Principals’ Perceptions of the Massachusetts Educator Evaluation System

A Doctoral Thesis

presented by Sheila A. Muir

to the College of Professional Studies

In partial fulfillment of the requirements

for the degree of Doctor of Education

Dr. Chris Unger

Advisor

Northeastern University

Boston, Massachusetts

June 2016
Abstract

This qualitative study investigated principals’ perceptions of the effects of the 2011 Massachusetts Educator Evaluation System on teacher growth and development, principals’ capacity to support teacher growth and development, and principal-teacher relationships. Participants were eleven elementary level public school principals who are the sole evaluators of teachers in their schools.

The study was guided by the following three research questions

1. How do principals perceive the Massachusetts Educator Evaluation System as supporting teacher growth and development, in theory and as they have experienced it?

2. How do principals perceive the effect of the Massachusetts Educator Evaluation System on their ability, capacity and opportunity to effectively support their teachers’ development?

3. How do principals perceive the effect of the Massachusetts Educator Evaluation System on the development and maintenance of trusting relationships between principals and teachers as they have experienced it?

The findings of this study suggest that the Massachusetts Educator Evaluation System provides some benefits to the growth and development of teachers and to principals in supporting that growth and development. Additionally, the study revealed some aspects of the System that principals perceived to be challenging to implement effectively, and some aspects that principals perceived to be detrimental to teacher growth and development.

Keywords: Teacher Evaluation, Massachusetts Educator Evaluation System, Evaluation, Supervision, Principal-Teacher Trust.
Acknowledgements

I would like to begin by thanking my family for all of their support through this journey. To my incredible husband Olen, I cannot imagine how I could have done this without you. From the special evenings out that you planned after each long Saturday session in Groton to your taking complete responsibility for all things home-related so that I couldn’t make any excuse not to write; your support through these long seven years has made it possible for this goal to be realized. To my children, Justine, Racheal and Kyle I am so proud of the adults that you have become. I hope that through this I have reminded you that each person’s potential is unlimited and that learning should never stop. To my mother Hazel and my late father William, you both taught me the value of hard work and perseverance. Those principles sustained me through the most challenging phases of this doctoral thesis.

I would like to thank my friend and colleague Dr. Kristin Campione for all of her encouragement during the last few months of this process. Kristin, I dragged you out of another doctoral program at the start, and you dragged me through the final months of the thesis. Those countless weekends spent at the office writing in tandem, guided by the goals that you insisted we establish each morning, were instrumental to my progress. Although I had hoped you would beat me to the finish line, I am thrilled that we will graduate together.

I would also like to thank all of my colleagues in the Quabbin Regional School District. I feel blessed to work in a school district where I am surrounded by high quality professionals who are always striving to improve outcomes for students. You always push my thinking and motivate me to continue learning.

I want to offer special thanks to Dr. Maureen Marshall, superintendent of schools for advising and encouraging me throughout this process and throughout my entire career in public
education. Maureen, you inspired me to pursue the study of science in 1976 when I was a student in your biology class, and you continue to inspire me each and every day with your steadfast passion and commitment to provide the highest possible quality education to the students in the Quabbin Regional School District. I am ever grateful to work with you.

To my good friend, mentor and third reader, Dr. Deborah Brady: The fact that you earned your doctorate as a single mom with two sets of twins at home made me feel incredibly inadequate whenever I dared to think I could not do this! I could never repay you for all you have done for me. There truly is never a time when I am with you that I am not learning something that makes me a better educator and leader. Your selflessness and commitment to improving the capacity of each and every person with whom you work is second to none.

To my study participants. As principals who work alone as educational leaders of your schools, you have tremendous responsibility. I remain overwhelmed by the commitment that each of you have to ensure that each educator and student with whom you work is able to reach his or her highest potential. You gave time that you didn’t have to support this work and I am so very thankful for your generosity.

To my second reader Dr. Bair, your thoughtful and detailed feedback helped tremendously.

