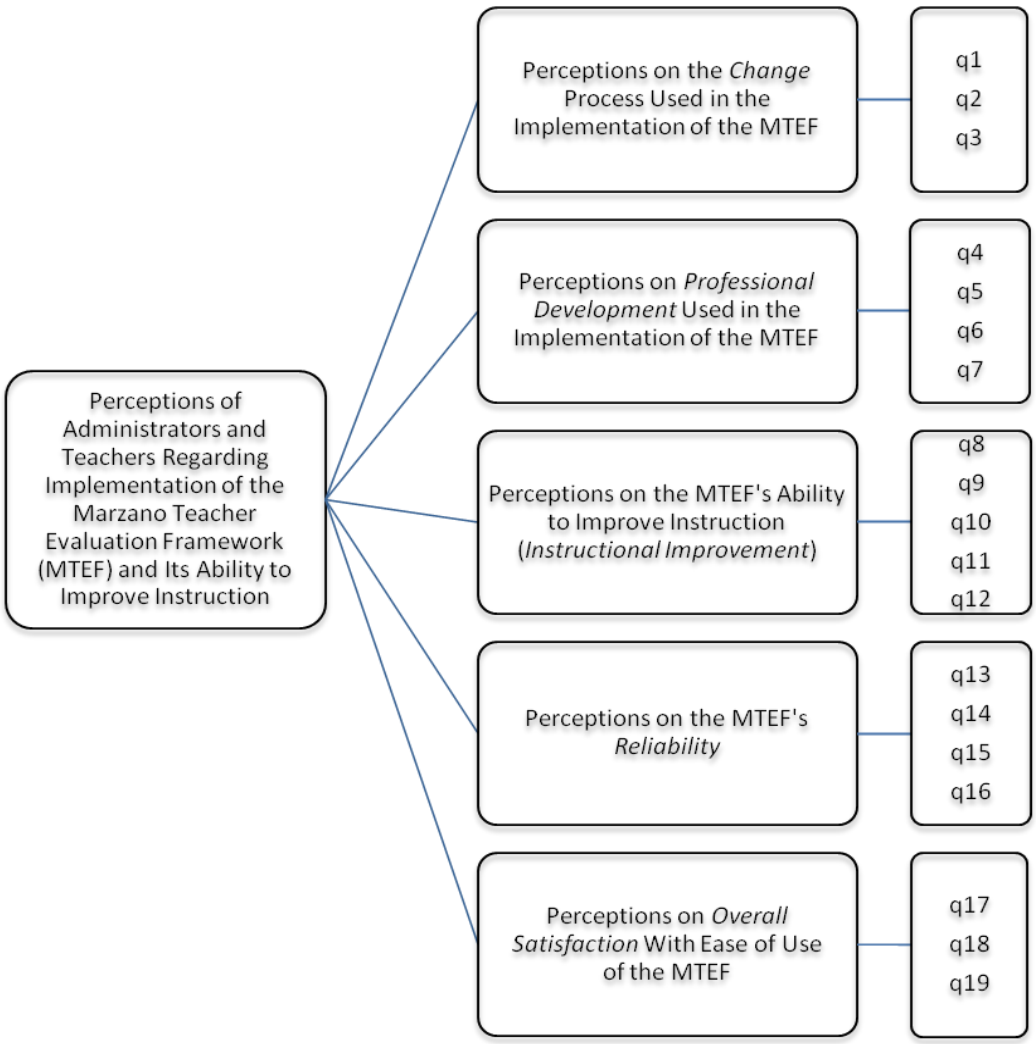

Figure 2. Research Constructs and Survey Question Alignment.

Change

Fullan's (2011) *Choosing the Wrong Drivers for Whole System Reform* examined "levers" that have the best chance of successfully implementing change. In Chapter I of this dissertation, the section titled *Theoretical Framework* outlined characteristics necessary to implement change. The following section on change explains in greater detail four drivers for whole system reform.

Capacity Building Versus Accountability

Whole system reform will only happen through intrinsic motivation and improved competencies of teachers working together purposefully for a common goal (Finkelstein, 2016; Fullan, 2011; Minnici, 2014; Sutton & Shouse, 2016). Unfortunately, in the United States, politicians and the public have led with the wrong drivers to enact change. The idea that increased accountability, assessments, standards, rewards, and punishment would motivate teachers and change education is false (Derrington & Campbell, 2015; Firestone, 2014; Fullan, 2011). Models being used at the time of this study to enact change have not been building widespread capacity, increasing intrinsic motivation, or addressing day-to-day culture of school systems. Changing the focus of improvement from accountability to building teachers' capacities is a process that takes time (Fullan, 2011). Unfortunately, policy makers do not like to take capacity into account, which is the reason for so many failed initiatives over the past 30 years (Fullan, 2011; Fullan, Cuttress, & Kilcher, 2005).

Building capacity starts with policies, resources, strategies, and actions that aim at increasing participants' collective power moving forward (Fullan et al., 2005; Goe, Holdheide, & Miller, 2011). Developing new knowledge, competencies, and skills designed to create new identities and increase motivation allow the change process to start moving forward (Firestone, 2014; Fullan et al., 2005). It is essential capacity building be done as a group, whether is it is an individual school, a district, or an entire system, all individuals must work together (Fullan, 2011; Saltzman, 2016; Sutton & Shouse, 2016). According to Fullan et al. (2005), building capacity is oftentimes the missing element when stakeholders agree on change.

Increasing Group Quality Versus Individual Quality

In order for change to take place, administrators of a system must focus on changing the culture of an organization, not individuals (Fullan, 2011; Robinson, 2015; Sun, 2011). Schools will not be successful by improving abilities of individual teachers sporadically throughout a school or district. Instead, systems that are successful improve 95% of their teaching staff (Fullan, 2011).

In a study conducted by Carrie Leana (2011), she discussed research findings as she examined differences between human capital and social capital and their effects on school reform. Human capital was defined as "factors such as teacher experience, subject knowledge, and pedagogical skills" and "'social capital' – the patterns of interactions among teachers" (Leana, 2011, p. 32). Historically, school change initiatives in the United States have focused on trying to increase human capital with very little attention paid to improving social capital as a way to improve schools (Leana, 2011).

Over the past 30 years, politicians have focused on improving teachers' human capital as a way of improving student achievement; politicians have had little success with this effort (Fullan, 2011; Goodwin, 2011; Leana, 2011; Sahlberg, 2011). According to Leana, researchers have found minimal correlation between teachers accumulating more education and student achievement. Instead, Leana concluded from her research findings that when teachers build trusting relationships with their colleagues and have frequent interactions that revolve around instruction and student learning, social capital is high and student achievement scores improve. Teaching is not a profession that can be practiced in isolation; if it is going to be done successfully, it must be an ongoing collaborative effort (Hargreaves & Fullan, 2011; Kilgore & Reynolds, 2011; Sahlberg, 2011). High social capital is essential

for student achievement; and when combined with teachers that have high human capital, that is when students showed the greatest gains in achievement (Leana, 2011). However, Leana was careful to point out that building social capital between teachers must come before a school focuses on increasing human capital, if there is going to be system-wide improvement and change.

According to Fullan (2011), "Social capital is measured in terms of the frequency and focus of conversations with peers that are centered on instruction, and that are based on feelings of trust and closeness between teachers" (p. 11). Social capital is so powerful it has the ability to make low-ability teachers perform as well as average teachers, and schools that lack social capital can make good teachers less effective (Fullan, 2011; Goodwin, 2015; Leana, 2011). When individuals work together with purpose for a common goal and are intrinsically motivated, they will produce better results and increase accountability (Coleman, 1988; Firestone, 2014; Fullan, 2011).

Numerous researchers have cited teachers as being the key factor to student success and to improving schools. However, a key essential missed in many research findings is that effective teaching starts with social capital. In a 2011 OECD (Oganisation for Economic Cooperation and Development) report, it was concluded that transforming schools into learning organizations must start with teachers leading. Peers working with peers and building trust increases accountability, and motivates teachers to improve their instructional practices (Coleman, 1988; Gruenert & Whitaker, 2015; Kilgore & Reynolds, 2011; Leana, 2011; Sutton & Shouse, 2016). According to Fullan (2011), "If you want the instructional practices-student engagement/achievement nexus to be the centre of attention do two things: name it as the focus, and use the group to get more of it" (p. 14).

Instruction as the Focus Versus Technology

Often in education, technology has been seen as a solution instead of a partner when schools are looking to improve. Over the last 40 years, technology has continued to improve while instructional practices have seen very little change (Fullan, 2011). The idea that putting a device in every student's hand will somehow make him or her smarter, more engaged, and more knowledgeable has not proven to be true (Fullan, 2011). Technology for the sake of technology has the potential to be a distraction in a school setting instead of an accelerator for learning (OECD, 2011). Teachers must be grounded in effective pedagogy before they can decide how to best implement technology into their classrooms (Fullan, 2011). According to Fullan, there are no research findings that technology should be the lead driver for educational inform, but it can be a highly useful tool when combined with effective instructional practices and highly motivated teachers.

The U.S. Department of Education, Office of Educational Technology (2010) challenged educators to "leverage technology to create relevant learning experiences that mirror students' daily lives and the reality of their futures" (p. 9). Today's students want to be active participants in their learning. The net generation is not interested in listening to lectures and filling in a worksheet, but would rather have conversations and choices of what they learn, and they want learning to be relevant to real-world experiences, interesting, and fun (Sheskey, 2010; Tapscott, 2009).

Students of today have been described as, "experimental, engaged, and constantly connected" and excel in "learning environments that are active, social, and learner-centered" (Ramaley & Zia, 2005, p. 8). Identification of the importance of experimental learning goes back to John Dewey (1938), in his book, *Experience in Education*. Dewey stated, "There is

an intimate and necessary relation between the process experience and education" (p. 7). Learning for net generation students should be participatory and challenging. According to McNeely (2005), if Net Generation students are not challenged appropriately, they get bored and become easily disengaged. Today's students learn best through discovery and exploration by themselves or with other students. The exploratory style of teaching helps students better retain material and use it in meaningful and creative ways (McNeely, 2005; Oblinger & Oblinger, 2005; Tapscott, 2009).

As previously mentioned by Fullan (2011), technology alone will not improve education. According to P21: Partnership for 21st Century Learning (2007), 21st-century teaching and learning prepares students for the more complex and ever-changing work environment students will face after their formal education. Skills that will be required of students as they enter the workforce include critical thinking, communication, collaboration, and creativity. To implement 21st century skills in a classroom, education experts generally are in agreement that "21st century competencies and expertise such as critical thinking, complex problem solving, collaboration, and multimedia communication should be woven into all content areas" (U.S. Department of Education, Office of Educational Technology, 2010, p. 13).

Systemic Versus Fragmented Implementation

The United States education system has traditionally implemented initiatives piece by piece and as a result failed to systemize successfully an approach to education (Fullan, 2011). NCLB started with highly qualified teachers and standardized tests, next came *Race to the Top*, followed by Elementary and Secondary Education Act (ESEA) waivers, and most recently, the Common Core standards. All of them had little to no effect on student

achievement. Fullan (2011) pointed out that when systematic approaches are not used, initiatives become fragmented and fall apart during implementation.

Fullan's (2011) definition of systemic does not mean that all elements need to be linked; instead, systemic strategies are those that "require and support on-the-ground improvement efforts in every school and every district" (p. 16). By focusing on right drivers, capacity building, group work, and improved pedagogy, all schools will engage in continuous improvement (Fullan, 2011; Sutton & Shouse, 2016). A systemic mindset creates a belief that quality education for all is crucial to our future (OECD, 2011) and that everyone is part of the solution (Fullan, 2011). When policy leaders understand that teachers are the key to improvement, and they can only be successful when they are supported, then change begins to take place (Fullan, 2011; Goe et al., 2014; Minnici, 2014). It is essential that education leaders work to hire, recruit, and train new teachers, support them through their early years, provide them with opportunities for growth during their career, and finally, provide good working conditions that include team development (Fullan, 2011; Sun, 2011). There is not one single thing that drives successful change; it is a systemic mindset that allows a system to improve and change (Fullan, 2011; Reeves, 2009).

A key belief that all successful system administrators have come to understand revolves around the idea of trusting and respecting teachers; without them change will not occur (Fullan, 2011). For the United States, this would require getting rid of low-trust strategies and start a process of engaging in professional discourse that leads to commitment and solutions (OECD, 2011). Following a systemic implementation plan that focuses on trust, support, relationships, and instructional improvements will build greater accountability

than any type of measure that is implemented (Finkelstein, 2016; Fullan, 2011; Stanulis, Cooper, Dear, Johnston, & Richard-Todd, 2016).

Professional Development

Research confirms the most important factor that contributes to students' success in school is the quality of a teaching staff (Baete & Hochbein, 2014; Looney, 2011; Marzano et al., 2011; Mizell, 2010; Pecheone & Whittaker, 2016). Everyone wants their children to go to schools that have excellent teaching for every child every day. The problem begins when individuals try to decide how to accomplish this task. Great teachers and great schools do not happen by chance. If teachers are to improve, they need to continually expand their knowledge base and skills to help implement the best and most current instructional strategies, and the only way to accomplish this task is through effective professional development that is designed with teachers as a guiding force (Seely Flint, Zisook, & Fisher, 2011; Mizell, 2010; Ritter & Barnett, 2016; Stewart, 2011; Sergiovanni & Starratt, 1971).

Unfortunately, many people do not understand the process and time needed to effectively collaborate and improve teaching and learning (Mizell, 2010). In a study conducted by Tucker (2011), he examined education policies and practices of the United States and compared them to countries that lead the world in student achievement. The countries studied included Finland, Japan, Canada, China, and Singapore. All have been far ahead of the United States in student performance. One of the areas examined was continuing professional development and its connection to instruction. Of countries studied, Tucker found teachers' professional development was teacher led, highly valued, informed by the latest and best research, and closely aligned with day-to-day teaching. Also, collaboration has been valued and expected in these countries. In Singapore, they have a policy that states,

"teach less, learn more," which frees up time outside the classroom to meet with students, plan, do research, and collaborate with colleagues (Darling-Hammond, 2010). Also, in Singapore, they value professional development so much that teachers are guaranteed 100 hours of professional development a year (Darling-Hammond, 2010). In the United States, both Tucker (2011) and Stewart (2011) concluded that professional development has been primarily top-down, driven with topics chosen by the administration that teachers most likely would not have selected, which leads to teachers devaluing professional development days.

Effective professional development provides educators with knowledge they need to address learning challenges of all students and improve their performance (Seely Flint et al., 2011; Learning Forward, 2011; McGuinn, 2015; Stanulis et al., 2016). "To be effective, professional development requires thoughtful planning followed by careful implementation with feedback to ensure it responds to the educators' learning needs" (Mizell, Hord, Killion, & Hirsh, 2011, p. 10). Implemented correctly, professional development has the ability to change educators' practice and increase student learning (Darling-Hammond, 2014; Stanulis et al., 2016). Learning Forward (2011) advocated for sustained implementation of a change that takes place over a 3-5 year time period; educators need time and on-going implementation support to allow them to deepen their knowledge and understanding of why new changes are occurring.

Over the years, at least in the United States, professional development has had very little effect on student learning (Diaz-Maggioli, 2004). Intentions are noble, to improve student learning, yet there is an enormous gap between what a teacher desires for professional development and what they receive (Seely Flint et al., 2011; Minnici, 2014; Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009). According to Diaz-

Maggioli, professional development should be approached as a career-long process that allows teachers to improve their teaching skills and meet the needs of their students. Too often, professional development is a top-down approach that does not take into consideration voices of teachers, which creates an element of distrust between teachers and school leaders (Diaz-Maggioli, 2004; Mizell et al., 2011; Tucker, 2014).

As far back as 1971, Sergiovanni and Starratt advocated for teachers to be involved in planning their own professional development; Sergiovanni and Starratt said it was morally wrong if teachers were not given the opportunity to select from a variety of options to improve their performance. Learning is an individual matter, and teachers need to be involved in the process by determining their course of action (Seely Flint et al., 2011; Sergiovanni & Starratt, 1971). Effective professional learning takes place when teachers are intrinsically motivated and perceive a need for change (Firestone, 2014; Kotter, 1996), not when there is external pressure to learn about a given topic designed with a top-down approach that does not take into account the individual needs of teachers (Ritter & Barnett, 2016; Sergiovanni & Starratt, 1971). As a whole, teachers view themselves as competent professionals that know what is best for their students, and when they work in an environment where they are not respected, distrust and apathy become the norm (Sergiovanni & Starratt, 1971). Developing a foundation of trust is complex, yet essential if schools are going to have open communication about what is best for student learning (Kilgore & Reynolds, 2011; Marzano, Waters, & McNulty, 2005).

Sergiovanni and Starratt (2002) defined school climate as "the enduring characteristics that describe the psychological character of a particular school, distinguish it from other schools, and influence the behavior of teachers and students, and is the

psychological ‘feel’ that teachers have for that school” (p. 82). These characteristics include goal focus, communication adequacy, optimal power equalization, resource utilization, cohesiveness, morale, innovativeness, autonomy, adaptation, and problem-solving competence (Sergiovanni & Starratt, 2002). How these characteristics operate within a school will determine whether it is an open climate that supports learning or a closed climate that hinders a teachers’ learning (Diaz-Maggioli, 2004).

Instructional Improvement

Researchers agree, teacher action in the classroom is the leading factor in student achievement (Kane & Stainger, 2008; Marzano, 2003; Rockoff, 2004; Tyler, Taylor, Kane, & Wooten, 2010; Wright, Horn, & Sanders, 1997). Improving teachers’ skills that translate into higher student achievement is no simple task. Darling-Hammond (2013) advocated that if we want to achieve greatness “we will have to teach our way to stronger student learning by supporting teachers’ collective learning” (p. 7). However, according to City, Elmore, Fiarman, and Tietel (2009), the greatest barriers education professionals encounter when looking to improve schools is a lack of an agreed upon definition of what is quality instruction. Danielson (1996), Dean et al. (2012), and Marzano (2003) recommended providing teachers with a research-based instructional framework that allows for variation in each teacher’s approach based on their students’ needs and a teacher’s strengths. Instructional frameworks provide teachers with highly effective research-based instructional strategies that help inform instruction and give teachers a framework for their professional growth, which is essential for improving student achievement (Darling-Hammond, 2013; Cherasaro, Brodersen, Reale, & Yanoski, 2016). According to Marzano (2003), studies have

proven effective teachers use more research-based instructional strategies than ineffective teachers use and have a wider variety of strategies at their disposal.