Dr. Unger, I could not have asked for a better advisor. I cannot thank you enough for all of the things that you did to support me. Your encouraging words and strategic deadlines helped me to accomplish each milestone and made it seem possible that I would reach this moment. I know that without your guidance I would not have made it.
# Table of Contents

Abstract .............................................................................................................................. 2
Acknowledgements ............................................................................................................ 3
Chapter 1: Introduction ........................................................................................................ 8
  Statement of the Problem ............................................................................................... 8
  Research Problem .......................................................................................................... 13
  Justification for Research ............................................................................................. 16
  Deficiencies in the Research ......................................................................................... 19
  Relating the Discussion to Audiences ........................................................................... 22
  Purpose Statement ........................................................................................................ 23
  Research Questions ...................................................................................................... 23
  Significance of the Research ........................................................................................ 24
  Positionality Statement ................................................................................................. 29
  Theoretical Framework .................................................................................................. 32
  Conclusion ..................................................................................................................... 38
Chapter II: Literature Review ............................................................................................. 39
  Overview ....................................................................................................................... 39
  Teacher Evaluation ........................................................................................................ 39
  The Link of Teacher Effectiveness to Student Achievement ........................................... 65
  Effectiveness of Teacher Evaluation for Improvement of Teaching .............................. 67
  Impact of Collaboration on Teacher Practice and Student Learning .............................. 70
  Trust in Schools .......................................................................................................... 71
  Role of the School Principal ........................................................................................ 74
  Trust in Supervision and Evaluation ............................................................................ 78
  Conclusion ..................................................................................................................... 80
Chapter III: Research Design ............................................................................................. 82
  Introduction ................................................................................................................... 82
  Purpose Statement ....................................................................................................... 83
Personal Comments .............................................................................................................172
References........................................................................................................................176
Appendix A: Initial Participant Recruitment Email.................................................................194
Appendix B: Signed Informed Consent Form for In-Person Interviews.................................195
Appendix C: Unsigned Informed Consent Form for Telephone Interviews...............................198
Appendix D: In-Person Interview Guide ..............................................................................201
Appendix E: Telephone Interview Guide ...............................................................................205
Appendix F: Massachusetts Model System for Educator Evaluation-Teacher Rubric.............209
Chapter I: Introduction

Statement of the Problem

The following vignette is a synthesis of numerous discussions with principals, district leaders, and teacher evaluation professional development consultants. This conversation demonstrates the challenges that principals across Massachusetts face as they implement the newly adopted (June, 2011) Massachusetts Educator Evaluation System.

The scenario presented here is a discussion among three principals who are in their second year of using the new Massachusetts Educator Evaluation System. The primary focus for the meeting described below was calibration of summative evaluation ratings. Meetings of this nature take place regularly in many districts to address challenging teacher evaluation situations, and to calibrate their assessments of teacher performance expectations across evaluators and schools. A district leader is often present at these meetings. Typically, this type of meeting takes place in the late spring as principals begin reviewing the data they have collected about each teacher and assign an overall summative or formative evaluation rating, and a rating for each of four standards of performance (Standards I-IV). Each teacher’s rating is submitted to the Massachusetts Department of Elementary and Secondary Education (DESE) annually. The ratings are used at the district level to place underperforming teachers on more rigorous evaluation plans to provide more support for improvement or, in cases in which improvement has not been sufficient, to dismiss teachers. The discussion below is an example of such an interaction among principals:

Principal 1: When I compare her performance to the rubric descriptors, I know she should be rated Unsatisfactory, but if I give her that rating she will hate me. A rating of
Unsatisfactory seems so harsh, but that’s what I see when I’m in her room and when I review her data. I don’t know what to do.

Principal 2: I know. I have a teacher like that, too. I thought about suspending the evaluation process for a few months so that I could coach her, but I decided against it because I would not get the required number of observations completed if I did that. She really needs help, and I think she knows that, but she won’t admit it so we can’t collaboratively discuss the improvement strategies that I have suggested. I can’t figure out how to help her improve.

Principal 3: When I had a teacher like that, I did not tell her how bad it really was. I rated her Needs Improvement in Standard I and Proficient in Standard II so that she wouldn’t go on an Improvement Plan. I had one teacher on an Improvement Plan in my building already and I knew I could not handle the workload if I had two. Even though the ratings she received weren’t really accurate, she did improve. She focused on the specific feedback that I gave her each time I was in her room, and little by little, she got better. I would recommend that you handle your situations that way.

Principal 1: Well what about dismissal? At this point what I am seeing is horrible. If I don’t accurately rate her, and if she doesn’t improve with my suggestions, then she won’t be on an Improvement Plan, and I couldn’t dismiss her, right?

Principal 3: Well, yes, that’s true. It would take an extra year, but what will the other teachers in your building think if you rate her Unsatisfactory and put her on an Improvement Plan? Won’t they be afraid to come to you with questions and to discuss challenging situations? Will they fear that disclosing their challenges to you will cause you to view them negatively?
Principal 1: I know. I have been thinking about that, too. This is a tricky situation for me as a new principal. I think some teachers are beginning to trust me. I don’t want to do anything that will erode that trust. On the other hand, I am worried that the students in her classroom aren’t learning. Her data is not showing adequate student growth. I can’t let that continue. I need to keep student achievement my highest priority, right?

For these principals, the struggle is not with their ability to recognize good teaching or teaching that must improve; the struggle is how to effectively implement a new evaluation system that is designed to foster improvements in teacher quality through two distinct simultaneously occurring processes: gathering evidence that will be used to judge teacher performance for accountability purposes, and gathering evidence that will be used as feedback to teachers for the purposes of improving their professional practice.

Numerous teacher evaluation consultants, district leaders, and principals with whom the researcher has interacted concur that conversations like the one above occur regularly in Massachusetts public schools. This is especially true since the state of Massachusetts adopted new regulations for educator evaluation in June of 2011 to comply with the requirements of the federal Race to the Top (RTTT) grant. The RTTT grant provided Massachusetts with 250 million dollars earmarked for education innovation and reform (Massachusetts Department of Elementary and Secondary Education, 2013). RTTT funding, however, came with some strings attached: “Improving principal and teacher effectiveness based on performance” by designing and implementing “rigorous, transparent and fair evaluation systems” was one of Massachusetts’ requirements for receiving funding (U.S. Department of Education, 2009, p.3).

The new regulations in Massachusetts differed significantly from the previous regulations. Adopted in 1995, the previous regulations required only that school committees
establish and collectively bargain an evaluation system that included performance standards (Massachusetts Department of Elementary and Secondary Education, 2011). From 1995 to the enactment of the 2011 regulations, educator evaluation in Massachusetts was highly variable from district to district, and it was ineffective in achieving the goals of “providing educators with adequate feedback for improvement and serving as an important accountability tool” (Massachusetts Department of Elementary and Secondary Education, 2011). The regulations enacted in 2011 required district school committees to establish evaluation systems that included the specific Performance Standards for the evaluation of teachers and administrators as well as a specific evaluation cycle that included five steps: self-assessment, goal setting and development of an Educator Plan, implementation of the Educator Plan, formative assessment or formative evaluation, and summative evaluation. For each of these five steps, the regulations further specified detailed requirements (Massachusetts Department of Elementary and Secondary Education, Final Regulations on Evaluation of Educators, 2011). In addition, these regulations indicated requirements about specific evidence that must be used in evaluation (including student performance measures and student feedback), and the four performance levels that were to be used for rating educators (Exemplary, Proficient, Needs Improvement and Unsatisfactory). Finally, it was required under the new regulations that the performance rating of each educator was to be submitted to the Massachusetts Department of Elementary and Secondary Education annually (Massachusetts Department of Elementary and Secondary Education, Final Regulations on Evaluation of Educators, 2011).