Numerous researchers over the past three decades have gone to great lengths to develop lists of effective teaching strategies that have had positive effects on student achievement. In 1986, former Secretary of Education, William Bennett, published a list of over 40 research-based instructional strategies that were designed to improve teachers' instructional strategies (Bennett, 1986). In 1994, Creemers produced a similar list of instructional strategies to assist teachers. Hattie (1992) also presented a list of instructional strategies; however, his list included effects size and percentile gain next to the strategies allowing educators to see the most effective strategy. In 1996, Danielson published *Enhancing Professional Practice: A Framework for Teaching,* which detailed teaching practices that were proven through empirical research studies to increase student learning. Marzano, Pickering, and Pollock (2001) followed with their book titled *Classroom Instruction That Works: Research-Based Strategies for Increasing Student Achievement,* which included nine categories broken down into 34 specific behaviors associated with an increase in student achievement. Marzano et al. identified these nine categories by synthesizing findings from prior meta-analyses. Each of the nine categories were reported out by average effect size, percentile gain, and standard deviation showing educators which strategies were most effective at improving student learning. In 2012, Dean and colleagues built on Marzano et al.'s (2001) work and published a second edition of *Classroom Instruction that Works: Research-Based Strategies for Increasing Student Achievement*. For the second edition, McREL researchers analyzed the literature that had been published since

2001, updated the teaching strategies to reflect teaching and learning in the 21st Century, and reorganized the structure of the previous nine categories.

Over the years, researchers have investigated numerous factors considered to affect student achievement including socioeconomic status, ability grouping, and heterogeneous versus homogeneous classroom structure, and the one factor that has continued to be the key determining factor in student achievement is the classroom teacher (Wright, Horn, & Sanders, 1997). Wright and colleagues studied third, fourth, and fifth grade students' standardized assessment results that measured academic progress from year to year in 54 school systems in middle and east Tennessee. Roughly 24,000 tests results were examined at each grade level in the areas of math, reading, language, social studies, and science. Results of the study concluded that teacher effectiveness was the dominant factor that affected students' academic gain (Wright et al., 1997). In another study, Rockoff (2004) researched two school districts in New Jersey made up of 10,000 students and almost 300 teachers. Based on his findings, he concluded that teacher quality, which is defined in this study as years of experience, had a substantial impact on improving student outcomes. However, it was not until 2010, that Tyler and his colleagues made the initial connection between classroom instructional strategies and their effects on increasing student achievement. Until that time little was known about which skills, practices, and characteristics caused an increase in student achievement (Tyler et al., 2010). Tyler et al. examined data from the Cincinnati Public School District (CPSD) Teacher Evaluation Systems (TESs), which was based on the Charlotte Danielson Teacher Evaluation Framework, and compared teachers' scores to students' state mandated test scores. Researchers focused on two categories of the TESs, "Creating an Environment for Learning" and "Teaching for Learning" (Tyler et al.,

2010). The results of the study provided evidence that as teachers moved through the TES rating scale from "basic" to "proficient" to "distinguished," student achievement scores also increased (Tyler et al., 2010).

With the realization that a teacher is the single most important variable in education, we are left with two options on how to improve teacher quality; (a) replace poor teachers with better ones, which is unlikely in the current state considering there is a nation-wide teacher shortage, or (b) improve the quality of teachers currently in the profession (Wiliam, 2011). Research on effective teaching strategies is evident. Danielson (2007), Dean et al. (2012), Marzano et al. (2001), and others that came before them have clearly outlined and succinctly defined what teaching strategies are effective for improving student achievement. However, with all we know about effective instructional practices, why do we not see this knowledge consistently being put into practice. According to Freppon (2001), schools need to create an environment of support where teacher collaboration and learning are the norms. Both Sahlberg (2011) and Hargreaves and Fullan (2012) have shared the same idea of creating a collaborative culture between teachers and principals. Schools that have open and supportive cultures understand that teaching is difficult and help from colleagues is necessary if teachers are going to continue to grow and learn (Chenoweth, 2016; Davis, Darling-Hammond, LaPointe, & Meyerson, 2005; Margolis & Huggins, 2012; Sahlberg, 2011). When teachers work together and support each other, they have greater confidence and certainty about what it is they are trying to accomplish and the best way to achieve their desired results (Hargreaves & Fullan, 2012). According to Rosenholtz (1991), "Improvement in teaching is a collective rather than individual enterprise, and that analysis, evaluation, and experimentation in concert with colleagues are conditions under which teachers improve" (p.

73). In his book *Finish Lesson: What Can the World Learn From Educational Change in Finland?* Sahlberg (2011) attributed the success of schools in Finland to a culture of trust and collaboration that has been cultivated between teachers and education authorities. When teachers work in collaborative cultures, they accumulate knowledge and circulate ideas, provide assistance and support, and this increases teachers' confidence and encourages them to be more open to change and improvement (Hargreaves & Fullan, 2012; Margolis & Huggins, 2012; Stanulis et al., 2016).

Reliability

Traditionally, teacher evaluations have consisted of a simplistic rating scale scoring teachers as *outstanding*, *satisfactory*, or *needs improvement*, and evaluations lacked specific feedback on what teachers could do to improve professionally (Danielson, 2010; Darling-Hammond, 2014). Unfortunately, without consistent objective feedback on performance, teachers are unlikely to see professional growth (Covey, 1991**;** Looney, 2011; Ritter & Barnett, 2016). A reliable evaluation model is defined by Looney (2011) as a model "that evaluators' judgments' are consistent across repeated observations" (p. 445). According to Warner (2013), "Internal-consistency reliability or homogeneity, assess the reproducibility of data by examining consistency across content: similar items or similar set of items" (p. 931). Evaluations must provide teachers with an accurate assessment of their teaching, meaningful feedback, and a productive dialogue with evaluators following an observation (Danielson, 2010; Darling-Hammond, 2014; Derrington & Campbell, 2015; Whitehurst, Chingos, & Lindquist, 2015).

In a study conducted by Sartain, Stoelinga, and Brown (2011) that examined a teacher evaluation process in Chicago Public Schools, Sartain et al. concluded that an agreed upon

definition and a common understanding of what good teaching looks like is an essential starting point. For a teacher evaluation system to be transparent and credible, both teachers and administrators must understand what constitutes "good teaching" (Danielson, 2010; Donaldson & Donaldson, 2012; Firestone, 2014; Ritter & Barnett, 2016). It can no longer be acceptable for an administrator to say, "I know good teaching when I see it." Teacher evaluation is most effective when it measures teachers' performances against a clearly defined set of competencies and standards that define high quality teaching and learning, and allows for reflective teaching practices that enable teachers to define strategies that will help them improve and meet standards (Looney, 2011). Developing a common language that is built around effective instruction allows teachers to analyze their own teaching methodology and invites administrators to ask probing question (Danielson, 2010). A common language built by teachers and administrators is a crucial first step in building trust and creating a reliable teacher evaluation model (Donaldson & Donaldson, 2012; Maslow & Kelley, 2012; Sartian et al., 2011).

Once a common language has been developed, evaluators must be trained. A teacher evaluation model is only credible if a highly skilled, competent evaluator is completing it (Cherasaro et al., 2016; Danielson, 2010; Darling-Hammond, 2014; Looney, 2011). Systems that have shown high levels of effectiveness are able to make similar independent judgments based on a set of evaluative criteria (Danielson, 1996). In the Chicago study, Sartain and colleagues (2011) found a discrepancy in scoring between school officials and trained evaluators when it came to scoring teachers in the highest category, "distinguished" as opposed to evaluators that scored teachers as "proficient." According to Danielson (1996), an inter-rater agreement is critical to the reliability of an evaluation model. Teachers need to

know the fundamental principle of equity is applied in a model, and there is consistency among evaluators when scoring and providing feedback (Bill and Melinda Gates Foundation, 2013; Danielson, 2010; Derrington & Campbell, 2015; McGuinn, 2015). District and school leaders are not able to leave this critical aspect of teacher evaluation to chance. According to Danielson (2012), ongoing training and consultation are essential for inter-rater reliability. Without proper training, objective feedback and reliability of results could be threatened (Stumbo & McWalters, 2011). Untrained evaluators are more likely than trained evaluators to have bias enter into their evaluations and have their expectations of teachers influence the results of their evaluations rather than actual teacher behaviors (Cherasaro et al., 2016; Darling-Hammond, 2014; Muijs, 2006). Evaluators must be given sufficient training and multiple opportunities to practice using an evaluation framework in order to calibrate their judgment with their colleagues (Danielson, 2010).

Correctly being able to identify quality of instruction is essential for an evaluation system to be reliable; however, being able to effectively provide teachers with potentially discomforting or unwelcome feedback based on an evaluation is a skill that is equally important (Donaldson & Donaldson, 2012). As far back as 1922, Cubberley stated, "The supervisor must first of all try to establish good personal relations with the supervised" (p. 241). He then goes on to say,

> Kindliness, consideration, and helpfulness are necessary to win the confidence of teachers, and unless teachers can feel that the supervisor is a friend interested in their success, instead of a critical representative of the board or of the central office, helpful relations are not likely to be established between them.
>
> (Cubberley, 1922, p. 241)

Fast forward almost 100 years, and not a lot has changed. Teachers must be willing to accept evaluators' judgments and engage in productive conversations about their performance for improvement and change to occur (Danielson, 2010; Ritter & Barnett, 2016). Teacher development is more than just a scoreboard about what happened during an evaluation; that is just the starting point.

According to Bambrick-Santoyo (2012), teacher development takes place when evaluators skillfully coach and work with instructors to develop concrete actions that will improve student achievement. Tuytens and Devos (2011) examined school level leadership activities and applied it to teacher evaluations and the feedback provided to teachers, and concluded that the tide is turning in the usefulness of feedback based on teachers' evaluations. Tuytens and Devos found school leaders that used active leadership supervision, charismatic leadership, and displayed leadership content knowledge were able to successfully and positively effect teachers' professional growth using a quality teacher evaluation framework.

In the last decade, researchers have argued that school leaders are crucial components of effective teacher evaluation (Blase & Blase, 1999; Davis, Ellett, & Annunziata, 2002; Robinson, Lloyd, & Rowe, 2008), and the research conducted by Tuytens and Devos (2011) confirmed the positive effects that school leaders can have on instructional improvement. Evaluators that are able to recognize different components of classroom practice, interpret evidence against a set of agreed upon standards, engage teachers in highly productive conversation regarding their classroom instruction, and provide them with ideas for improving instructional strategies can have a positive effect on student achievement (Danielson, 2010; Donaldson & Donaldson, 2012; Looney, 2011). Leadership strategies

provided by Tuytens and Devos' (2011) research have provided school leaders with specific strategies that, when used correctly, will reduce teachers' fear of evaluation, convince them of its usefulness, and create a positive overall evaluation experience.

Overall Satisfaction with Ease of Use of an Instrument

Darling-Hammond (2013) stated, "The final requirement of a productive evaluation system is that it be feasible to implement well—on the part of both the evaluators and those being evaluated—and that it be adequately resourced to be effective" (p. 132). Systems should have a user-friendly design and not overwhelm teachers or administrators. According to Darling-Hammond (2013), there are three major issues that require attention when implementing a new teacher evaluation model:

- there must be adequate human resources available to provide necessary support to implement a teacher evaluation system,
- the sustainability of a teacher evaluation system must be balanced with other aspects of a school district's operations, and
- the appropriateness and manageability of measures of teaching must be reviewed and agreed upon as positively affecting students. (Darling-Hammond, 2013)

Principals in the United States tend to become overwhelmed with the numerous responsibilities expected of them compared to their colleagues in other countries that are not responsible for non-academic issues that occur in school (e.g. bussing, business matters, building issues, and so on) (Darling-Hammond, 2013; Shakman, Breslow, Kochanek, Riordan, & Haferd, 2012). It is easy to see why principals are not able to commit the

necessary time needed to adequately support and evaluate teachers under conditions present at the time of this study (Kelleher, 2016). One of the failures of teacher evaluation systems in the United States has been a reliance on principals to be the sole evaluator of teachers, expected to observe, mentor, coach, document, and make the final call on dismissal, if necessary (Darling-Hammond, 2013; White, Cowhy, Stevens, & Sporte, 2012). However, at the time of this report, researchers had began advocating for districts to share the responsibility of teacher evaluation among varies stakeholders including: district personnel, associate principals, department chairs, and master and mentor teachers (Bill and Melinda Gates Foundation, 2013; Darling-Hammond, 2013, White el al., 2012). The idea has been to reduce principals' workloads and provide teachers with more than one person's point of view in order to increase reliability of evaluations.

The overall goal of teacher evaluation is to provide educators with actionable feedback for improving practice. Therefore, standards must include a common language that is clear and understood by all stakeholders (Kane, 2012; Ritter & Barnett, 2016; White el al., 2012). At the time of this study, recent literature had called for teacher evaluations to make use of a theoretical framework to identify effective teaching, and to implement a process that ensures reliability, with the end goal of improving instructional practices and student learning (Coggshall, Rasmussen, Colton, Milton, & Jacques, 2012; Goe, Biggers, & Croft, 2012). White et al. (2012) found that teachers in five Illinois School Districts using the Danielson Teacher Evaluation Framework described the standards and rubrics as precise and appropriate for all teachers regardless of grade level and subject taught. The end result of the new evaluation system was a common language that centered around quality instruction creating an environment of accountability between colleagues.

Summary

Chapter II examined the history of teacher evaluation, evaluation trends being used in schools at the time of this study, and an examination of three models that school districts across the country were using with teachers at the time of this report. Additionally, the relationships of the five constructs – *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with ease of use of the Marzano Teacher Evaluation Framework model – framing this study were examined as they relate to the implementation of a new teacher evaluation model.

This study investigated the perceptions of administrators and teachers as they relate to the implementation of the Marzano Teacher Evaluation Framework. Chapter III will present a description of the survey instrument used in this study and the methodology utilized in this study's data collection process.

CHAPTER III

RESEARCH METHODS

The purpose of this study was to test the effectiveness of implementation of the Marzano Teacher Evaluation Framework (MTEF), which was used in a school district in the Midwest. Teachers and administrators were surveyed to gain an understanding of their perceptions of implementation of the MTEF. Participants were surveyed on the constructs that framed the study, which included: *change*, *professional development*, and the model's ability to improve instruction (*instructional improvement*), its *reliability*, and *overall satisfaction* with the instrument. The *change* construct examined the process the school district used to implement the MTEF. The *professional development* construct examined the training used to inform teachers and administrators about how to use the new model. The *instructional improvement* construct examined teachers' and administrators' beliefs in the ability of the MTEF model to improve teachers' instructional skills. The *reliability* construct examined beliefs of teachers and administrators regarding whether or not the MTEF accurately describes and measures effective instructional practices. The *overall satisfaction* construct examined teachers' and administrators' overall comfort with the new teacher evaluation model.

Chapter III presents methods used to collect and analyze data. Description of research population, survey instrument, collection of data, and data analysis follow.

Research Questions

Research questions include:

1. What are the perceptions of one school district's administrators regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?
2. What are the perceptions of the school district's teachers regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?
3. Was there a difference between the school district's administrators and teachers regarding the five research constructs that frame this study, which include *change, professional development, instructional improvement, reliability*, and *overall satisfaction* with the new Marzano Teacher Evaluation Framework model?

Description of the Research Population

The research population includes teachers and administrators from a Midwest school district considered a leader in their state regarding implementation of a standards-based teacher evaluation model, the MTEF. The school district's implementation leaders have traveled to school districts around the state sharing, training, and leading other school districts through the MTEF teacher evaluation implementation process. Several members from the implementation committee have also presented at local and state-level meetings and conferences, and via webinars, explaining the implementation process their school district used to transition to a new MTEF teacher evaluation model.

Potential research participants were defined as all building-level principals, associate principals, building resource coordinators (BRCs), and teachers in the participating school district. Research participants at the administrative level included three high school principals, four high school associate principals, four middle school principals, three middle school associate principals, ten elementary principals, and two BRCs.

Research participants at the teaching level included 336 elementary teachers from twelve different schools, 167 middle school teachers from four different schools, and 179 high school teachers from three different schools.

Survey Instrument

An on-line survey instrument for both administrators and teachers (Appendix A and Appendix B, respectively) was posted using SurveyMonkey®. The survey was developed based on the researcher's review of literature, feedback from colleagues, the theoretical framework for this study, a "Stages of Concern" questionnaire (SoCQ; Hall, Wallace, & Dossett, 1973), and a dissertation conducted by Canelake (2012) on the implementation of a teacher evaluation model. In addition, both principal and teacher surveys went through several drafts while the researcher's cohort instructor, advisor, and school district personnel reviewed each of them and provided feedback.

The teacher and the administrator surveys were given at different times in the implementation phases of the MTEF teacher evaluation model. The teacher survey was given in the Fall of 2013, and the administrator survey was given in the Spring of 2014, approximately 4 months apart. The researcher's intent was to compare teachers' and administrators' results verbatim, but at slightly different stages of the MTEF implementation.

The survey consisted of three different sections. Section 1 included three questions regarding participants' demographics. Questions in this section asked participants about the school level where they worked (elementary, middle, or high school), their highest degree earned (bachelor, master, specialist, or doctorate), and their years of experience working in education.

Section 2 consisted of five "yes" or "no" questions. Each of the questions were directly aligned with the five research constructs: *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with the new MTEF teacher evaluation model. The purpose of the five questions was to determine participants' overall perceptions of each construct. An example item was, "I have been satisfied with the professional development that has been conducted during the implementation of the Marzano Teacher Evaluation Framework at the building and district level." The questions in Section 2 were identical for both teachers and administrators.

Section 3 consisted of 19 questions aligned with the five research constructs. The number of questions per research construct ranged from three to five depending on information needed to gain an understanding of each of the specific constructs. Participants answered each question based on a six-point Likert-type scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *slightly disagree*, 4 = *slightly agree*, 5 = *agree*, 6 = *strongly agree*).

The *change* construct questions were based on work of Michael Fullan (2011) and his Drivers for Whole System Reform. An example question from both the teacher and administrator survey is, "The process used by the school district to determine the need for a new teacher evaluation was appropriate."