The Massachusetts Department of Elementary and Secondary Education developed a Model System that complied with the regulations. School districts were allowed to adopt the Model System, adapt the Model System, or develop their own system provided it complied with
all requirements of the regulations (Massachusetts Department of Elementary and Secondary Education, 2015a). Most school districts adopted or adapted the Massachusetts Model System. The year of initial implementation varied from district to district based on DESE requirements, but by the 2014-2015 school year, all districts were required to have begun the first of three phases of implementation with a required completion date of the 2017-2018 school year. Implementation of new evaluation systems that meet the rigorous requirements of the regulations necessitates significant shifts for both principals and teachers, and as such, presents many potential challenges. Other noteworthy changes, such as student performance measures and student feedback, the formal and structured nature of the evaluation process, and the high-stakes impact of the evaluation ratings that can lead to dismissal, may create further difficulties. In addition, the regulations require both formative assessment and evaluation, and summative evaluation. The formative feedback provided to educators is non-judgmental and intended to help educators improve their practice (Marzano, Frontier & Livingston, 2011). Formative processes are often referred to as supervision rather than as evaluation and, when effectively implemented to improve teacher practice, resemble coaching (Danielson & McGreal; Showers, 1985), wherein a teacher feels comfortable discussing his or her shortcomings or the challenges he or she is facing to receive support for overcoming such things (Nolan & Hoover, 2011). Strong, trusting professional relationships are necessary for formative feedback (supervision) to be effective (Tomal, Wilhite, Phillips, Sims & Gibson, 2015). In contrast, the summative evaluation represents a judgment about the performance of the educator, It is used for accountability purposes, and for employment decisions (Nolan & Hoover, 2011; Sergiovanni & Starratt, 2007). These formative and summative requirements may also represent an implementation challenge, particularly for principals. Providing both judgmental and non-
judgmental feedback to teachers within the same system of evaluation is considered by many scholars to be a conflict (Cogan, 1973; Danielson & McGreal, 2000; Nolan, 1997; Marzano, 2012; Peterson, 2000; Popham, 1998; Wise et al, 1984), because an educator who knows that their evaluator is judging their performance is likely to be reluctant to freely discuss and reflect on his or her practice in the presence of the evaluator (Danielson & McGreal, 2000; Peterson, 2000), thereby limiting the ability of the evaluator to provide the support that may be needed to improve the teacher’s practice.

Principals’ primary responsibility is instructional leadership (Fullan, 2009; NAESP, 2008), so they must be able to implement practices that help teachers grow professionally so student achievement will increase. The presence of a trusting relationship between teachers and their principal has been shown to improve student achievement (Byrk & Schneider, 2003). Given that the Massachusetts Educator Evaluation System formally links formative assessment and evaluation with summative evaluation, and that implementation of these practices is mandated (Massachusetts Department of Elementary and Secondary Education Final Regulations on the Evaluation of Educators, 2011), it is important that the effect of the required practices on teacher-principal relationships is investigated.

Research Problem

In recent years the role of principal has changed (City, Elmore, Fiarman & Teitel, 2009). The traditional role focused primarily on building management (Elmore, 2000), but with increased evidence of the significant effect principals can have on teacher performance, and in turn, student achievement (Leithwood, Louis, Anderson & Wahlstrom, 2004; Peterson, 2002), the primary role of a principal is now that of instructional leader. Thus, principals are in part responsible for increasing student achievement by improving teacher performance (Cross &
In recent years, improving teacher performance and student achievement has become ever more challenging for principals. Changes in curriculum, specifically the adoption of the Common Core Standards for literacy and mathematics, advances in educational research that provide new and more effective instructional practices for teachers, and shifts in the demographics of student populations that result in more children with learning challenges in classrooms have significantly affected the complexity of the role of classroom teacher. This subsequently affects the role of the instructional leader who is charged with ensuring that teacher performance is effective.

The challenges of instructional leadership have been further complicated by recent mandated changes in educator supervision and evaluation systems. In response to the federal No Child Left Behind Act (NCLB) of 2001 and the U.S. Department of Education Race to the Top program, many states have adopted comprehensive, standards-based educator evaluation systems (Darling-Hammond, 2013). These systems are designed to meet two primary goals: measurement of teacher effectiveness and improvement of teacher effectiveness (Danielson, 2002; Hinchey, 2010). These new educator evaluation systems require principals to make judgments about teacher performance against a set criteria of best practices that were identified in the research for effective teaching (Darling-Hammond, 2013). The evaluative judgments and feedback provided by principals are intended to improve teacher performance by informing teachers of areas in need of improvement (Heneman and Milanowski 2006; Hoe & Kane, 2013; Odden 2004; Odden & Wallace 2008). Principals use these same judgments to inform employment decisions and, in states such as Massachusetts, they are submitted to the Department
of Education to meet accountability requirements (Massachusetts Department of Elementary and Secondary Education, Final Regulations on Evaluation of Educators, 2011).