The second construct examined *professional development* used by the district to implement the new MTEF teacher evaluation model. An example question for teachers asked, "To date, the professional development on the Marzano Teacher Evaluation Framework has met my needs as a teacher"; whereas, the question for administrators was, "To date, the professional development on the Marzano Teacher Evaluation Framework has met my needs as an administrator." The professional development questions were based on work of Mizell (2010) and the professional development standards of Learning Forward (2011).

The third construct, *instructional improvement*, examined teachers' and administrators' perceptions of the MTEF's ability to improve teachers' instructional practice. An example question for teachers asked, "The Marzano Teacher Evaluation Framework is an effective tool to help me improve my development as a teacher." Whereas the question for administrators asked, "The Marzano Teacher Evaluation Framework is an effective tool to influence my teachers' development." The instructional improvement construct questions were based on work of Danielson (1996), Dean et al. (2012), and Marzano (2003).

Reliability was the fourth construct and examined teachers' and administrators' perceptions of the MTEF model's ability to consistently rate teachers. This construct was based on work of Looney (2011), who stated that reliability of evaluators across evaluations must be consistent. An example question for both teachers and administrators asked, "I am confident that most administrators' ratings would be similar if they were rating the same teacher while using the Marzano Teacher Evaluation Framework."

The fifth and final construct measured teachers' and administrators' *overall satisfaction* with the MTEF evaluation tool. Questions on satisfaction were drawn from the

work of Darling-Hammond (2010). According to Darling-Hammond, an evaluation system must be feasible to implement well and adequately resourced for both teachers and administrators. An example question for both teachers and administrators stated, "Overall, the new Marzano Teacher Evaluation Framework is easy to use."

Data Collection

Research and data collection was conducted using quantitative methods. An on-line survey was used to gather data from K-12 teachers and building level administrators. An advantage to the online survey approach, in the case of this study, was to keep participants' responses confidential, which may have helped to ensure honest responses (Rudestam & Newton, 2007). There were two different forms of the survey used. One survey was designed for teachers and one was designed for administrators. While there were two different surveys, they both asked the same questions using wording appropriate for the two separate audiences.

Online surveys are known to present problems of both coverage and nonresponse bias (Gay, Mills, & Airasian, 2009). "Nonresponse error arises through the fact that not all people included in the sample are willing or interested in completing the survey" (Couper, 2000, p. 473). In this study, the sample was limited to teachers and administrators in one school district. It is possible participants in this survey were motivated to complete the survey due to their interest in the topic.

Data used to assess *teachers'* perceptions of the MTEF evaluation model and the implementation process was gathered in the Fall of 2013 by the same researcher. Research was conducted with IRB approval (Project Number IRB-201310-116, Appendix C) and with school district approval (Appendix D). Participants asked to be part of this study included all district elementary, middle, and high school teachers. In the school district, elementary

teachers have been defined as teachers of Grades K-5; middle school teachers, Grades 6-8; and high school teachers, Grades 9-12. An email was sent to teachers describing the purpose of the research, the researcher's background, and, if they chose to participate, instructions and a link to the survey (Appendix E). The survey was open for 14 days. When 2 days remained, the researcher sent out a single reminder email to all teachers. The survey-return rate was 48% (321/682). Out of the three-school levels surveyed, elementary teachers had the lowest return rate at 33.6% (n = 110). High school teachers returned the survey at a rate of 65.4% (n = 114), and middle school teachers returned the survey at a rate of 56.7% (n = 97).

The researcher received IRB approval (Appendix F) and school district approval from the assistant superintendent (Appendix G) to conduct a second round of research with administrators, as well as a letter of support from the director of curriculum, instruction, assessment, and professional development, who was responsible for the implementation of the Marzano Teacher Evaluation Framework in the school district (Appendix H).

The second set of data was collected from the school district's administrators using the same process used to survey the teachers. An email was sent to administrators describing the research, the researcher's background, and instructions along with a link to the survey, if they chose to participate (Appendix I). The survey was open for 14 days. With 2 days remaining before the survey closed, the researcher sent a reminder email to administrators informing them they still had time to participate in the study if they were interested. Administrators' return rate was 100% (26/26).

Data Analysis

All analysis of data was conducted using SPSS statistical software (Version 21, IBM, 2012). The researcher first examined demographic categories including school level, highest

degree earned, and years of experience for both teachers and administrators. Each category was broken down into percentages for individual subgroups for comparison purposes. Next, Section 2 containing five general questions, one for each construct, was examined to determine percentage of agreement with each statement.

To answer Research Questions #1 and #2, descriptive statistics, including percentage of agreement, standard deviation, and mean scores were examined for each of the five constructs: *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with ease of use of the MTEF instrument. Research Question #1 asked, "What are the perceptions of one school district's administrators regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?" and Research Question #2 asked, "What are the perceptions of one school district's teachers regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?" In addition, data were examined and compared across research constructs to determine differences, if any, between administrators' and teachers' perceptions. Subscales were tested for internal consistency, reflected by Cronbach alpha scores.

Lastly, individual *t*-tests were conducted to answer the third and final research question that compared differences between administrators' and teachers' perceptions regarding the five individual constructs that framed the study. An independent variable for this study was the role an educator played in the district, i.e. administrator or teacher, while dependent variables were participants' perceptions of the implementation process. Statistically significant differences were determined at the $p < .05$ level, rejecting the null

hypothesis that teachers and administrators would have the same perceptions regarding implementation of the MTEF model.

Summary

Chapter III included the purpose of the study, research questions, a description of the research population, an explanation of the survey instrument, and a discussion of the process used for data analysis. In Chapter IV, data gathered in this study will be presented. In Chapter V, the researcher will include a summary and discussion of the findings.

CHAPTER IV

RESULTS

The purpose of this study was to test the effectiveness of the implementation of the Marzano Teacher Evaluation Framework (MTEF), which was used in a school district in the Midwest. The researcher examined teachers' and administrators' perceptions regarding *change, professional development, instructional improvement, reliability*, and *overall satisfaction* with ease of use of the Marzano model of teacher evaluation. The study was comprised of quantitative methods to assess the effectiveness of the MTEF.

Chapter IV includes results used to answer the following research questions:

1. What are the perceptions of one school district's administrators regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?
2. What are the perceptions of one school district's teachers regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?
3. Was there a difference between the school district's administrators and teachers regarding the five research constructs that frame this study, which include *change, professional development, instructional improvement, reliability*, and *overall satisfaction* with the new Marzano Teacher Evaluation Framework model?

Research Population

Research participants were all teachers and administrators in a Midwest school district. Administrators included all building level principals, associate principals, and BRCs. Administrative research participants included three high school principals, four high school associate principals, four middle school principals, three middle school associate principals, 10 elementary principals, and two BRCs. The teacher research participants included 336 elementary teachers from 12 different schools, 167 middle school teachers from four different schools, and 179 high school teachers from three different schools. The district's official approval of the study was granted by the assistant superintendent of schools (Appendix D and Appendix G), and endorsed by the director of curriculum, instruction, assessment, and professional development (CIAPD; Appendix H). The CIAPD director and teacher evaluation committee were in charge of implementing the MTEF.

Research participants were each sent an email inviting them to participate in the study. The email consisted of an explanation of the study, consent for participation in the study, assurance that responses would be anonymous, a link to the survey, estimation of time needed to complete the survey, directions for the survey, and the opportunity to opt out of the research study **(**Appendix E and Appendix I).

Research Questions 1 and 2 examined teachers' and administrators' perceptions of implementation of the MTEF and its ability to improve instructional practices. Results pertaining to the first two research questions will be combined and examined by comparing the constructs of *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with the MTEF model's ease of use as perceived by administrators and teachers.

Data in Table 3 show an overview of administrators who responded to the survey. Administrative participants consisted of 26 building level leaders, all of who completed the survey for a 100% response rate. The demographic breakdown is included in Table 3.

Table 3. Administrators' Demographic Information.

Demographic Category	Overall Sample Count (n = 26)	%
School Level		
Elementary School	12	46.2
Middle School	7	26.9
High School	7	26.9
Highest Degree Earned		
Master's Degree	13	50.0
Specialist's Degree	6	23.1
Doctorate Degree	7	26.9
Years of Administrative Experience		
0-5	9	34.6
6-10	4	15.4
11-15	10	38.5
15 or more	3	11.5

Table 4 shows demographic information of teachers who participated in this study. The school district has 682 teachers. A total of 328 teachers completed the survey for a response rate of 48%. High school teachers returned the survey at a rate of 65.4% (117/179). The lowest return rate came from elementary teachers, 33.6% (113/336). Middle school teachers returned 56.7% of their surveys (98/173).

Table 4. Teachers' Demographic Information.

Demographic Category	Overall Sample Count (n = 328)	%
School Level		
Elementary School	113	34.5
Middle School	98	29.9
High School	117	35.6
Highest Degree Earned		
Bachelor's Degree	108	32.9
Master's Degree	216	65.9
Doctorate Degree	4	1.2
Years of Teaching Experience		
0-8	65	19.8
9-16	88	26.8
17-24	96	29.3
25 or more	79	24.1

Table 5 shows school district administrators' overall percentage of agreement concerning the five constructs of the research study: *change* (c1), *professional development* (c2), *instructional improvement* (c3), *reliability* (c4), and *overall satisfaction* (c5) with the MTEF model's ease of use. In each of the survey questions, school administrators overwhelmingly agreed with the process the school district used to implement the MTEF. Participants were given the option to answer the question with either a "yes" or a "no" response. Answering "yes" meant that administrators agreed with the statement and were satisfied with the school district's implementation process. In the areas of *change* and *professional development* there was 100% agreement that a new teacher evaluation model was needed and the professional development process used by the district to train administrators was appropriate.

Table 5. Questions On Five Constructs Regarding *Administrator* Perceptions Concerning Implementation of the Robert Marzano Teacher Evaluation Framework.

Survey Questions	% Yes
c1. I believe there was a need for the newly adopted Marzano Teacher Evaluation Framework in the [redacted] Public School District.	100
c2. I have been satisfied with the professional development that has been conducted during the implementation of the Marzano Teacher Evaluation Framework.	100
c3. I believe the Marzano Teacher Evaluation Framework will have a positive effect on my ability to improve my teachers' performance.	96.2
c4. I believe the Marzano Teacher Evaluation Framework is a reliable instrument for evaluating teacher effectiveness.	96.2
c5. Overall, I am satisfied with the Marzano Teacher Evaluation Framework instrument and its features.	96.2

Table 6 shows school district teachers' overall percentage of agreement concerning statements about the five constructs of the research study: *change* (c1), *professional development* (c2), *instructional improvement* (c3), *reliability* (c4), and *overall satisfaction* (c5) with the MTEF model's ease of use. Participants were given the option to answer the question with either a "yes" or a "no" response. Answering with a "yes" meant teachers agreed with the statement and were satisfied with the schools district's implementation process. The most positive response from teachers was in the construct of the MTEF model's ability to improve teachers' classroom performance (c3), 72.1% answered "yes" to that statement. The construct of *reliability* (c4) scored the lowest in percentage of teachers agreeing with the statement (61.8%). The *change* (c1) and *professional development* (c2) constructs also received positive responses from staff with 71.5% and 70.9% of staff in

agreement. *Overall Satisfaction* (c5) with ease of use of the model received a 66.1% agreement response from teachers.

Table 6. Questions on Five Constructs Regarding *Teacher* Perceptions Concerning Implementation of the Robert Marzano Teacher Evaluation Framework.

Survey Questions	% Yes
c1. I believe there was a need for the newly adopted Marzano Teacher Evaluation Framework in the [redacted] Public School District.	71.5
c2. I have been satisfied with the professional development that has been conducted during the implementation of the Marzano Teacher Evaluation Framework.	70.9
c3. I believe the Marzano Teacher Evaluation Framework will have a positive effect on my teaching performance.	72.1
c4. I believe the Marzano Teacher Evaluation Framework is a reliable instrument for evaluating teacher effectiveness.	61.8
c5. Overall, I am satisfied with the Marzano Teacher Evaluation Framework instrument and its features.	66.1

Survey Questions

The survey was designed using a six-point Likert-type scale. Participants responded to the survey using *strongly agree* (1), *agree* (2), *slightly agree* (3), *slightly disagree* (4), *disagree* (5), and *strongly disagree* (6). Answers showing some form of agreement with a survey statement included *strongly agree*, *agree*, and *slightly agree*, while answers showing some form of disagreement with a survey statement included *slightly disagree*, *disagree*, and *strongly disagree*. As shown in Table 7, administrators showed some form of agreement with all 18 of the survey questions. Mean scores for administrators ranged from 3.9, as a low score for Question 6 (q6), to a high score of 5.7 for Question 11 (q11).

Table 7. Administrators' Agreement, Mean (*M*) Score, and Standard Deviation (*SD*) for Each Survey Statement.

Survey Questions	% Showing Some Form of Agreement	*M**	*SD*
Change (c1)			
q1. The process used by the [redacted] Public School District to determine the need for a new teacher evaluation model was appropriate.	96.1	5.1	1.0
q2. The process used by the Teacher Evaluation Committee to determine the new teacher evaluation model for the [redacted] Public School District was appropriate.	96.1	5.1	1.1
q3. The process used by the Teacher Evaluation Committee to update and inform stakeholders on the new model was effective in gaining support for the initiative.	88.4	4.8	1.1
Professional Development (c2)			
q4. The use of staff meetings to communicate the phases of the implementation was beneficial.	100	5.0	0.9
q5. To date, the professional development on the Marzano Teacher Evaluation Framework has met my needs as an administrator.	92.3	5.1	0.9
q6. The video vignettes in iObservation are an effective tool in helping my teachers improve their instruction.	69.2	3.9	1.1
q7. I am satisfied with the implementation being slowly phased in over a three year time period.	88.4	5.2	1.1
Instructional Improvement (c3)			
q8. The Marzano Teacher Evaluation Framework is an effective tool to influence my teachers' development.	100	5.2	0.6
q9. The feedback from the Marzano Teacher Evaluation Framework is an effective tool in helping me improve my teachers' instructional performance.	100	5.2	0.7
q10. I use the Marzano Teacher Evaluation Framework feedback to help me effectively guide my teachers' performances.	100	5.2	0.6
q11. Compared to the "old evaluation system," observational feedback is more relevant and meaningful to affirm or alter instruction.	100	5.7	0.7

Table 7. cont.

Survey Questions	% Showing Some Form of Agreement	*M**	*SD*
Instructional Improvement (c3) Continued			
q12. Compared to the "old evaluation system," observational feedback is more immediate.	100	5.5	0.7
Reliability (c4)			
q13. I believe the Marzano Teacher Evaluation Framework will result in consistent ratings among teachers.	84.6	4.3	1.2
q14. I am confident that most administrators' ratings would be similar if they were rating the same teacher while using the Marzano Teacher Evaluation Framework.	80.7	4.1	1.1
q15. The Marzano Teacher Evaluation Framework's scales scoring system allows for consistent scoring of teachers.	80.7	4.3	1.2
q16. I am confident in the consistency of the Marzano Teacher Evaluation Framework.	84.6	4.3	1.2
Overall Satisfaction (c5)			
q17. Overall, the new Marzano Teacher Evaluation Framework is relatively easy to use.	100	4.9	0.6
q18. The teaching standards that are measured by the new Marzano Teacher Evaluation Framework are focused on what is necessary to raise student achievement.	100	5.2	0.7
q19. The rubrics used to measure the teaching standards in our new Marzano Teacher Evaluation Framework are adequately descriptive.	92.3	4.6	1.0

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5; *strongly disagree* = 1, *strongly agree* = 6).

Table 7 shows administrators' responses to each of the survey questions and how questions are aligned with each construct. Survey statement Question 6 (q6), "The video vignettes in iObservation are an effective tool in helping my teachers improve their instruction," had the lowest percentage of some form of agreement among administrators. Only 69.2% of administrators agreed or had some form of agreement with the statement.

Question 6 also had the lowest mean score at 3.9. All other questions had a mean score of 4.1 or above on the administrators' survey. The next lowest score showing percentage of agreements was 80.7% for Question 14 (q14). For the construct *instructional improvement*, 100% of administrators agreed positively to each of the statements about the MTEF and its ability to improve their teachers' instructional practice. The *instructional improvement* construct also showed the highest mean scores with each question scoring 5.2 or above.

Table 8 displays responses for each of the questions in the *professional development* construct. Question 6 (q6) shows the least number of administrators agreeing with the statement that the video vignettes in iObservation are an effective tool in helping teachers improve their performance. The 3.9 mean score for Question 6 is the lowest mean score in the *professional development* construct. The other three statements in the *professional development* construct had mean scores of 5.0 or higher.

Table 8. Administrators: *Professional Development* Construct—Detailed Report/Response.

	Strongly Agree (6)	*Agree* (5)	*Slightly Agree* (4)	*Slightly Disagree* (3)	*Disagree* (2)	*Strongly Disagree* (1)	Mean (*M**)	*n*
q4	10	9	6	1	0	0	5.0	26
q5	9	12	2	3	0	0	5.1	26
q6	0	9	9	4	4	0	3.9	26
q7	15	6	2	2	1	0	5.2	26

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5).

Table 9 gives overall responses to questions in the *instructional improvement* construct. In each of the questions, a majority of the administrators responded either *agree* or *strongly agree*: q8 (84.6%), q9 (84.6%), q10 (96.2%), q11 (88.5%), q12 (88.5%).

Table 9. Administrators: *Instructional Improvement* Construct—Detailed Report/Response.

	Strongly Agree (6)	*Agree* (5)	*Slightly Agree* (4)	*Slightly Disagree* (3)	*Disagree* (2)	*Strongly Disagree* (1)	Mean (*M**)	*n*
q8	7	15	4	0	0	0	5.2	26
q9	9	13	4	0	0	0	5.2	26
q10	8	16	2	0	0	0	5.2	26
q11	20	3	3	0	0	0	5.7	26
q12	17	6	3	0	0	0	5.5	29

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5).

Table 10 shows a detailed breakdown of administrators' responses to statements q13, q14, and q15. The *reliability* construct had the lowest overall mean scores of constructs with three questions having a mean score of 4.3, and one question having a mean score of 4.1. All other constructs had a majority of their statements score above 5.0 for a mean.

Table 10. Administrators: *Reliability* Construct—Detailed Report/Response.

	Strongly Agree (6)	*Agree* (5)	*Slightly Agree* (4)	*Slightly Disagree* (3)	*Disagree* (2)	*Strongly Disagree* (1)	Mean (*M**)	*n*
q13	0	15	5	3	2	1	4.3	26
q14	0	11	9	3	2	1	4.1	26
q15	1	15	5	2	2	1	4.3	26
q16	0	15	5	3	2	1	4.3	26

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5).

Table 11 gives teachers' responses to each of the questions and the constructs questions are aligned with.

Table 11. Teachers' Agreement, Mean (*M*) Score, and Standard Deviation (*SD*) for Each Survey Statement.

Survey Questions	% Showing Some Form of Agreement	*M**	*SD*
Change (c1)			
q1. The process used by the [redacted] Public School District to determine the need for a new teacher evaluation model was appropriate.	81.4	4.3	1.2
q2. The process used by the Teacher Evaluation Committee to determine the new teacher evaluation model for the [redacted] Public School District was appropriate.	80.8	4.3	1.2
q3. The process used by the Teacher Evaluation Committee to update and inform stakeholders on the new model was effective in gaining support for the initiative.	74.1	4.1	1.3
Professional Development (c2)			
q4. The use of staff meetings to communicate the phases of implementation was beneficial.	82	4.4	1.2
q5. To date, the professional development on the Marzano Teacher Evaluation has met my needs as a teacher.	68.3	3.9	1.3
q6. The video vignettes in iObservation help improve my instruction.	60.4	3.5	1.3
q7. I am satisfied with the implementation being slowly phased in over a three year time period.	89	4.8	1.2
Instructional Improvement (c3)			
q8. The Marzano Teacher Evaluation Framework is an effective tool to influence my development as a teacher.	75.6	4.0	1.3
q9. The feedback from the Marzano Teacher Evaluation Framework is an effective tool in helping me to improve my performance.	73.5	4.0	1.4
q10. I use the Marzano Teacher Evaluation Framework feedback to help effectively guide my teaching performances.	79.3	4.2	1.3

Table 11. cont.

Survey Questions	% Showing Some Form of Agreement	*M**	*SD*
Instructional Improvement (c3) Continued			
q11. Compared to the "old evaluation system," observational feedback is more relevant and meaningful to affirm or alter instruction.	78	4.4	1.4
q12. Compared to the "old evaluation system," observational feedback is more immediate.	82.3	4.6	1.4
Reliability (c4)			
q13. I believe the Marzano Teacher Evaluation Framework will result in consistent ratings among teachers.	63.2	3.6	1.4
q14. I am confident that most administrators' ratings would be similar if they were rating the same teacher while using the Marzano Teacher Evaluation Framework.	59.8	3.6	1.4
q15. The Marzano Teacher Evaluation Framework's scales scoring system allows for consistent scoring of teachers.	64.9	3.7	1.3
q16. I am confident in the consistency of the Marzano Teacher Evaluation Framework.	60.4	3.6	1.3
Overall Satisfaction (c5)			
q17. Overall, the new Marzano Teacher Evaluation Framework is relatively easy to use.	73.8	4.1	1.3
q18. The teaching standards that are measured by the new Marzano Teacher Evaluation Framework are focused on what is necessary to raise student achievement.	78.4	4.2	1.3
q19. The rubrics used to measure the teaching standards in our new Marzano Teacher Evaluation Framework are adequately descriptive.	80.5	4.2	1.2

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5; *strongly disagree* = 1, *strongly agree* = 6).

Survey statement (q7), "I am satisfied with the implementation being slowly phased in over a three year time period," had the highest percent (89%) of some form of agreement among teachers. The q7 statement also had the highest mean score on the survey (4.8) and

one of the lowest standard deviation scores (1.18). Teachers scored the *reliability* construct the lowest in percentage of some form of agreement. Survey statement q14, "I am confident that most administrators' ratings would be similar if they were rating the same teacher while using the Marzano Teacher Evaluation Framework," scored the lowest percent rating (59.8%) of some form of agreement. Overall, the *reliability* construct showed the lowest percentage of some form of agreement among teachers with the highest percent score for a statement being 64.9%.

Table 12 shows responses from teacher participants regarding the *change* construct. Survey statement "q1" had the highest percentage of some form of agreement from respondents with 81.4% of respondents agreeing with the statement, "The process used by the [redacted] Public School District to determine the need for a new teacher evaluation model was appropriate."

Table 12. Teachers: *Change* Construct—Detailed Report/Responses.

	Strongly Agree (6)	*Agree* (5)	*Slightly Agree* (4)	*Slightly Disagree* (3)	*Disagree* (2)	*Strongly Disagree* (1)	Mean (*M**)	*n*
q1	18	178	72	31	33	6	4.3	328
q2	19	174	71	22	36	6	4.3	328
q3	18	148	76	26	48	12	4.1	328

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5).

Survey Question 7 (q7), elicited an 89% favorable response rate from respondents that agreed with the statement in some form regarding the three-year implementation process of the MTEF. While survey Question 7 had the highest percentage of agreement among teachers and a mean score (4.8), survey Question 6 (q6) showed the least amount of

agreement among respondents with a 60.4% response rate and the lowest mean score of the survey at 3.5 (Table 13).

Table 13. Teachers: *Professional Development* Construct—Detailed Report/Responses.

	Strongly Agree (6)	*Agree* (5)	*Slightly Agree* (4)	*Slightly Disagree* (3)	*Disagree* (2)	*Strongly Disagree* (1)	Mean (*M**)	*n*
q4	30	166	68	27	34	3	4.4	328
q5	14	116	93	37	48	20	3.9	328
q6	8	72	119	40	67	21	3.5	328
q7	83	158	51	8	18	9	4.8	328

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5).

Table 14 is a further breakdown of the overall scale of responses to questions associated with the *instructional improvement* construct. The highest percentage of some form of agreement with a statement from respondents was in Question 12 (q12). The q12 survey question asked if the new teacher evaluation model was able to provide more timely feedback than the "old evaluation system." Eighty-two point three percent (82.3%; 272/328) of teachers agreed with the statement.

Table 15 displays the respondents' survey results for the *reliability* construct. Overall, survey statements associated with the *reliability* construct had the lowest means of the survey with none of the questions scoring a mean above 3.7. No other construct had more than two questions with means below 4.0.

Table 14. Teachers: *Instructional Improvement* Construct—Detailed Report/Responses.

	Strongly Agree (6)	*Agree* (5)	*Slightly Agree* (4)	*Slightly Disagree* (3)	*Disagree* (2)	*Strongly Disagree* (1)	Mean (*M**)	*n*
q8	21	124	102	24	40	17	4.0	328
q9	28	121	91	25	43	20	4.0	328
q10	33	133	93	21	34	14	4.2	328
q11	67	132	57	25	30	17	4.4	328
q12	77	148	45	20	22	16	4.6	328

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5).

Table 15. Teachers: *Reliability* Construct—Detailed Report/Responses.

	Strongly Agree (6)	*Agree* (5)	*Slightly Agree* (4)	*Slightly Disagree* (3)	*Disagree* (2)	*Strongly Disagree* (1)	Mean (*M**)	*n*
q13	9	97	100	41	47	34	3.6	328
q14	10	90	95	52	48	34	3.6	328
q15	9	102	101	48	42	26	3.7	328
q16	9	85	103	52	48	31	3.6	328

* The higher a mean (*M*) number, the stronger the agreement with a statement (5.5 indicates a stronger agreement than 4.5).

To test for internal consistency (Cohen, Manion, & Morrison, 2007**)**, Cronbach's alpha tests were conducted for each construct to ensure participants' responses were consistent within the constructs of *change, professional development, instructional improvement, reliability*, and *overall satisfaction* with ease of use of the MTEF. For the administrator survey, three out of the five constructs in Table 16 indicate an overall "acceptable to high rate" of internal consistency.

Table 16. Administrators: Correlation of Subscale Constructs and Measures of Internal Consistency.

Construct Number	Subscale Constructs	c1.	c2.	c3.	c4.	c5.	α
c1.	Change q1, q2, q3						.90
c2.	Professional Development q4, q5, q6, q7	.73*					.60
c3.	Instructional Improvement q8, q9, q10, q11, q12	.76*	.66*				.86
c4.	Reliability q13, q14, q15, q16	.44	.48	.49			.85
c5.	Overall Satisfaction q17, q18, q19	.57*	.55*	.51*	.61*		.65

*Correlation is significant at the .05 level.

Professional development (.60) and *overall satisfaction* (.65) with the MTEF instrument were a bit low. In addition, Table 16 shows the correlation of subscales that were used to study relationships between constructs. Four out the five constructs in Table 16 show pairwise positive correlations between the constructs, which include *change*, *professional development*, *instructional improvement*, and *overall satisfaction* with ease of use of the MTEF instrument. The one construct that failed to meet the .05 level of significance is the *reliability* construct. Based on available research, we might expect to see correlation between constructs due to the fact that each one supports the other and all are necessary for a successful implementation of a model (Finkelstein, 2016; Fullan, 2011; Minnici, 2014; Tucker, 2011).

In the teacher survey, the five constructs show pairwise positive correlations between all constructs (Table 17).

Table 17. Teachers: Correlation of Subscale Constructs and Measures of Internal Consistency.

Construct Number	Subscale Constructs	c1.	c2.	c3.	c4.	c5.	α
c1.	Change q1, q2, q3						.92
c2.	Professional Development q4, q5, q6, q7	.70*					.82
c3.	Instructional Improvement q8, q9, q10, q11, q12	.75*	.81*				.93
c4.	Reliability q13, q14, q15, q16	.58*	.74*	.72*			.96
c5.	Overall Satisfaction q17, q18, q19	.72*	.76*	.76*	.67*		.89

*Correlation is significant at the .05 level.

Correlation of subscales was used to study the relationships between constructs. As shown in Table 17, all correlations are significant at the .05 level. In addition to correlation of subscales, Table 17 shows the Cronbach Alpha coefficients for each set of questions in the five constructs.

Research Question 3 examined whether or not there was a difference between the school district's administrators and the district's teachers regarding the five research constructs that framed this study. Constructs examined included: *change, professional development, instructional improvement, reliability,* and *overall satisfaction* with ease of use of the MTEF model. The purpose of this question was to identify whether or not there existed any differences in perceptions between administrators and teachers regarding the implementation of the MTEF, and its ability to improve instruction. Table 18 shows results from independent *t*-tests comparing mean scores for both administrators and teachers.

Table 18. Comparison Between Administrators and Teachers (*strongly disagree* = 1, *strongly agree* = 6).

Constructs	*M*	*SD*	*t*	*df*	*p*	*d*
Change			3.96	360	.001*	0.78
Administration	5.0	.95				
Teachers	4.2	1.11				
Professional Development			4.68	34.52	.001*	0.78
Administration	4.8	.67				
Teachers	4.1	1.01				
Instructional Improvement			9.10	48.50	.001*	1.21
Administration	5.5	.53				
Teachers	4.3	1.19				
Reliability			2.35	352	.019*	0.51
Administration	4.2	1.02				
Teachers	3.7	1.26				
Overall Satisfaction With Model's Ease of Use			5.53	41.03	.001*	.81
Administration	4.9	.60				
Teachers	4.2	1.14				

* $p < .05$

The researcher used SPSS software to conduct independent sample *t* tests to determine if there was statistical significance between teachers' and administrators' perceptions of the MTEF implementation process regarding the five dependent variables of *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with the MTEF model's ease of use. Table 18 shows independent *t*-test results for the administrators and teachers for each of the constructs. Differences were statistically significant for all five constructs at the $p < .05$ level, with administrators having a higher mean score in all of the constructs. The *instructional improvement* construct showed the

largest effect size for statistically significant factors with administrators (M = 5.36, SD = .53) and teachers (M = 4.25, SD = 1.19); t (48.50) = 9.10, p = .001, $p < .05$, d = 1.21.

Summary

Chapter IV presented the qualitative data that was used to answer the research questions of the study regarding administrators' and teachers' perceptions in reference to the implementation of the MTEF. Quantitatively, descriptive analysis of items, reliability testing of subscales, internal consistency, and independent sample t tests were conducted to assess participants' perceptions. The results of the study showed there to be a statistical significance in each of the five constructs between teachers and administrators perceptions.

Chapter V presents a summary of the study, conclusions drawn from the results, limitations or the student, and recommendations for future research.

CHAPTER V

SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Chapter V is divided into four sections: summaries of results, conclusions, limitations, and recommendations for further study. The findings from this study were based on a literature review, quantitative data analysis, and the researcher's knowledge and background as it relates to implementation of the Marzano Teacher Evaluation Framework model.

Summary

The purpose of this study was to test the effectiveness of implementation of the Marzano Teacher Evaluation Framework (MTEF) in a Midwest school district. The process was two-fold: first, to examine perceptions of teachers as they relate to the implementation of the model, specifically looking at *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with the model's ease of use; and second, to compare the perceptions of administrators using the same questions and constructs. Quantitative data was collected and analyzed to examine perceptions of teachers and administrators as the school district moved to bring about change in the way teacher evaluation was conducted throughout the school district.

When this study began in 2013, there was a significant amount of literature on what an effective teacher evaluation should consist of, but very little information about how to

implement a new model and bring about change to improve instruction, which in turn should bring about an increase in student achievement. Results from this study could have a considerable impact on how school districts implement new teacher evaluation models to bring about positive change and improve teachers' instructional practices.

The surveys for teachers and administrators contained three demographic questions that identified teachers' and administrators' years of experience, school level employment (elementary, middle school, high school), and education level (bachelor, master, or doctorate degree). The next section of the survey included five "yes" or "no" questions regarding each of the individual constructs: *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with the model's ease of use. Following the five construct questions there were 19 questions, based on a six-point Likert-type scale (*strongly agree*, *agree*, *slightly agree*, *slightly disagree*, *disagree*, *strongly disagree*) that assessed teacher perceptions of the five constructs. The researcher analyzed quantitative data generated by the teachers' and administrators' responses to examine whether or not there were differences in their perceptions of the MTEF implementation, and whether or not they believed in its ability to improve instruction.

Conclusions With Discussion

For the discussion purposes of this study, the three research questions will be combined and discussed by comparing results of this study; to the literature based on the constructs of *change*, *professional development*, *instructional improvement*, *reliability*, and *overall satisfaction* with ease of use of the MTEF instrument as perceived by administrators and teachers.

Research Questions

Research questions are given for the readers review.

1. What are the perceptions of one school district's administrators regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?
2. What are the perceptions of one school district's teachers regarding the implementation of the Marzano Teacher Evaluation Framework and supervision process and its ability to improve instructional practices?
3. Was there a difference between the school district's administrators and teachers regarding the five research constructs that frame this study, which include *change, professional development, instructional improvement, reliability*, and *overall satisfaction* with the new Marzano Teacher Evaluation Framework model?

Change

Findings in this study indicated that administrators viewed implementation of the new teacher evaluation model and its ability to improve instruction more favorably than did teachers. An independent *t*-test showed there to be statistically significant differences in administrators' and teachers' perceptions of the MTEF implementation regarding the *change* construct.

The first general question on the survey asked administrators and teachers a "yes" or "no" question as to whether or not they believed the school district needed a new teacher evaluation model. Of the administrators surveyed, 100% believed there was a need to adopt a new teacher evaluation model, compared to 71.5% of teachers surveyed. After the general

"yes" or "no" question, respondents answered three questions specific to the *change* construct using a six-point Likert-type scale. Responses to questions about the *change* construct were distributed with a $M = 5.0$ and $SD = .95$ for administrators, compared to teachers' average scores of $M = 4.2$ and $SD = 1.11$. The three survey questions/statements that addressed the *change* construct had good internal consistency for both administrators ($\alpha = .90$) and teachers ($\alpha = .92$).

Based on findings and the statistically significant difference in perceptions between administrators ($M = 5.0$, $SD = .95$) and teachers ($M = 4.2$, $SD = 1.11$) regarding *change* ($t(360) = 3.96$, $p = .001$, $p < .05$, $d = .78$), a closer examination of the possible reasons for this significant difference in perceptions will analyzed later in this section. One possible reason might be administrators were all involved with the change process from the start and teachers were not. On February 14, 2011, the school district engaged in a strategic planning activity that included all district-level administrators and building-level administrators, a small sampling of lead teachers, and a variety of community stakeholders, including parents, business people, and university officials. During this meeting, a priority area, "Promote practices which attract and retain high quality staff," emerged as an area of emphasis and included a goal to, "Develop a staff evaluation model that promotes effectiveness" ([Strategic plan of participating school district – Name of school left off to preserve confidentiality], 2011, p. 10). In August 2011, the CIAPD Department formed a committee consisting of three district-level administrators, five building-level administrators, and ten teachers to carry out the goal of developing a staff evaluation model to promote effectiveness. They were to implement the following strategies:

- Review the current research on evaluation models.
- Develop a plan for collaborative involvement of stakeholders in model development and implementation.
- Provide professional development for staff of the school district.

([Teacher handbook of participating school district], 2015, p. 3)

During review of evaluation models currently being used at that time, the committee studied Charlotte Danielson's and Robert Marzano's teacher evaluation models. In the Spring of 2012, the MTEF was selected by the Teacher Evaluation Committee as a good model to implement ([Teacher handbook of participating school district], 2015).

As Fullan (2011) pointed out in his theory of change, one of the key drivers to system-wide reform is building capacity of teachers and empowering them as the central driving force behind change. Successful change revolves around the idea of trusting teachers and engaging them in the process of professional discourse regarding problems and potential solutions (Fullan, 2011; OECD, 2011). Unfortunately, in the United States, educational change initiatives have oftentimes overlooked the importance of building capacity and empowering teachers to be part of a change process, which is why so many initiatives have failed over the years (Fullan, Cuttress, & Kilcher, 2005; Fullan, 2001; Fullan, 2011). Instead, changes oftentimes come from the top down (politicians, national- and state-level leaders, and district- or building-level leaders) instead of the bottom-up approach (teacher to principal to superintendent; Fullan, 2011). This tendency for implementing change in this manner is evident in the survey results**.** The question, "The process used by the Teacher Evaluation Committee to update and inform stakeholders on the new model was effective in gaining

support for the initiative" had the lowest percentage of some form of agreement among teachers with 74.1% agreement, the lowest mean (M = 4.1), and the highest standard deviation (SD = 1.28). The same question for administrators also scored the lowest of all constructs with 88.4% of respondents showing some form of agreement, the lowest mean (M = 4.8), and the second highest standard deviation (SD = 1.05).