Implementation of these new requirements for educator evaluation is an added challenge for instructional leaders. First, it requires principals and teachers to understand the new policies and procedures that govern the implementation practices. Second, in most cases it requires evaluators and teachers to invest significantly more time in the evaluation function. The third, and perhaps most significant, challenge is the need to balance the dual purposes of accountability with teacher growth and development.

Teacher and student performance improve when principals couple effective evaluation with opportunities for leaders and teachers to engage in reflective, collaborative, and collegial discussions focusing on improved teacher practice (Blase & Blase, 1999; Cross & Rice, 2000; Fullan, 2014; Glickman 2002; Gruenert, 2005; NAESP, 2008). However, creating opportunities for teachers and leaders to have such discussions requires an environment of psychological safety and trust (Cross & Rice, 2000; Fullan, 2014). The requirement that evaluators continually collect evidence to support their judgments of teacher performance for accountability purposes makes it difficult to achieve an environment that supports safety and growth (Barth, 1990; Danielson & McGreal, 2000; Gimpl, 2003). When educators know that their evaluator is judging their performance, they are likely to be reluctant to discuss and reflect on their practice freely in the presence of the evaluator (Danielson & McGreal, 2000), thereby limiting the ability of the educational leader to provide the support that may be needed to improve the teacher’s practice.

The role of principal as evaluator and the role of principal as supporter of collegial, constructive, professional dialogue are both essential but conflicting roles (Danielson 2001;
Danielson & McGreal, 2000; Marzano, 2012). Although this conflicting role for principals is identified in the research, there have been few studies that examine the perceptions of principals as they implement an evaluation system that merges supervision and evaluation roles, such as the system implemented in Massachusetts as a result of the Final Regulations on Evaluation of Educators, 603 CMR 35.00 enacted in 2011. This study will examine principals’ perceptions of the effect of the implementation of the Massachusetts Educator Evaluation System on teacher growth and development, their capacity as instructional leaders to support teacher growth and development, and the trusting relationship between principal and teacher.

**Justification for Research**

Research on teacher evaluation is inconclusive with respect to its effect on improved teaching practices and student achievement.

Two empirical studies of note indicate that teachers with higher ratings under comprehensive evaluation systems produced greater gains in student test scores (Milanowski & White, 2004; Pianta 2000). However, Kimball, White, Milanowski and Borman (2004) found mixed results in the relationship between scores in standard-based evaluations and average achievement of the teacher’s students. More recently, several studies have shown a positive relationship between evaluation ratings and student achievement (Heneman & Milanowski, 2004; Holtzapple, 2003). These studies link teacher evaluation scores with student achievement and, therefore, indicate that the teacher evaluation process accurately identifies effective teachers. However, the studies do not address whether or not the evaluation process itself affected the student achievement outcomes or the effectiveness of the teachers.

There have been very few studies focused on comprehensive evaluation systems’ effect on teacher performance, and the studies that have been conducted leave doubt as to the influence
of the evaluation process on increased teacher effectiveness. Kimball (2002) studied three
districts with standard-based evaluation systems. His study indicated that feedback provided to
teachers reaffirmed their performance, but it did not encourage any changes in practice or
professional development. A 2006 policy brief by Henman, Milanowski, Kimball and Odden
concluded that standards-based evaluation systems produced relatively few changes in
instructional practice. A comprehensive study of teacher evaluation conducted in 2009 by
Weisberg et al. found only 43 percent of teachers said evaluation helped teachers improve. In
contrast, one recent study showed promising results for the improvement of teaching through
evaluation: Taylor and Tyler (2012) studied the Cincinnati Public Schools’ Teacher Evaluation
System (TES) to see what effect it had on improvements in performance of experienced teachers
as measured by student achievement gains. They found student achievement in mathematics
increased during the year teachers were evaluated, and that the improvement was sustained in
future years. Student achievement in reading, however, did not show improvement in this study.
With inconclusive evidence that teacher evaluation processes cause improvements to teacher
effectiveness and lead to increases in student achievement, it is important to study how the
mandated practice of evaluation may alter other factors known to have a positive influence on
teacher quality.