Implementing substantial, lasting change is a time-consuming process and is not something that can be done quickly (Fullan, 2011). Research by Fullan et al. (2005), identified building capacity of all stakeholders affected by change as a key component to successfully implementing a change initiative. Without a necessary solid foundation, stakeholders lack motivation and the necessary skills to successfully implement change (Fullan, 2011). All teachers must be engaged in a change process in order to feel as though they are part of the change, instead of feeling as though a change is something being done to them.

Professional Development

An independent *t*-test showed there to be a statistically significant difference in administrators' and teachers' perceptions on the *professional development* construct regarding the implementation of the MTEF. The first general question about the *professional development* construct asked administrators and teachers a "yes" or "no" question as to whether or not they were satisfied with the professional development they had been provided by the school district in regards to the implementation of the MTEF. Of the administrators surveyed, 100% were satisfied with the professional development that was conducted during the implementation of the MTEF at both the district and the building level, compared to 71.5% of teachers surveyed. When the four statements from the online survey addressing the

professional development construct were combined in a single construct score, administrators (M = 4.8, SD = .67) outscored teachers (M = 4.1, SD = 1.01).

Based on findings and the statistically significant difference in perceptions between administrators (M = 4.8, SD = .67) and teachers (M = 4.1, SD = 1.01) regarding professional development ($t(34.52) = 4.86$, $p = .001$, $p < .05$, $d = .78$), a closer examination of the possible reasons for this statistical difference in perceptions will be analyzed later in this section. The four statements addressing the *professional development* construct showed the widest range of "percentage of some form of agreement" for administrators and teachers. Administrators' scores ranged from 100% of some form of agreement on Question 4 (q4), to a low of 69.3% of some form of agreement on Question 6 (q6), while teachers' scores ranged from a high of 89% of some form of agreement on Question 7 (q7), to a low of 60.4% of some form of agreement on Question 6 (q6).

Question 5 (q5): "To date, the professional development of the MTEF has met my needs as a teacher/administrator" had the largest percentage of disagreement between teachers and administrators in the *professional development* construct. Among teachers, 68.3% agreed with the statement, with M = 3.9, and SD = 1.33. The same item had 92.3% of some form of agreement with the statement among administrators, with M = 5.1, and SD =.89. Another question that had a sizeable difference in the "percentage of some form of agreement" was Question 4 (q4), "The use of staff meetings to communicate the phases of implementation was beneficial." Among teachers, there was 82% of some form of agreement with the statement, with M = 4.4, and SD = 1.15. The same question had 100% of some form of agreement with the statement among administrators, with M = 5.0, and SD = .87.

The two questions that showed the greatest difference in agreement between teachers and administrators require a closer look. The method used to explain the process of the MTEF implementation was conducted through the use of staff meetings at each of the school district's 18 buildings. Each of the presentations was created by the school districts' teacher evaluation committee for consistency purposes and sent to building administrators to present to their staffs. The presentations followed a PowerPoint format and took approximately 45 minutes to share with teachers. While administrators favored this approach, teachers' perceptions were not as positive. According to the standards developed by Learning Forward (2011) and the research conducted by Stewart (2011) and Tucker (2011), professional development is effective when it is teacher lead, collaborative, and closely aligned to their day-to-day instructional practices. While the consistency was in place across the district and all teachers were provided the same information, the process may have lacked teacher engagement and felt like a top-down approach, creating apathy and passive learning by teachers (Learning Forward, 2011; Sergiovanni & Starratt, 1971; Tucker, 2014).

Instructional Improvement

The third construct, *instructional improvement*, showed a statistically significant difference between administrators' and teachers' perceptions of the MTEF ability to improve teacher's instructional skills. The first general question on the *instructional improvement* construct asked administrators and teachers a "yes" or "no" question as to whether or not they believed the MTEF would have a positive effect on teaching performance. Of the administrators surveyed, 96.2% believed the new teacher evaluation model would have a positive effect on teaching performance, compared to 72.1% of teachers surveyed. According

to the quantitative data yielded from the survey, a statistically significant difference (p = .001) was found between administrators and teachers.

Responses to survey statements addressing the *instructional improvement* construct showed the most significant difference between administrator and teacher perceptions. In each of the five statements, administrators had a 100% "form of agreement," while teachers' responses ranged between a high of 82.3% and a low of 73.5%. After averaging mean scores of the five statements addressing *instructional improvement*, the overall mean score for administrators was M = 5.36 and the teachers overall mean score was M = 4.25, for a difference of 1.11, the highest of all constructs.

Based on findings and the statistically significant difference in perceptions of administrators (M = 5.4, SD = .53) and teachers (M = 4.3, SD = 1.91) regarding instructional improvement ($t(48.50) = 9.10$, $p = .001$, $p < .05$, $d = 1.21$), a closer examination of the possible reasons for this statistical difference will be analyzed later in this section. Question 9 (q9), is at the heart of this study. For teachers, q9 stated, "The feedback from the Marzano Teacher Evaluation Framework is an effective tool in helping me to improve my performance." For administrators, q9 stated, "The feedback from the Marzano Teacher Evaluation Framework is an effective tool in helping me to improve my teachers' instructional performance." Administrators agreed 100% with the statement, while 73.5% of the teachers agreed with the statement. The difference of 26.5% is the largest disparity between teachers and administrators for any of the questions.

Based on research, instructional frameworks have provided teachers with strategies that help guide their instruction using highly effective research-based strategies proven to increase students' achievement (Cherasaro et al., 2016; Darling-Hammond, 2013; Marzano

& Pickering, 2003). While administrators support the idea that instructional frameworks can help guide and improve teachers' classroom performance, there is a significant difference between administrators' perceptions and teachers' beliefs. Though, instructional improvement has been viewed as a separate construct in this study, one needs to examine this construct in the light of research on change and professional development since they go hand in hand with improving teachers' instructional skills. According to City et al. (2009), the greatest barriers teachers face when looking to improve student achievement is an agreed upon definition of quality instruction.

While an instructional framework such as the MTEF does provide teachers with effective research-based teaching strategies, in this study, the framework was decided on with input from eight teachers, which represents .012 percent (8/682) of the district's teachers**.** Based on Fullan's (2011) research, any new initiative in education must start with teachers and students as the driving force. When teachers feel part of a team and have a major voice in the change process, they are more likely to become intrinsically motivated and driven to make change happen (Darling-Hammond, 2013; Fullan, 2011; Sutton & Shouse, 2016). Moving initiatives forward requires stakeholders to work together collectively and purposefully towards a common goal to enact change (Finkelstein, 2016, Fullan, 2011, Minnici, 2014).

According to the literature, the design of the MTEF meets guidelines of a highly effective teacher evaluation model that has the ability to improve teachers' instructional strategies. However, the difference in beliefs of teachers and administrators on the model's ability to improve instruction is significant. Based on the implementation method used by the school district in this study, and on research on change theory, effective professional

development, and instructional improvement, it is possible that more time spent building all teachers' capacity as it relates to creating a common understanding of "what is quality instruction" would have been a good starting point for improving instruction, prior to implementing the MTEF. Using a backward design approach (Wiggins & McTighe, 2005) during planning would have given teachers a substantial voice in the change process, empowering them and increasing the likelihood of a smooth transition to the MTEF. Combining collaboration, intrinsic motivation, and instructional improvement are essential elements needed in whole system reform (Fullan, 2011).

Reliability

Responses to survey statements about the *reliability* construct showed a statistically significant difference between administrators' and teachers' perceptions. The first general question addressing the *reliability* construct asked administrators and teachers a "yes" or "no" question as to whether or not they believed the MTEF was a reliable instrument for evaluating teacher effectiveness. Of the administrators surveyed, 96.2% believed the new MTEF teacher evaluation model would lead to reliable scoring of teachers, compared to 61.8% of the teachers. The *reliability* construct showed the lowest "percentage of agreements" among teachers. This variable held strong for internal consistency with both the teachers (a = .85) and the administrators (a = .96) meeting the .70 or higher level considered "acceptable" for internal consistency.

Survey statements addressing the *reliability* construct had the lowest overall mean scores and largest standard deviations of all the constructs for both administrators and teachers. A statistically significant difference was found between administrators' (M = 4.2, SD = 1.02) and teachers' (M = 3.7, SD = 1.02) perceptions regarding reliability (t(352) =

2.35, $p = .019$, $p < .05$, $d = .51$) of the MTEF. Based on the data, a closer examination for possible reasons for this significant difference should be analyzed.

Questions addressing the *reliability* construct centered around the idea of consistency of scoring teachers' instructional performance. In other words, when using the MTEF to evaluate teacher performance, would all teachers be evaluated fairly and scored consistently regardless of which administrator would be completing an evaluation. Inter-rater reliability is an essential component of any teacher evaluation model. Teachers need to know the principle of equity is being applied throughout a school and district when administrators are scoring and providing feedback. Being able to evaluate teachers correctly and uniformly is essential for an instrument, process, or method to be reliable (Danielson, 1996; Derrington & Campbell, 2015; Donaldson & Donaldson, 2012, McGuinn, 2015). Experts agree, on-going training, consultation, and practice for administrators are essential to ensure consistent and reliable feedback for teachers regarding performance (Cherasaro et al., 2016; Danielson, 2012; Darling-Hammond, 2014; Muijs, 2006). Failure to include these elements in any teacher evaluation model can allow bias and preconceived expectations to influence an administrator's evaluation, either positively or negatively (Cherasaro et al., 2016; Danielson, 2010; Muijs, 2006). According to Donaldson and Donaldson (2012), being able to correctly identify instructional practices is essential for reliability of a teacher evaluation process.

Experts in the field of teacher evaluation clearly identify reliability as an essential component of a successful model. According to results from this study, reliability is clearly an area that showed a discrepancy between teachers' and administrators' responses and mean scores were low within each of the statement groupings relative to other constructs in the study. While the school district implemented the MTEF using an expert's advice on training,

practice, and consultation with administrators, clearly, somewhere in the process there was a disconnect. One possible explanation might include lack of communication between district-level administration and principals regarding the extensive training that building-level administrators were engaged in to ensure reliability and consistency of scoring. While administrators participated in trainings with national experts and engaged in instructional rounds with colleagues and trainers, this information was not formally shared with teachers.

As previously stated, understanding the MTEF and applying it correctly to teachers' instructional performance is essential for reliability, but so is skillfully providing teachers with feedback regarding their instructional practice. Administrators that are not able to build trust and be viewed as a resource will struggle with teachers accepting their feedback. In order for evaluations to be productive and have a positive effect on teachers' performances, teachers must be willing to engage in productive conversations about their performance (Danielson, 2010; Ritter & Barnett, 2016). In research conducted by Tuytens and Devos (2011), administrators who used active leadership supervision, were charismatic, and knowledgeable about content were able to positively engage in discussion with teachers about their performance with their feedback more likely to be viewed as reliable by teachers.

Overall Satisfaction With Ease of Use of the MTEF

Four statements on the survey addressed the *overall satisfaction* construct, and three items were specific to teachers and administrators' perceptions of the MTEF model's ease of use, ability to raise student achievement, and clarity of standards. First, a general question on the *overall satisfaction* construct was a "yes" or "no" question as to whether or not respondents were satisfied with the MTEF instrument and its features. Of the administrators

surveyed, 96.2% were satisfied with the overall features of the framework, compared to 66.1% of the teachers.

A statistically significant difference was found between administrators (M = 4.9, SD = .60) and teachers (M = 4.2, SD = 1.14) regarding the *overall satisfaction* construct (t(41.03) = 5.53, p = .001, p < .05, d = .81). The question that presented the highest discrepancy between teachers' and administrators' responses was Question 17 (q17), "Overall, the new Marzano Teacher Evaluation Framework is relatively easy to use." For administrators, responses showed 100% of "some form of agreement," with q17 compared to teachers' responses showing 73.8% of "some form of agreement," for a difference of 26.2%. Tornero and Taut (2010) concluded that there are three reasons why teachers may have a negative attitude towards teacher evaluation: lack of experience and knowledge regarding teacher evaluations, questions of competency regarding an evaluator, and the extra work new evaluation systems may create. Prior to the MTEF, teachers and administrators in the school district used a subjective evaluation tool. It was an open-ended narrative of what was taking place in the classroom and was not based on any standards. In addition, teachers were only observed in their classroom once every 3 years. The other 2 years of a 3-year cycle, teachers would meet with their supervising administrator to review teachers' progress towards goals they had developed at the start of the year. In addition, no pre-conference, informal observations, or walk-throughs were required of teachers. With the increase in supervision and a model that defines effective teaching standards, it is possible that teachers may feel they are losing their autonomy and their professional judgment is being questioned, which is consistent with findings of Tornero and Taut (2010).

Critical Analysis

Through my experiences as a classroom teacher, principal, doctoral student, and member of district and state-level committees working with teacher evaluation models, I have come to the conclusion that teacher evaluation models are a small piece of a big puzzle when it comes to improving teachers' skills and students' learning. As I reflect on statistical findings of this study, my review of the literature, and my experiences as an educational leader, it becomes clear that any major change in schools must start from the ground level with teachers being the driving force, and then move up. Applying Fullan's (2011) theoretical framework to this study highlights the importance of starting a change process with teachers through capacity building, increasing collaboration, and improving pedagogy, and linking all of them together to create a systemic change.

While the purpose of this study was to test the implementation of the MTEF, throughout the process, I found myself asking, "If there was a chance to do this implementation over again, what recommendations would I make to the committee?" Based on findings of this study, I would first recommend creating a vision and building the process backwards from that point. An understanding of stakeholders' beliefs about the educational experiences we want our students to have is an essential starting point, creating a common goal that everyone is working towards purposefully. Based on the vision, I would focus on building teachers' capacity, providing the necessary resources and support stakeholders require to move towards reaching their vision. As Fullan (2011), Firestone (2014), and Derrington and Campbell (2015) pointed out, this is the most often overlooked step. It is a process that takes time and cannot be rushed. Once the necessary support systems were in place and the vision was clearly articulated and understood by all stakeholders, the process of

selecting a teacher evaluation framework that aligned with the vision could begin. It is imperative that all stakeholders have a voice in the selection of a teacher evaluation model, are aware of what it measures, and know how it works to help teachers improve their instructional practice. Communication, transparency, and trust building are all essential components of a successful implementation. Throughout this process, it is crucial that teachers are treated as professionals and their individual strengths and opportunities for growth are taken into consideration, especially during the professional development stage of implementation. The bottom line is professional development should be led from the bottom-up, not the top down and closely aligned to teachers' daily instructional practices (Diaz-Maggioli, 2004; Mizell et al., 2011; Tucker, 2014). The one-size-fits-all approach does not work.

Envision a school district where students, teachers, and school and district leaders are all working together towards a common vision of preparing students for their future through the use of highly effective instructional practices. Creating that type of environment takes time, effort, resources, communication, and vision. As Fullan (2011) pointed out with his Drivers of Whole System Reform, building teachers capacity, supporting their collaboration, focusing on effective pedagogy, and tying it together in a systemic, district-wide approach can provide school districts with the necessary framework for successful, long-term education change with teachers as the driving force.

Limitations

Results of this study provide insight into the complex issue of implementing a new teacher evaluation model. Nonetheless, findings of this study must be evaluated in light of the limitations that existed at the time of the study.

The first limitation of the study is its cross-sectional design as opposed to a longitudinal design. Cross-sectional studies are designed to provide a snapshot of participants current attitudes or perceptions regarding a topic (Gay, Mills, & Airasian, 2009). A single-point-in-time snapshot limits a researcher's ability to make an informed, reliable decision about changes that should be made in an implementation process. A longitudinal study conducted throughout a 5-year implementation process would potentially provide better data and results.

Secondly, the study included a small sample of administrators and teachers from one school district in the Midwest. Survey participants were members of a specific group and are not representative of a broad cross-section of administrators or teachers, due to the fact that the research took place in one school district. As a result, study findings are limited in their generalizability, and there is the possibility of sample bias.

Implications for Practice

The most significant findings in this study are the differences in perceptions of teachers and administrators in each of the five constructs. As a result, findings of this study should be a call for administrators at both the district and school level to examine steps required to implement wide-scale change. Implementing sweeping change is a challenging task, and results of this study indicate there are ways to proceed that ensure teachers will view a process positively.

Results of this study and of research found in the literature that supports these findings showed teacher involvement is crucial for successful implementation of any type of change (Learning Forward, 2011; Stewart, 2011; Tucker, 2011). Teachers involvement, leadership, and voice are essential in a large-scale implementation of a change, allowing

teachers to feel as though they are part of the process, instead of feeling as though it is something that is being done to them (Fullan et al., 2005; Learning Forward, 2011; Sergiovanni & Starratt, 1971). The recommendation section that follows provides specific ideas for implementing district-wide change based on research and findings of this study.

Recommendations

With a goal of implementing a new teacher evaluation model as a way to improve teacher performance and in return improve student achievement, it is essential that teachers and administrators work together to develop a common understanding of what constitutes effective instruction (Watson, Miller, Davis, & Carter, 2010). A well designed evaluation system aligns with teacher beliefs, encourages teachers to be reflective, and creates dialogue between teachers and administrators about effective teaching practices (Taylor & Tyler, 2012). According to Zimmerman and Deckert-Pelton (2003), teachers value constructive feedback, advice and encouragement, and pedagogical feedback in a timely manner. While the administrators surveyed felt satisfied with the MTEF, the teachers did not. As stated in the research literature, teacher involvement in design of a change is essential for buy-in and acceptance. Range, Scherz, Holt, and Young (2011) suggested that change will only happen if teachers have ownership of the change.

The following recommendations emerged after analysis of data from this study and a review of the literature.

1. In order to bring about large-scale change, all stakeholders must be involved in the process, first determining if there is a need for change, and if so, having a voice throughout the process. Ideally, change should come from the bottom up, not the top down.

2. Communication is essential for all stakeholders that are involved in a change process. Keeping stakeholders up to date on progress and decisions being made throughout a process ensures transparency and buy-in.
3. Professional development should be driven by teachers' needs. When professional development is teacher led, collaborative, and closely aligned to teachers' needs, there is a greater chance of success.
4. Prior to any major change, a district should start with a vision of what it wants to create. Empower all teachers to be involved in the process by giving them a voice and creating a common understanding of what the end product will look like by using a backward design approach.
5. Trust is an essential component of teacher evaluation. Without it, constructive and productive conversations regarding teachers' instructional performance will not happen. Teachers must know their administrators are there to support them and provide them with feedback that will improve their instructional performance.

Recommendations for Additional Research

Research presented in this study sets a broad foundation for understanding of and implementing a new teacher evaluation model. Upon completion of the study survey and examination of the results, it was evident there are additional areas that would be appropriate for further study. Recommendations for additional research include:

1. Conduct a longitudinal study throughout an implementation process to assess the full effects of a teacher evaluation model. A change process can take up to 5 years (Fullan et al., 2005). Surveying teachers at the beginning, middle, and end

of an implementation process would provide a researcher with a clearer picture of administrators' and teachers' perceptions of a new teacher evaluation model.

2. Expand the research across geographic areas of the country using a random stratified sample of teachers and administrators would be beneficial.
3. Conduct a qualitative study to examine teachers' and administrators' specific perceptions regarding implementation of a new teacher evaluation model. This type of study might identify specific details that teachers and administrators perceive as positive or negative regarding an implementation, and what they would like to see changed as a process moves forward.
4. Add an additional construct to research on a teacher evaluation model on *trusting relationships*. This would allow a researcher to examine the relationship between a model's ability to improve instruction and its reliability based on the relationship between an administrator and teacher. The idea that trusting relationships are necessary between teachers and administrators goes as far back as Cubberley (1922) and is still addressed in the literature by Danielson (2010) and Ritter and Barnett (2016).
5. Expand the scope of the study to include a variety of teacher evaluation models (i.e. Marzano, Marshall, Danielson, McREL) comparing administrators' and teachers' perceptions regarding the effectiveness of each model using the constructs from this study. Comparing models would provide a researcher with valuable information regarding the perceived effectiveness of each model.

Conclusion

The purpose of this study was to test administrators' and teachers' perceptions regarding implementation of the MTEF model. Central to the research was examining the process from the beginning stages of change to how the implementation of the new teacher evaluation model affected teachers' ability to improve their instructional practice. In each of five constructs studied, the researcher found administrators' and teachers' perceptions were significantly different.

Based on results of this study, current research at the time of this study, and the researcher's educational experience, it was evident that implementation of a new teacher evaluation model is an extremely challenging process. It is the hope of this researcher that this study will provide school districts with valuable data and information as they look to make changes in their teacher evaluation models over years to come.

APPENDICES

Appendix A
Administrator's Survey

Marzano Teacher Evaluation--Administration

Please take a moment to complete the survey below. The purpose of this survey is to analyze administrators' perceptions regarding the implementation of the Marzano Teacher Evaluation Framework process.

Current Level of Administration	Years of Administration Experience
___ Elementary School ___ Middle School ___ High School Highest Degree Earned ___ Master's Degree ___ Specialist's Degree ___ Doctorate Degree	___ 0-5 ___ 6-10 ___ 11-25 ___ 15-above

c1.	I believe there was a need for the newly adopted Marzano Teacher Evaluation Framework in the [redacted] Public School District.	**Y**	**N**
c2.	I have been satisfied with the professional development that has been conducted during the implementation of the Marzano Teacher Evaluation Framework?	**Y**	**N**
c3.	I believe the Marzano Teacher Evaluation Framework will have a positive effect on my ability to improve my teachers' performance.	**Y**	**N**
c4.	I believe the Marzano Teacher Evaluation Framework is a reliable instrument for evaluating teacher effectiveness.	**Y**	**N**
c5.	Overall, I am satisfied with the Marzano Teacher Evaluation Framework instrument and its features.	**Y**	**N**

Please think of the following questions in the context of the new teacher evaluation model. Rate each of the questions to the best of your ability.		**Strongly Disagree**	**Disagree**	**Slightly Disagree**	**Slightly Agree**	**Agree**	**Strongly Agree**
q1.	The process used by the [redacted] Public School District to determine the need for a new teacher evaluation model was appropriate.	**1**	**2**	**3**	**4**	**5**	**6**
q2.	The process used by the Teacher Evaluation Committee to determine the new teacher evaluation model for the [redacted] School District was appropriate.	**1**	**2**	**3**	**4**	**5**	**6**
q3.	The process used by the Teacher Evaluation Committee to update and inform stakeholders on the new model was effective in gaining support for the initiative.	**1**	**2**	**3**	**4**	**5**	**6**
q4.	The use of staff meetings to communicate the phases of implementation was beneficial.	**1**	**2**	**3**	**4**	**5**	**6**
q5.	To date, the professional development on the Marzano Teacher Evaluation Framework has met my needs as an administrator.	**1**	**2**	**3**	**4**	**5**	**6**
q6.	The video vignettes in iObservation are an effective tool in helping my teachers improve their instruction.	**1**	**2**	**3**	**4**	**5**	**6**
q7.	I am satisfied with the implementation being slowly phased in over a three year time period.	**1**	**2**	**3**	**4**	**5**	**6**
q8.	The Marzano Teacher Evaluation Framework is an effective tool to influence my teachers' development.	**1**	**2**	**3**	**4**	**5**	**6**
q9.	The feedback from the Marzano Teacher Evaluation Framework is an effective tool in helping me to improve my teachers' instructional performance.	**1**	**2**	**3**	**4**	**5**	**6**

q10.	I use the Marzano Teacher Evaluation Framework feedback to help me effectively guide my teachers' performances.	**1**	**2**	**3**	**4**	**5**	**6**
q11.	Compared to the "old evaluation system," observational feedback is more relevant and meaningful to affirm or alter instruction.	**1**	**2**	**3**	**4**	**5**	**6**
q12.	Compared to the "old evaluation system," observational feedback is more immediate.	**1**	**2**	**3**	**4**	**5**	**6**
q13.	I believe the Marzano Teacher Evaluation Framework will result in consistent ratings among teachers.	**1**	**2**	**3**	**4**	**5**	**6**
q14.	I am confident that most administrators' ratings would be similar if they were rating the same teacher while using the Marzano Teacher Evaluation Framework.	**1**	**2**	**3**	**4**	**5**	**6**
q15.	The Marzano Teacher Evaluation Framework's scales scoring system allows for consistent scoring of teachers.	**1**	**2**	**3**	**4**	**5**	**6**
q16.	I am confident in the consistency of the Marzano Teacher Evaluation Framework.	**1**	**2**	**3**	**4**	**5**	**6**
q17.	Overall, the new Marzano Teacher Evaluation Framework is relatively easy to use.	**1**	**2**	**3**	**4**	**5**	**6**
q18.	The teaching standards that are measured by the new Marzano Teacher Evaluation Framework are focused on what is necessary to raise student achievement.	**1**	**2**	**3**	**4**	**5**	**6**
q19.	The rubrics used to measure the teaching standards in our new Marzano Teacher Evaluation Framework are adequately descriptive.	**1**	**2**	**3**	**4**	**5**	**6**

Appendix B
Teacher's Survey

Marzano Teacher Evaluation--Teachers

Please take a moment to complete the survey below. The purpose of this survey is to analyze teachers' perceptions regarding the implementation of the Marzano Teacher Evaluation Framework process.

Current Level of Teaching	What is your current teaching category defined as
___ Elementary School ___ Middle School ___ High School	___ Regular Education Teacher ___ Special Education Teacher ___ Other
Highest Degree Earned	Years of Teaching Experience
___ Bachelor's Degree ___ Master's Degree ___ Doctorate Degree	___ 0-8 ___ 9-16 ___ 17-24 ___ 25-above

c1.	I believe there was a need for the newly adopted Marzano Teacher Evaluation Framework in the [redacted] Public School District.	**Y**	**N**
c2.	I have been satisfied with the professional development that has been conducted during the implementation of the Marzano Teacher Evaluation Framework.	**Y**	**N**
c3.	I believe the Marzano Teacher Evaluation Framework will have a positive effect on my teaching performance.	**Y**	**N**
c4.	I believe the Marzano Teacher Evaluation Framework is a reliable instrument for evaluating teacher effectiveness.	**Y**	**N**
c5.	Overall, I am satisfied with the Marzano Teacher Evaluation Framework instrument and its features?	**Y**	**N**

	Please think of the following questions in the context of the new teacher evaluation model. Rate each of the questions to the best of your ability.	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
q1.	The process used by the [redacted] Public School District to determine the need for a new teacher evaluation model was appropriate.	**1**	**2**	**3**	**4**	**5**	**6**
q2.	The process used by the Teacher Evaluation Committee to determine the new teacher evaluation model for the [redacted] School District was appropriate.	**1**	**2**	**3**	**4**	**5**	**6**
q3.	The process used by the Teacher Evaluation Committee to update and inform stakeholders on the new model was effective in gaining support for the initiative.	**1**	**2**	**3**	**4**	**5**	**6**
q4.	The use of staff meetings to communicate the phases of implementation was beneficial.	**1**	**2**	**3**	**4**	**5**	**6**
q5.	To date, the professional development on the Marzano Teacher Evaluation has met my needs as a teacher.	**1**	**2**	**3**	**4**	**5**	**6**
q6.	The video vignettes in iObservation help improve my instruction.	**1**	**2**	**3**	**4**	**5**	**6**
q7.	I am satisfied with the implementation being slowly phased in over a three year time period.	**1**	**2**	**3**	**4**	**5**	**6**
q8.	The Marzano Teacher Evaluation Framework is an effective tool to influence my development as a teacher.	**1**	**2**	**3**	**4**	**5**	**6**
q9.	The feedback from the Marzano Teacher Evaluation Framework is an effective tool in helping me to improve my performance.	**1**	**2**	**3**	**4**	**5**	**6**

q10.	I use the Marzano Teacher Evaluation Framework feedback to help effectively guide my teaching performances.	**1**	**2**	**3**	**4**	**5**	**6**
q11.	Compared to the "old evaluation system," observational feedback is more relevant and meaningful to affirm or alter instruction.	**1**	**2**	**3**	**4**	**5**	**6**
q12.	Compared to the "old evaluation system," observational feedback is more immediate.	**1**	**2**	**3**	**4**	**5**	**6**
q13.	I believe the Marzano Teacher Evaluation Framework will result in consistent ratings among teachers.	**1**	**2**	**3**	**4**	**5**	**6**
q14.	I am confident that most administrators' ratings would be similar if they were rating the same teacher while using the Marzano Teacher Evaluation Framework.	**1**	**2**	**3**	**4**	**5**	**6**
q15.	The Marzano Teacher Evaluation Framework's scales scoring system allows for consistent scoring of teachers.	**1**	**2**	**3**	**4**	**5**	**6**
q16.	I am confident in the consistency of the Marzano Teacher Evaluation Framework.	**1**	**2**	**3**	**4**	**5**	**6**
q17.	Overall, the new Marzano Teacher Evaluation Framework is relatively easy to use.	**1**	**2**	**3**	**4**	**5**	**6**
q18.	The teaching standards that are measured by the new Marzano Teacher Evaluation Framework are focused on what is necessary to raise student achievement.	**1**	**2**	**3**	**4**	**5**	**6**
q19.	The rubrics used to measure the teaching standards in our new Marzano Teacher Evaluation Framework are adequately descriptive.	**1**	**2**	**3**	**4**	**5**	**6**

Appendix C
IRB Approval for Fall Teacher Study

REPORT OF ACTION: EXEMPT/EXPEDITED REVIEW
University of North Dakota Institutional Review Board

Date: 10/8/2013 **Project Number:** IRB-201310-116

Principal Investigator: Arason, Kristopher

Department: Educational Leadership

Project Title: Implementation of the Marzano Teacher Evaluation Framework in a Rural School District

The above referenced project was reviewed by a designated member for the University's Institutional Review Board on 10/10/2013 and the following action was taken:

☐ Project approved. **Expedited Review** Category No. _____
Next scheduled review must be before: _____

☐ **Copies of the attached consent form with the IRB approval stamp dated _____ must be used in obtaining consent for this study.**

☒ Project approved. **Exempt Review** Category No. 2
This approval is valid until DEC 24 2013 as long as approved procedures are followed. No periodic review scheduled unless so stated in the Remarks Section.

☐ **Copies of the attached consent form with the IRB approval stamp dated N/A must be used in obtaining consent for this study.**

☐ Minor modifications required. The required corrections/additions must be submitted to RDC for review and approval. **This study may NOT be started UNTIL final IRB approval has been received.**

☐ Project approval **deferred. This study may not be started until final IRB approval has been received.**
(See Remarks Section for further information.)

☐ Disapproved claim of exemption. This project requires Expedited or Full Board review. The Human Subjects Review Form must be filed out and submitted to the IRB for review.

☐ Proposed project is not human subjects research as defined under Federal regulations 45 CFR 46 or 21 CFR 50 and does not require IRB review.

☐ Not Research ☐ Not Human Subject

PLEASE NOTE: Requested revisions for student proposals MUST include adviser's signature. All revisions MUST be highlighted and submitted to the IRB within 90 days of the above review date.

☒ Education Requirements Completed. (Project cannot be started until IRB education requirements are met.)

cc: Dr. Steve Lemire

Michelle L. Baker 10/10/2013
Signature of Designated IRB Member Date
UND's Institutional Review Board

If the proposed project (clinical medical) is to be part of a research activity funded by a Federal Agency, a special assurance statement or a completed 310 Form may be required. Contact RDC to obtain the required documents.

(Revised 10/2006)

Appendix D
School District Approval – Teacher Survey

Request to Conduct Research in the [redacted] Public Schools

Date: 9/30/2013	**Name:** Kristopher G. Arason	**Phone:** (701) 215-2591
Fax or Email: kris.arason@gfschools.org **Address:** 549 Mighty Acres Drive Grand Forks, ND 58201		**Research Advisor:** Dr. Steven Lemire **College or Dept.:** Educational Leadership

Research Title: A Rural Districts Implementation Process of a New Teacher Evaluation Model

Give a brief description of your research. Attach additional papers if necessary. Please attach sample copies of assessment instrument, tests, or communications to be used:

The purpose of this study will be to collect and analyze quantitative data regarding the Marzano Teacher Evaluation Framework. I am interested in examining the implementation of the new evaluation model and teachers' perceptions regarding its effectiveness. More specifically, I will examine the need for change, professional development, instructional improvement, and the evaluation model's effectiveness and ease of use.

Number of students needed for research: 0	**Number of teachers needed for research:** I would be sending out the survey to all teachers in the [redacted] Public School District	**Grade Level or Dept.:** Elementary, Middle, and High School

What schools are you interested in conducting the research in? All of the schools in the [redacted] Public School District will be used.

Will confidential records be required? (If yes, indicate type.) **No**	**Length of time required to complete the research:** The survey will take teachers approximately 10 minutes

To be completed by School District Official:

Approved: [redacted]	
Assistant Superintenden	**Date:** 10-4-13
Approved to conduct res	Those listed

Appendix E
Email to Teachers Asking Them to Participate in the Study

From: Kris Arason October 21, 2013 9:04:06 AM

Subject: Marzano Teacher Evaluation Survey

To: Elementary Principals [redacted]

Principals,

First of all, thank you so much for your help in distributing this survey. Below is the email and the link to the survey that I am asking you send to your teaching staff regarding the Marzano Teacher Evaluation Framework. If you would like to put a sentence or two at the top of your email introducing the survey that would be great. If you just want to forward this email and erase the writing in blue that would be fine. If you have any questions before you send the email, please let me know. Thanks again for you help.

Kris

Greetings [redacted] Teachers,

My name is Kris Arason. I am a graduate student at the University of North Dakota in the Department of Educational Leadership. I am inviting you to participate in a reseach project on the implementation of the Marzano Teacher Evaluation Framework (MTEF) in the [redacted] Public School District.

As the school district continues to move forward with the MTEF, I am interested in teachers' insight. The purpose of this study is to examine the implementation process to date and look for ways to improve the process going forward. For this research project, I am will be examining the implementation of the model and teachers' perceptions of the model's ability to improve instruction, its reliability, and ease of use. Below is a link to a short survey that will give valuable feedback regarding the MTEF.

The survey will take approximately five minutes to complete and is competely anonymous. Participation in this study is voluntary and you may refuse to participate without consequences. Response to the survey will only be reported in aggregate form to protect the identity of the respondents.

http://www.surveymonkey.com/s/TMZHSGF

Thank you for your consideration. Your time and help is greatly appreciated.

Regards,

Kris Arason

Mr. Kris Arason, Principal
Red River High School, 2211 17th Avenue South, Grand Forks, ND 58201
Office Phone: 701-746-2205 Ext. 6800 Office Fax: 701-746-2406
kris.arason@gfschools.org

Appendix F
IRB Approval for Spring Administrator Study

REPORT OF ACTION: EXEMPT/EXPEDITED REVIEW
University of North Dakota Institutional Review Board

Date: 5/28/2014 **Project Number:** IRB-201405-472

Principal Investigator: Arason, Kristopher

Department: Educational Leadership

Project Title: Implementation of the Marzano Teacher Evaluation Framework in a Rural School District: 2014

The above referenced project was reviewed by a designated member for the University's Institutional Review Board on 5/28/2014 and the following action was taken:

☐ Project approved. **Expedited Review** Category No. _____
Next scheduled review must be before: _____

☐ **Copies of the attached consent form with the IRB approval stamp dated _____ must be used in obtaining consent for this study.**

☒ Project approved. **Exempt Review** Category No. 2 + 4
This approval is valid until DEC 24 2014 as long as approved procedures are followed. No periodic review scheduled unless so stated in the Remarks Section.

☐ **Copies of the attached consent form with the IRB approval stamp dated N/A must be used in obtaining consent for this study.**

☐ Minor modifications required. The required corrections/additions must be submitted to RDC for review and approval. **This study may NOT be started UNTIL final IRB approval has been received.**

☐ Project approval **deferred. This study may not be started until final IRB approval has been received.**
(See Remarks Section for further information.)

☐ Disapproved claim of exemption. This project requires Expedited or Full Board review. The Human Subjects Review Form must be filled out and submitted to the IRB for review.

☐ Proposed project is not human subjects research as defined under Federal regulations 45 CFR 46 or 21 CFR 50 and does not require IRB review.