Many scholars claim a collaborative school culture fosters greater teacher and student
performance (Barth, 1990; Peterson & Deal, 1999; Fullan & Hargreaves, 1996; Kerstan & Israel,
and experimentation are made relevant to, an integral part of, the occupation and career of
teaching” (p.9), and argued in 1984, that in order for schools to improve the overall quality of
teaching, it was imperative to require teachers to collaborate. In a study of high-poverty urban
schools from across the U.S., Johnson and Asera (1999) concluded one strategy that many of the high-performing schools used was development of structured time to collaborate about instructional issues during the school day. In 2005, a study involving the faculties of 81 schools in Indiana found a positive correlation between a collaborative school culture and student achievement in mathematics and language arts (Gruenert, 2005). Collegiality research conducted by Little (1983) suggests more change in instructional practice occurs when teachers have the opportunity to reflect with a colleague or peer. Danielson (2016) claims that “the main vehicle for improving student learning is professional conversation among teachers and between teachers and administrators” (p.90).

Collaborative cultures can only exist in school environments where there are trusting relationships (Byrk & Schneider, 2003). Many scholars claim that teacher evaluation processes negatively affect the ability of a principal to develop trusting relationships and collegial cultures. Sergiovanni and Starratt (2007) write, “evaluating teachers can dampen, if not betray, the collegiality and trust that is needed for teacher learning to take place” (p.168). Danielson and McGreal (2000) state that, “evaluation appears inconsistent with more supportive non-judgmental demeanor that most people associated with coaching” (p.9). Bullis (2014) found the unintended negative consequences of teacher evaluation in Massachusetts and Florida were teacher morale, teacher-principal relationships and teacher growth.

The influence teacher evaluation has on improved student achievement is unclear, yet it is evident in the research that schools with collaborative cultures see improvements in student outcomes when educators trust one another and work together to improve their practices (Vescio, Ross & Adams, 2008). It is important to examine teacher evaluation more closely to understand
more about the possible effect evaluation has on the trusting relationships that are necessary for collaboration.

**Deficiencies in the Research**

A number of research studies have claimed that traditional systems do not accurately measure or address teacher quality (Toch & Rothman, 2008; Weisburg et al., 2009). Empirical studies examining the correlation between teacher performance as measured through evaluation systems and student achievement outcomes provide little compelling evidence to support a claim that evaluation ratings correlate with teacher effectiveness. Holtzapple (2003) studied the standards-based teacher evaluation system implemented in the Cincinnati public schools and found a positive correlation between teachers’ evaluation ratings and the learning and achievement of their students. However, Milanowski’s (2004) study of Cincinnati found only “small to moderate positive correlations” (p.34). Likewise, White’s (2004) study of Coventry, Rhode Island found only small correlations in reading and no correlation in math. However, there were significant variances in the results throughout the study. Kimball, White, Milanowski and Borman (2004) examined standards-based teacher evaluation in Washoe County, Nevada. Their results were mixed in the predictive value of teacher evaluation ratings on student achievement. They concluded that there was only “tentative evidence for the criterion related validity of the evaluation system and the use of evaluation scores as measures of classroom effects” (p.70). *The Widget Effect*, a report generated for The New Teacher project outlining the findings from a landmark study of teacher evaluation in 12 districts in Colorado, Illinois, Arkansas and Ohio conducted by Weisberg et.al., (2009) found that, “a teacher’s effectiveness – the most important factor for schools in improving students’ achievement – is not measured, recorded, or used to inform decision-making in any meaningful way” (p.3). They
identified many contributing factors related to system design and implementation, including failure to identify clear performance standards, regular monitoring of administrator judgments, inadequate evaluator training, and lack of consequences (including dismissal) for poor performance. After studying leadership decision making in evaluation, Kimball and Milanowski (2009) found that there was “substantial variation in the relationships of evaluators’ ratings of teachers and value-added measures of the average achievement of the teachers’ students” (p.34), and concluded that “unique combinations of evaluator and context factors…govern decision making” (p.63). More recently in the *Measures of Effective Teaching Study* (Ho & Kane, 2013), a positive relationship between evaluation ratings and student achievement were reported. However, the conclusion has been criticized because of the methodology that examined results for groups of teachers rather than for individual teachers.