☐ Not Research ☐ Not Human Subject

PLEASE NOTE: Requested revisions for student proposals MUST include adviser's signature. All revisions MUST be highlighted and submitted to the IRB within 90 days of the above review date.

☒ Education Requirements Completed. (Project cannot be started until IRB education requirements are met.)

cc: Dr.Sherryl Houdek

[signature] 5/28/2014
Signature of Designated IRB Member Date
UND's Institutional Review Board

If the proposed project (clinical medical) is to be part of a research activity funded by a Federal Agency, a special assurance statement or a completed 310 Form may be required. Contact RDC to obtain the required documents.

(Revised 10/2006)

Appendix G
School District Approval – Administrator Survey

Request to Conduct Research in the [redacted] Public Schools

Date: 3/24/2014	**Name:** Kristopher G. Arason	**Phone:** (701) 215-2591
Fax or Email: kris.arason@gfschools.org		**Research Advisor:** Dr. Sherry Houdek
Address: 549 Mighty Acres Drive Grand Forks, ND 58201		**College or Dept.:** Educational Leadership
Research Title: A Rural Districts Implementation Process of a New Teacher Evaluation Model		
Give a brief description of your research. Attach additional papers if necessary. Please attach sample copies of assessment instrument, tests, or communications to be used: The purpose of this study will be to collect and analyze quantitative data regarding the Marzano Teacher Evaluation Framework. I am interested in examining the implementation of the new evaluation model and administrators' perceptions regarding its effectiveness. More specifically, I will examine the need for change, professional development, instructional improvement, and the evaluation model's effectiveness and ease of use.		
Number of students needed for research: 0	**Number of admininstrators needed for research:** All principals, associate principals, and building resource coordinators	**Grade Level or Dept.:** Elementary, Middle, and High School
What schools are you interested in conducting the research in? All of the schools in the [redacted] Public School District will be used.		
Will confidential records be required? (If yes, indicate type.) **No**		**Length of time required to complete the research:** The survey will take teachers approximately 10 minutes

To be completed by School District Official:

Approved: [redacted] **ved:**	
Assistant Superinten	**Date:** 3-25-14
Approved to conduct **ing schools:** Listed	

Appendix H
Letter of Support for Implementing Administrator Study

[redacted]

[redacted]

[redacted]
Director of Curriculum, Instruction, Assessment, and Professional Development
[redacted]

March, 2014

Institutional Review Board
Twamley Hall, Room 106
264 Centennial Drive Stop 7134
Grand Forks, ND 58202-7134

Dear Institutional Review Board:

Please accept this communication as a letter of endorsement for doctoral candidate, Mr. Kris Arason, as he pursues his research around teacher evaluation and the role of the principal and teachers, specifically as it relates to the Dr. Robert Marzano Teacher Evaluation Model that the [redacted] Public School District has implemented this school year (2012-13). As the chairperson of this initiative, Mr. Arason's research of our preliminary work, implementation year, and post-follow-up data gathering will be most beneficial to our process. In fact, because North Dakota will now likely be a state that requires much more rigor in teacher evaluation and the new and emerging role of the principal as the chief evaluator, Mr. Arason's research may have wide reaching effects that will assist other school districts.

If I can be of further assistance, please do not hesitate to contact me at [redacted].

Respectfully,

[redacted]
Director, Curriculum, Instruction, Assessment & Professional Development

Appendix I
Email to Administrators Asking Them to Participate in the Study

From: Kris Arason June 4, 2014 9:26:38 AM

Subject: Marzano Teacher Evaluation Survey

To: Secondary Principals Elementary Principals [redacted]

[redacted] Administrators,

Below is a link to a survey for the research I am conducting for my dissertation. My topic is the Marzano Teacher Evaluation Framework (MTEF) and the implementation process. This is a district-wide study that will include all the elementary, middle, and high school administrators.

As the school district continues to move forward with the MTEF, I am interested in administrators' insight. The purpose of this study is to examine the implementation process to date and look for ways to improve the process going forward. For this research project, I will be examining the implementation of the model and administrators' perceptions of the model's ability to improve instruction, its reliability, and ease of use. Below is a link to a short survey that will give valuable feedback regarding the MTEF.

https://www.surveymonkey.com/s/63DDS6L

The survey will take approximately five minutes to complete and is completely anonymous. Participation in this study is voluntary and you may refuse to participate without consequences. Response to the survey will only be reported in aggregate form to protect the identity of the respondents.

Thank you for your consideration. Your time and help is greatly appreciated.

Regards,

Kris Arason

Mr. Kris Arason, Principal
Red River High School, 2211 17th Avenue South, Grand Forks, ND 58201
Office Phone: 701-746-2205 Ext. 6800 Office Fax: 701-746-2406
kris.arason@gfschools.org

REFERENCES

American Recovery and Reinvestment Act (ARRA) of 2009, Pub. L. No. 111-5, 123 Stat. 115-521 (U.S. Government Publishing Office, 2009). Retrieved from https://www.gpo.gov/fdsys/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf

Aseltine, J. M., Faryniarz, J. O., & Rigazio-DiGilio, A. J. (2006). *Supervision for learning: A performance-based approach to teacher development and school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Ayer, F. C. (1954). *Fundamentals of instructional supervision*. New York, NY: Harper.

Baete, G. S., & Hochbein, C. (2014). Project proficiency: Assessing the independent effects of high school reform in an urban district. *Journal of Educational Research, 107*(6), 493-511. doi:10.1080/00220671.2013.823371

Bambrick-Santoyo, P. (2012). Perfecting practice: Instructional leadership requires the principal to work closely with teachers to practice the skills they need to improve their work. *Phi Delta Kappan, 94*(1), 70.

Barr, A., Burton, W., & Brueckner, L. (1947). *Supervision: Democratic leadership in the improvement of learning* (2nd ed.; Appleton series in supervision and teaching). New York, NY: D. Appleton-Century Company.

Bennett, W. (1986). *What works: Research about teaching and learning*. Washington, DC: United States Department of Education.

Bill & Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study*. Retrieved from http://k12education.gatesfoundation.org/wp-content/uploads/2015/05/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

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

Blumberg, A. (1985). *The school superintendent : Living with conflict*. New York, NY: Teachers College, Columbia University.

Bobbitt, J. F., Parker S. C., & Monahan, A. C. (1913). *The supervision of city schools*. Chicago, IL: University of Chicago Press.

Borman, G. D., & Kimball, S. M. (2005, September). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *Elementary School Journal, 106*(1), 3. doi:10.1086/496904

Bridges, D. (1979). *Education, Democracy, & Discussion*. Atlantic Highlands, NJ: Humanities Press.

Brown, J. L. (2008, November 14). *Improving student achievement using Robert Marzano's,* The art and science of teaching*: Part One—Instructional Factors*. Presented at the Diocese of Cleveland, OH. Retrieved from http://www.leonschools.net/cms/lib7/FL01903265/Centricity/domain/547/marzano/marzano.pdf

Burke, P., & Krey, R. D. (2005). *Supervision: A guide to instructional leadership* (2nd ed.). Springfield, IL: C. C. Thomas Publisher.

Burkhardt, G., Monsour, M., Valdez, G., Gunn, C., Dawson, M., Lemke, C., . . . Martin, C. (2003). *enGauge® 21st century skills (for 21st century learners): Literacy in the digital age*. Retrieved from http://pict.sdsu.edu/engauge21st.pdf

Burrup, P. E. (1960). *The teacher and the public school system*. New York, NY: Harper.

Callister Everson, K., Feinauer, E., & Sudweeks, R. R. (2013). Rethinking teacher evaluation: A conversation about statistical inferences and value-added models. *Harvard Educational Review*, *83*(2), 349-370.

Canelake, C. (2012). *Implementing a standards-based teacher evaluation system: Learning experiences for administrators in an urban school district* (Unpublished doctoral dissertation). Arizona State University, Phoenix, AZ.

Cash, A. H., Hamre, B. K., Pianta, R. C., & Myers, S. S. (2012, 3rd Quarter). Rater calibration when observational assessment occurs at large scale: Degree of calibration and characteristics of raters associated with calibration. *Early Childhood Research Quarterly*, *27*(3), 529-542.

Chenoweth, K. (2016). ESSA offers changes that can continue learning gains. *Phi Delta Kappan*, *97*(8), 38-42.

Cherasaro, T. L., Brodersen, R. M., Reale, M. L., & Yanoski, D. C. (2016, November). *Teachers' responses to feedback from evaluators: What feedback characteristics matter?* (REL 2017-190). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/central/pdf/REL_2017190.pdf

City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education: A network approach to improving teaching and learning*. Cambridge, MA: Harvard Education Press.

Cogan, M. L. (1972). *Clinical supervision*. Boston, MA: Houghton Mifflin.

Coggshall, J. G., Rasmussen, C., Colton, A., Milton, J., & Jacques, C. (2012, May). *Generating teacher effectiveness: The role of job-embedded professional learning in teacher evaluation* (Research & policy brief). Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/GeneratingTeachingEffectiveness.pdf

Cohen, J. (2015). Challenges in identifying high-leverage practices. *Teachers College Record*, *117*(7), 1-41.

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, *45*(6), 378-387.

Cohen, L., & Manion, L. (1985). *Research methods in education* (2nd ed.). New York, NY: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2007*). Research methods in education* (6th ed.). New York, NY: Routledge.

Coleman, J. S. (1988). Social capital in the creation of human capital [Supplement: Organizations and Institutions: sociological and Economic Approaches to the Analysis of Social Structure]. *The American Journal of Sociology*, *94*, S95-S120.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: National Center for Educational Statistics (Available from the U.S. Government Printing Office).

Common Core. (2009). *Why we're behind: What top nations teach their students but we don't*. Washington, D.C.: Author. Retrieved from http://www.giarts.org/sites/default/files/Why-Were-Behind.pdf

Couper, M. P. (2000, December). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*, 464–494.

Covey, S. R. (1991). *Principle-centered leadership*. New York, NY: Simon & Schuster.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Cremin, L. A. (1976). *Public education*. New York, NY: Basic Books.

Cubberley, E. P. (1922). *Public school administration: A statement of the fundamental principles underlying the organization and administration of public education.* Boston, MA: Houghton Mifflin Company.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2001, February). New trends in teacher evaluation. *Educational Leadership, 58*(5), 12-15.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2010). Evaluations that help teachers learn. *Educational Leadership, 68*(4), 35-39.

Danielson, C. (2012). Observing classroom practice (classroom observations). *Educational Leadership*, *70*(3), 32-37.

Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (1998). Standards for assessing teaching effectiveness are key. *Phi Delta Kappan*, *79*(5), 471-472.

Darling-Hammond, L. (2010, Summer). Steady work: Finland builds a strong teaching and learning system. *Rethinking Schools*, *24*(4), 30-35.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement.* New York, NY: Learning Forward.

Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator, 38*(1), 4-13.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009, February). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad.* National Staff Development Council and The School Redesign Network at Stanford University. Retrieved from https://learningforward.org/docs/pdf/nsdcstudy2009.pdf

Davis, D., Ellett, C., & Annunziata, J. (2002). Teacher evaluation, leadership and learning organizations. *Journal of Personnel Evaluation in Education*, *16*(4), 287-301. doi:10.1023/A:1021791907098

Davis, S., Darling-Hammond, L., LaPointe, M., & Meyerson, D. (2005). *School leadership study: Developing successful principals*. Stanford, CA: Stanford University, Stanford Educational Leadership Institute.

Dean, C. B., Hubbell, E. R., Pitler, H. & Stone, B. (2012). *Classroom instruction that works: Research-based strategies for increasing student achievement* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Derrington, M. L., & Campbell, J. W. (2015, Fall). Principal concerns and superintendent support during teacher evaluation changes. *AASA Journal of Scholarship & Practice*, *12*(3), 11-23.

Dewey, J. (1938). *Experience and education*. New York, NY: Macmillan.

Díaz-Maggioli, G. (2004). *Teacher-centered professional development*. Alexandria, VA: Association for Supervision and Curriculum Development.

Donaldson, M. L. (2009, June). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/issues/2009/06/pdf/teacher_evaluation.pdf

Donaldson, M. L., & Donaldson, G. A., Jr. (2012). Strengthening teacher evaluation: What district leaders can do. *Educational Leadership*, *69*(8), 78-82.

Elementary and Secondary Education Act of 1965, 20 U.S.C. § 6301 *et seq*. (Office of the Law Revision Counsel, 2017). Retrieved from http://uscode.house.gov/

Ellett, C. D. (1997). Classroom-based assessments of teaching and learning. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 107-128). Thousand Oaks, CA: Corwin Press.

Ellett, C. D. & Teddlie, C. (2003, March). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education, 17*(1), 101-128. doi:10.1023/A:1025083214622

Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363-406.

Ericsson, K. A., Charness, N., Hoffman, R. R., & Feltovich, P. J. (Eds.). (2006). *The Cambridge handbook of expertise and expert performance*. New York: Cambridge University Press.

Every Student Succeeds Act of 2015, 20 U.S.C. § 6301 *et seq*. (United States Copyright Office, 2015). Retrieved from https://congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Feeney, E. J. (2007). Quality feedback: The essential ingredient for teacher success. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *80*(4), 191-197.

Ferguson, R. F. (1991, Summer). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28*(2), 465-498.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher*, *31*(8), 4-14.

Finkelstein, C. (2016, April). Thank you so much for the truth! *Phi Delta Kappan*, *97*(7), 19-24.

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher, 43*(2), 100-107.

Freppon, P. A. (2001). *What it takes to be a better teacher: The role of personal and professional development.* Portsmouth, NH: Heinemann.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco, CA: Jossey-Bass.

Fullan, M. (2011). *Choosing the wrong drivers for whole system reform*. Retrieved from: http://michaelfullan.ca/wp-content/uploads/2016/06/13396088160.pdf

Fullan., M., Cuttress, C., & Kilcher, A. (2005, Fall). 8 Forces for leaders of change. *Journal of Staff Development, 26*(4), 54-64.

Garrett, K. (2011, October). Value added: Do new teacher evaluation methods make the grade? *Education Digest: Essential Readings Condensed for Quick Review, 77*(2), 40-45.

Gay, L. R., Mills, G. E., & Airasian, P. (2009). *Educational research: Competencies for analysis and applications* (9th ed.). Upper Saddle River, N.J.: Pearson Merrill Prentice Hall.

Glickman, C. D. (2002). *Leadership for learning: how to help teachers succeed*. Alexandria, VA: Association for Supervision and Curriculum Development.

Goe, L., Biggers, K., & Croft, A. (2012, May). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning* (Research & policy brief). Washington, DC: National Comprehensive Center for Teacher Quality Retrieved from http://www.gtlcenter.org/sites/default/files/docs/LinkingTeacherEval.pdf

Goe, L., Holdheide, L., & Miller, T. (2011, May). A practical guide to designing comprehensive teacher evaluation systems: A tool to assist in the development of teacher evaluation systems. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.lauragoe.com/lauragoe/practicalguideevalsystems.pdf

Goldhaber, D. D., Goldschmidt, P., & Tseng, F. (2013, June). Teacher value-added at the high-school level: Different models, different answers? *Educational Evaluation and Policy Analysis*, *35*(2), 220-236. doi:10.3102/0162373712466938

Goldhammer, R., Anderson, R., & Krajewski, R. (1980). *Clinical supervision : Special methods for the supervision of teachers* (2nd ed.). New York: Holt, Rinehart, and Winston.

Goldman, S. (1966). *The school principal*. New York, NY: Center for Applied Research in Education.

Goodwin, B. (2011). *Simply better: Doing what matters most to change the odds for student success*. Alexandria, VA: Association for Supervision and Curriculum Development.

Goodwin, B. (2015, November). Research says / Mindsets are key to effective data use. *Educational Leadership*, *73*(3), 78-79.

Greene, J. P. (2002, Summer). The business model. *Education Next*, *2*(2), 20-22.

Greenstone, M., Looney, A., & Shevlin, P. (2011, September). *Improving student outcomes: Restoring America's education potential* (Strategy paper for the Hamilton Project). Washington, DC: Brookings. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/092011_education_greenstone_looney_shevlin.pdf

Grieder, C., & Romine, S. (1965). *American education: An introduction to the teaching profession* (3rd ed.). New York, NY: Ronald Press.

Gruenert, S., & Whitaker, T. (2015). *School culture rewired: How to define, assess, and transform it*. Alexandria, VA: Association for Supervision and Curriculum Development.

Hall, G. E., Wallace, R. C., Jr., & Dossett, W. F. (1973, September). *A developmental conceptualization of the adoption process within educational institutions*. Austin, TX: The University of Texas at Austin, Research and Development Center for Teacher Education.

Hanushek, E. A. (1992, February). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84-117.

Hanushek, E. A., & Rivkin, S. G. (2010, May). Generalizations about using value-added measures of teacher quality. *The American Economic Review, 100*(2), 267-271.

Hargreaves, A. (2009). The fourth way of change: Towards an age of inspiration and sustainability. In A. Hargreaves & M. Fullan (Eds.), *Change wars* (pp. 11-43). Bloomington, IN: Solution Tree.

Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York, NY: Teachers College Press.

Harris, D. N. (2010, May). Clear away the smoke and mirrors of value-added. *Phi Delta Kappan, 91*(8), 66-69.

Hastings, J. S., Kane, T. J., & Staiger, D. O. (2006). Gender and performance: Evidence from school assignment by randomized lottery. *The American Economic Review, 96*(2), 232-236.

Hattie, J. (1992). *Self-concept*. Hillsdale, NJ: L. Earlbaum Associates.

Hazi, H. M., & Rucinski, D. A. (2009, March). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *Education Policy Analysis Archives*, *17*(5), 1-22.

Heneman, H., & Milanowski, A. (2003, June). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(2), 173-195. doi:10.1023/B:PEEV.0000032427.99952.02

Heneman, H. G., III, Milanowski, A., Kimball, S. M., & Odden, A. (2006, May). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay* (CPRE Policy Briefs). Philadelphia, PA: University of Pennsylvania, Graduate School of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED493116.pdf

Hill, H. C., Charalambous, C, Y., & Kraft, M. A. (2012, March). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, *41*(2), 56-64.