There have been few studies assessing the effect of evaluation on improvements in teaching practice. Those that have been conducted largely show little evidence of “capacity development of teachers” (Hallinger, Heck & Murphy, 2014, p.22).

Outcomes from studies related to implementation of evaluation systems point to the importance of examining implementation strategies and their influence on evaluation effectiveness. Painter (2000) suggests evaluators’ “attitudes and beliefs” (p. 370) are factors that need further investigation. O’Pry and Schumaker (2012) identified a need to investigate implementation challenges. A more recent study explored the challenges districts face in implementing evaluation systems implemented as a result of the US Department of Education’s Race to the Top program, as well as the strategies those districts developed to address these challenges (Shakeman, Breslow, Kochanek, Riordan & Haferd, 2012). The study identified the
need for research in three main areas: implementation studies, measurement studies and impact studies (Shakeman, Breslow, Kochanek, Riordan & Haferd, 2012).

There has been little research conducted specifically on the implementation of the Massachusetts Educator Evaluation System. This is likely due to the fact that it was first introduced in June 2011. This multi-year, three-phase mandated implementation plan began in the 2013-2014 school year, and it will not be fully implemented until the 2017-2018 school year in all public school districts in Massachusetts. Bullis (2014) has studied the perception of Massachusetts and Florida principals to the effect of teacher evaluation ratings on the teacher evaluation process. His findings suggest that although principals understand the intended results of such systems, there are unintended negative consequences related to teacher morale, teacher-principal relationships, and teacher growth. This research points to the need to understand these negative consequences more precisely to find ways to minimize or eradicate them.

The unintended negative consequences identified in Bullis’ study may stem from the conflicting roles that principals attempt to fulfill. Although the challenge that principals face as they try to balance and merge conflicting responsibilities is identified in the literature (Danielson & McGreal, 2000; Danielson, 2010; Himmelein, 2009; Range, Scherz, Holt & Young, 2011), studies specifically designed to understand how principals perceive this conflict, how it can change their relationships with teachers and how principals address this conflict in their work as school leaders are absent from the literature.

This study’s focus on the experiences of principals implementing the Massachusetts Educator Evaluation System may begin to inform practitioners and researchers about effective implementation strategies that will address teacher growth and development, the ability of principals to support teacher growth, and development of trusting principal/teacher relationships.
Relating the Discussion to Audiences

The insights gained through understanding the experiences of Massachusetts’ public school principals who are working to balance the dual responsibilities of providing summative evaluation ratings for accountability purposes and formative evaluation feedback intended to foster improvements in practice to teachers may inform national, state and local leaders who are interested in supporting the implementation of effective instructional leadership practices.

The national focus on teacher quality and educator evaluation stems from NCLB and RTTT. These federal initiatives prompted state and local officials to implement new teacher evaluation systems aimed at improving teacher practice and student achievement (No Child Left Behind Act, 2001; U.S. Department of Education, 2009). Many such systems have been implemented recently (Darling-Hammond, 2013; Platt & Tripp, 2014; Tomal et al., 2015). It is imperative to evaluate the effects of these systems over time and to make changes if necessary. The outcomes of this study may provide insight about laws, regulations and policies that may need to be changed at the national and state level to improve effectiveness of the teacher evaluation process.

In schools in Massachusetts that must implement the new teacher evaluation system, school district and union leaders will undoubtedly continually evaluate the effectiveness of these new processes. As changes are considered, the findings from this study may provide a lens through which the effectiveness of the Massachusetts Educator Evaluation System implemented in 2011 will be viewed.

This research may also have value at the practitioner level. Principals and those responsible for principal professional development and training in Massachusetts and in other
states that are requiring schools to implement similar systems may gain insights from this research that will affect leadership practices related to teacher evaluation.