Hill, H. C., Kapitula, L., & Umland, K. (2011, June). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*(3), 794-831. doi:10.3102/0002831210387916

Johnson, S. M. (1990). *Teachers at work: Achieving Success in our schools.* New York, NY: Basic Books.

Kane, T. J. (2012, Fall). Capturing the dimensions of effective teaching. *Education Next*, *12*(4), 34-41.

Kane, T. J., & Stainger, D. O. (2008, December). *Estimating teacher impacts on student achievement: An experimental evaluation* (NBER Working Paper No. 14607). Cambridge, MA: National Bureau of Economic Research. Retrieved from http:/www.nber.org/papers/w14607

Kelleher, J. (2016). You're OK, I'm OK. *Phi Delta Kappan*, *97*(8), 70-73.

Kilgore, S. B., & Reynolds, K. J. (2011). *From silos to systems: Reframing schools for success.* Thousand Oaks, CA: Corwin.

Kimball, S. M., & Milanowski, A. (2009, February). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly, 45*(1), 34-70. doi:10.1177/0013161X08327549

Kotter, J. P. (1996). *Leading change*. Boston, MA: Harvard Business School Press.

Learning Forward. (2011). *Standards for professional learning.* Oxford, OH: Author.

Leana, C. R. (2011, Fall). The missing link in school reform. *Stanford Social Innovation Review, 9*(4), 30-35. Retrieved from https://www2.ed.gov/programs/slcp/2011progdirmtg/mislinkinrfm.pdf

Lin, E., Wang, J., Klecka, C. L., Odell, S. J., & Spalding, E. (2010, September-October). Judging research in teacher education. *Journal of Teacher Education*, *61*(4), 295-301. doi:10.1177/0022487110374013

Looney, J. (2011, December). Developing high-quality teachers: Teacher evaluation for improvement. *European Journal of Education, 46*(4), 440-455. doi:10.1111/j.1465-3435.2011.01492.x

Lucio, W. H., & McNeil, J. D. (1968). *Supervision: A synthesis of thought and action* (2nd ed.). New York, NY: McGraw-Hill.

Margolis, J., & Huggins, K. S. (2012). Distributed but undefined: New teacher leader roles to change schools. *Journal of School Leadership, 22*(5), 953-982.

Marks, J. R., Stoops, E., & King-Stoops, J. (1978). *Handbook of educational supervision : A guide for the practitioner* (2nd ed.). Boston, MA: Allyn and Bacon.

Marks, J. R., Stoops, E., & King-Stoops, J. (1985). *Handbook of educational supervision : A guide for the practitioner* (3rd ed.). Boston, MA: Allyn and Bacon.

Marshall, K. (2005, June). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan, 86*(10), 727-735.

Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, Va.: Association for Supervision and Curriculum Development.

Marzano, R. J. (2007). *The art and science of teaching : A comprehensive framework for effective instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2010, December / 2011, January). Art & science of teaching / What teachers gain from deliberate practice. *The Effective Educator, 68*(4), 82-85.

Marzano, R. J. (2012, May). Art & science of teaching / Teaching self-efficacy with personal projects. *Educational Leadership, 69*(8), 86-87.

Marzano, R. J., & Pickering, D. (2003). *Classroom management that works: Research-based strategies for every teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., & Waters, T. (2009). *District leadership that works: Striking the right balance*. Bloomington, IN: Solution Tree.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.

Maslow, V. J., & Kelley, C. J. (2012). Does evaluation advance teaching practice? The effects of performance evaluation on teaching quality and system change in large diverse high schools. *Journal of School Leadership, 22*(3), 600-632.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*(1), 67-102.

McGuinn, P. (2015, September). Complicated politics to the core. *Phi Delta Kappan, 97*(1), 14-19.

McNeely, B. (2005). Using technology as a learning tool, not just the cool new thing. In D. G. Oblinger & J. L. Oblinger (Eds.), *Educating the Net generation* (pp. 4.1 -4.10). Boulder, CO: EDUCAUSE.

Medley, D. M., & Coker, H. (1987, March-April). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research, 80*(4), 242-247.

Milanowski, A. T., & Heneman, H. G., III. (2001, September). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study*. *Journal of Personnel Evaluation in Education, 15*(3), 193-212. doi:10.1023/A:1012752725765

Millman, J., & Darling-Hammond, L. (Eds.). (1990). *The new handbook of teacher evaluation : Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin Press, Inc. [A Sage Publications Company].

Minnici, A. (2014, Spring). The mind shift in teacher evaluation: Where we stand – and where we need to go. *American Educator, 38*(1), 22-26.

Mizell, H. (2010, March). School boards should focus on learning for all. *Phi Delta Kappan, 91*(6), 20-23. doi: 10.2307/27755662

Mizell, H., Hord, S., Killion, J., & Hirsh., S. (2011). New standards putting the spotlight on professional learning. *Journal of Staff Development, 32*(4), 10-14.

Muijs, D. (2006). New directions for school effectiveness research: Towards school effectiveness without schools. *Journal of Educational Change, 7*(3), 141-160. doi:10.1007/s10833-006-0002-7

Myers, A. F., Kifer, L. M., Merry, R. C., & Foley, F. (1938). *Cooperative supervision in the public schools*. New York, NY: Prentice-Hall.

National Commission on Excellence in Education (NCEE). (1983, April). *A nation at risk: The imperative for educational reform*. Washington, D.C.: US Department of Education. Retrieved from https://www2.ed.gov/pubs/NatAtRisk/risk.html

National Commission on Teaching and America's Future, The. (1996, September). What matters most: Teaching for America's future (Report of the National Commission on teaching & America's future). Woodbridge, VA: National Commission on Teaching and America's Future. Retrieved from: http://nctaf.org/research/publications/

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Oblinger, D., & Oblinger, J. (2005). Is it age or IT: First steps toward understanding the Net generation. In D. G. Oblinger & J. L. Oblinger (Eds.), *Educating the Net generation* (pp. 2.1-2.20). Boulder, CO: EDUCAUSE.

Odden, A. (2004, October). Lessons learned about standards-based teacher evaluation systems. *Peabody Journal of Education, 79*(4), 126-137.

OECD. (2011). *Strong performers and successful reformers in education: Lessons from PISA for the United States*. OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264096660-*en*

O'Pry, S. C., & Schumacher, G. (2012, November). New teachers' perceptions of a standards-based performance appraisal system. *Educational Assessment, Evaluation and Accountability*, *24*(4), 325-350. doi:10.1007/s11092-012-9148-4

Papay, J. P. (2012, Spring). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review, 82*(1), 123-141.

P21: Partnership for 21st Century Learning. (2007). *Framework for 21st century learning* [Web page]. Retrieved from http://www.p21.org/our-work/p21-framework

Pecheone, R. L., & Whittaker, A. (2016, April). Well-prepared teachers inspire student learning. *Phi Delta Kappan, 97*(7), 8-13.

Peterson, P. (2006). *Choice and competition in American Education.* Lanham, MD: Rowman & Littlefield.

Ramaley, J. A., & Zia, L. (2005). The real versus the possible: Closing the gaps in engagement and learning. In D. G. Oblinger & J. L. Oblinger (Eds.), *Educating the Net generation* (pp. 8.1-8.21). Retrieved from http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1004&context=pubadmin_fac

Ramirez, A., Lamphere, M., Smith, J., Brown, S., & Pierceall-Herman, J. (2011, July). Teacher development and evaluation: A study of policy and practice in Colorado. *Management in Education, 25*(3), 95-99. doi:10.1177/0892020610387956

Range, B. G., Scherz, S., Holt, C. R., & Young, S. (2011, August). Supervision and evaluation: The Wyoming perspective. *Educational Assessment, Evaluation and Accountability*, *23*(3), 243-265.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education.* New York, NY: Basic Books.

Reagan, T. G., Case, C. W., & Brubacher, J. W. (2000). *Becoming a reflective educator: How to build a culture of inquiry in the schools* (2nd ed.). Thousand Oaks, CA: Corwin Press, Inc. (A Sage Publications Company).

Reeves, D. B. (2009). *Leading change in your school: How to conquer myths, build commitment, and get results*. Alexandria, VA: Association for Supervision and Curriculum Development.

Ritter, G. W., & Barnett, J. H. (2016, April). Learning on the job: Teacher evaluation can foster real growth. *Phi Delta Kappan*, *97*(7), 48-52. doi:10.1177/0031721716641649

Robinson, C. (2015). Human capital, education, achievement and learning. *Journal of Economic and Social Measurement, 40*(1-4), 69-97.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 635-674. doi:10.1177/0013161X08321509

Rockoff, J. E. (2004, May). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*(2), 247-252.

Rosenholtz, S. J. (1991). *Teacher's workplace: The social organization of schools*. New York, NY: Teachers College Press.

Rudestam, K. E., & Newton, R. R. (2007). *Surviving your dissertation: A comprehensive guide to content and process* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Rury, J. L. (1991). *Education and women's work: Female schooling and the division of labor in urban America, 1870-1930*. Albany, NY: State University of New York Press.

Ryan, F., & Kuhlen, R. G. (1958, June). Trait ratings of high school students by teachers. *Journal of Educational Psychology, 49*(3), 124-128.

Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland?* New York, NY: Teachers College Press.

Saltzman, A. (2016). *The power of principal supervisors: How two districts are remaking an old role* (Stories from the field). New York: The Wallace Foundation. Retrieved from http://www.wallacefoundation.org/knowledge-center/Documents/The-Power-of-Principal-Supervisors.pdf

Sarason, S. B. (1995). *School change: The personal development of a point of view*. New York, NY: Teachers College Press, Columbia University.

Sartain, L., Stoelinga, S. R., & Brown., E. R. (with Luppescu, S., Matsko, K. K., Miller, F. K., Durwood, C., Jiang, J. Y., & Glazer, D.). (2011, November). Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation (Research report). Chicago, IL: The University of Chicago, Urban Education Institute, Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/publications/rethinking-teacher-evaluation-chicago-lessons-learned-classroom-observations-principal

Schumacher, G. (2011). Key factors for successfully implementing and sustaining quality improvement in K-12 education. *Journal for Quality & Participation, 33*(4), 17-20.

Seely Flint, A., Zisook, K., & Fisher, T. R. (2011, November). Not a one-shot deal: Generative professional development among experienced teachers. *Teaching and Teacher Education: An International Journal of Research and Studies*, *27*(8), 1163-1169.

Sergiovanni, T. J., & Starratt, R. J. (1971). *Emerging patterns of supervision: Human perspectives*. New York, NY: McGraw-Hill.

Sergiovanni, T. J., & Starratt, R. J. (2002). *Supervision: A redefinition* (7th ed.). Boston, MA: McGraw-Hill.

Shakman, K., Breslow, N., Kochanek, J., Riordan, J., & Haferd, T. (2012). *Changing cultures and building capacity: An exploration of district strategies for implementation of teacher evaluation systems*. Learning and Teaching Division, Education Development Center, Inc. Retrieved from http://ltd.edc.org/sites/ltd.edc.org/files/District%20Strategies%20for%20Implementation%20of%20Teacher%20Evaluation%20Systems.pdf

Sheskey, B. (2010). Creating learning connections with today's tech-savvy student. In H. H. Jacobs (Ed.), *Curriculum 21: Essential education for a changing world* (pp. 195-209). Alexandria, VA: Association for Supervision and Curriculum Development.

Shinkfield, A. J., & Stufflebeam, D. L. (1995). *Teacher evaluation: Guide to effective practice*. Boston, MA: Kluwer Academic Publishers.

Stanulis, R. N., Cooper, K. S., Dear, B., Johnston, A. M., & Richard-Todd, R. R. (2016). Teacher-led reforms have a big advantage – Teachers. *Phi Delta Kappan*, *97*(7), 53-57.

Stewart, V. (2011). Raising teacher quality around the world. *Educational Leadership*, *68*(4), 16-20.

Stiggins, R. J., & Duke, D. L. (1988). *The case for commitment to teacher growth : Research on teacher evaluation*. Albany, NY: State University of New York Press.

Stigler, J. W., & Hiebert, J. (2009, November). Closing the Teaching Gap. *Phi Delta Kappan Magazine*, *91*(3), 32-37.

Strong, M., Gargani, J., & Hacifazlioglu, O. (2011). Do we know a successful teacher when we see one?: Experiments in the identification of effective teachers. *Journal of Teacher Education*, *62*(4), 367-382. doi: 10.1177/0022487110390221

Stronge, J. H., & Hindman, J. L. (2006). *Teacher quality index: A protocal for teacher selection*. Alexandria, VA: Association for Supervision and Curriculum Development.

Stumbo, C., & McWalters, P. (2011). Measuring effectiveness: What will it take? *Educational Leadership*, *68*(4), 10-15.

Sun, C. (2011, August). School leadership: Improving state systems for leader development (Discussion guide). Arlington, VA: National Association of State Boards of Education. Retrieved from http://www.wallacefoundation.org/knowledge-center/Documents/NASBE-Discussion-Guide-School-Leadership-Improving-State-Systems-for-Leader-Development.pdf

Sutton, P. S., & Shouse, A. W. (2016, April). Building a culture of collaboration in schools. *Phi Delta Kappan*, *97*(7), 69-73.

Swearingen, M. E. (1962). *Supervision of instruction: Foundations and dimensions*. Boston: Boston, Allyn and Bacon.

Tanner, D., & Tanner, L. N. (1987). *Supervision in education: Problems and practices*. New York: Collier Macmillan.

Tapscott, D. (2009). *Grown up digital: How the net generation is changing your world*. New York, NY: McGraw-Hill.

Taylor, E. S., & Tyler, J. H. (2012, December). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628-3651.

Thompson, M. M. (1958). *The history of education* (3rd ed.). New York, NY: Barnes & Noble.

Toch, T. (2008, October). Fixing: Teacher evaluation. *Educational Leadership, 66*(2), 32-37.

Torff, B., & Sessions, D. (2009, Summer). Principals' Perceptions of the Causes of Teacher Ineffectiveness in Different Secondary Subjects. *Teacher Education Quarterly*, *36*(3), 127-148.

Tornero, B. & Taut, S. (2010, December). A mandatory, high-stakes National Teacher Evaluation System: Perceptions and attributions of teachers who actively refuse to participate. *Studies in Educational Evaluation*, *36*(4), 132–142.

Tracy, S. J. (1995, May-June). How historical concepts of supervision relate to supervisory practices today. *Clearing House*, *68*(5), 320-325.

Tucker, M. S. (2011, May 24). *Standing on the shoulders of giants: An American agenda for education reform*. Washington, D.C.: National Center on Education and the Economy®. Retrieved from http://ncee.org/wp-content/uploads/2011/05/Standing-on-the-Shoulders-of-Giants-An-American-Agenda-for-Education-Reform.pdf

Tucker, M. S. (2014). *Fixing our national accountability system*. Washington, D.C.: National Center on Education and the Economy®. Retrieved from http://www.ncee.org/wp-content/uploads/2014/08/FixingOurNationalAccountabilitySystemWebV4.pdf

Tuytens, M., & Devos, G. (2011, July). Stimulating professional learning through teacher evaluation: An impossible task for the school leader? *Teaching and Teacher Education: An International Journal of Research and Studies*, *27*(5), 891-899. doi:10.1016/j.tate.2011.02.004

Tyler, J. H., Taylor, E. S., Kane, T. J., & Wooten, A. L. (2010). Using student performance data to identify effective classroom practices. *The American Economic Review*, *100*(2), 256-260.

U.S. Department of Education. (2009, November). *Race to the top program: Executive summary*. Washington, D.C.: Author. Retrieved from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education. (2010, March). *A blueprint for reform: The reauthorization of the elementary and secondary education act.* Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development. Retrieved from https://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf

U.S. Department of Education. (2011). *Bringing flexibility & focus to education law: Looking back and moving forward* [Brochure]. Retrieved from https://www.ed.gov/sites/default/files/looking-back-moving-forward.pdf

U.S. Department of Education. (2015, December 2). *Fact sheet: Congress acts to fix No Child Left Behind.* Retrieved from https://www.ed.gov/news/press-releases/fact-sheet-congress-acts-fix-no-child-left-behind

U.S. Department of Education, Office of Educational Technology. (2010, March 5). *Transforming American education: Learning powered by technology* (Draft National Educational Technology Plan 2010). Retrieved from https://www.ed.gov/sites/default/files/NETP-2010-final-report.pdf

Warner, R. M. (2013). *Applied statistics from bivariate through multivariate techniques* (2nd ed.). Los Angeles, CA: Sage.

Watson, S., Miller, T., Davis, L., & Carter, P. (2010, Fall). Teachers' perceptions of the effective teacher. *Research in the Schools, (17)*2, 11-22.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (contributing authors Schunck, J., Palcisco, A., & Morgan, K.). (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project. Retrieved from http://widgeteffect.org

White, B. R., Cowhy, J., Stevens, W. D., & Sporte, S. E. (2012, November). *Designing and implementing the next generation of teacher evaluation systems: Lessons learned from case studies in five Illinois districts* (Research brief). The University of Chicago Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/sites/default/files/publications/Teacher%20Evaluation%20Policy%20Brief1_0.pdf

Whitehurst, G., Chingos, M. M., & Lindquist, K. (2015). Getting classroom observations right. *Education Next, 15*(1), 62-68.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Wiles, K. (1967). *Supervision for better schools* (3rd ed.). Englewood Cliffs, N. J.: Prentice-Hall.

Wiles, J., & Bondi, J. (1980). *Supervision, a guide to practice*. Columbus, OH: Merrill.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14.

Williams, J. (2009). *McREL's teacher evaluation system*. Denver, CO: Mid-continent Research for Education and Learning. Retrieved from https://www.medford.k12.nj.us/cms/lib/NJ01001377/Centricity/Domain/3/McRELTeacher%20Evaluation%20Users%20Guide.pdf

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1985). Teacher evaluation: A study of effective practices. *The Elementary School Journal, 86*(1), 61-121. doi:10.1086/461437

Whitehead, M. J. (1952, November). Teachers look at supervision. *Educational Leadership*, *10*(2), 100-106.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57-67. Retrieved from https://pdfs.semanticscholar.org/1677/01cd3fba9fdf5ef850eaca75b1b101334ed2.pdf

Zimmerman, S., & Deckert-Pelton, M. (2003, September). Evaluating the evaluators: Teachers' perceptions of the principal's role in professional evaluation. *NASSP Bulletin, 87*(636), 28-37.