The findings from this study will add to the body of research related to comprehensive, standards-based teacher evaluation systems. It will also address the paucity of research related to the Massachusetts Educator Evaluation System, and as such, may be applicable in other states that are implementing educator evaluation systems similar to the Massachusetts system. A deeper understanding of how principals perceive the effect of the Massachusetts Educator Evaluation System on teacher growth and development, on their capacity to support teacher growth and development in their role as instructional leaders, and in the development and maintenance of trusting relationships with their teachers will also inform theories of effective school leadership, particularly those that focus on the importance of trust and collaboration.

**Purpose Statement**

The aim of this study was to explore the experiences of elementary principals who are the sole evaluators of teachers in their schools as they implement the Massachusetts Educator Evaluation System.

**Research Questions**

The research addresses the following three questions:

1. How do principals perceive the Massachusetts Educator Evaluation System as supporting teacher growth and development in theory and as they have experienced it?

2. How do principals perceive the effect of the Massachusetts Educator Evaluation System on their ability, capacity and opportunity to effectively support their teachers’ development?
3. How do principals perceive the effect of the Massachusetts Educator Evaluation System on the development and maintenance of trusting relationships between principals and teachers as they have experienced it?

Significance of the Research

The Common Core Standards for literacy and mathematics adopted by many states, including Massachusetts, place more rigorous demands on students. These increased expectations for students coupled with more challenging student populations create increased challenges for educators. In Massachusetts alone, state-wide data for the 2014-2015 school year indicate that 17.2 percent of students have disabilities; English is not the first language for 19 percent of students, and 9 percent are classified as English Language Learners (ELL). In the 2013-2014 school year, the last year that Massachusetts reported the percentage of low income students, 38.3 percent of students state-wide were considered low income, with 33.6 percent qualifying for free lunch and 4.7 percent qualifying for reduced lunch prices. Overall, in the 2013-2014 school year, 48.8 percent of all students in Massachusetts fell into the category of high needs by virtue of the fact that they have one or more of the factors described above that may affect their academic achievement. Although state-wide averages paint an overall picture of the demographic challenges educators face, these percentages and the challenges they present to educators vary widely from district to district. In many rural districts in Massachusetts, populations of low income, ELL, and special education students are increasing at a rapid pace, making the challenges educators face more pronounced; many teachers do not have a strong skill set to address these needs. In the rural, central Massachusetts regional school district where the researcher is employed, the percentage of low-income students in two elementary schools has increased more than 16 percent in the last five years. The percentage of students who qualify for
free lunch across this same school district has increased from percent in the 2007-2008 school year to 21 percent in the 2013-2014 school year. In the ten-year span from the 2005-2006 school year to the 2015-2016 school year, the percentage of students with disabilities in the district has increased 6 percent. In addition to more rigorous academic standards and increasingly challenging student populations, educational research continues to advance, providing educators with new, more effective practices designed to improve student outcomes (McLaughlin & Talbert, 2006).

Teachers are in the process of making major shifts to adjust their curriculum and instructional practices to address the increased academic demands of the Common Core Standards (Fullan, 2014), the challenges of instructing increasing numbers of high-need students, and the new, more effective approaches to teaching identified in educational research. In many cases teachers must not only change their curriculum, but also deepen their content knowledge to meet the more rigorous Common Core Standards (Tomal et al., 2015). In addition to learning new content, helping a greater number of disadvantaged students meet significantly higher academic demands requires teachers to learn and refine new instructional and assessment strategies. For veteran Massachusetts teachers, many of whom have been teaching the same standards for nearly a decade or more, these changes require significant new learning.

The quality of classroom instruction has a profound influence on student achievement (Haycock & Huang, 2001; Milanowski, 2004; Nye, Konstantopoulas & Hedges, 2004; Odden, 2004; Odden & Wallace, 2003 Odden, Borman & Fermanich 2004; Sanders & Horn, 1994; Sanders & Rivers, 1996). Schmoker (2006) and Marshall (2013) both claim that the quality of instruction is the most significant factor in student achievement. In this era of expected significant change in teaching practice, it has become increasingly difficult for teachers to meet
the demands of their profession at a high level of quality. In order for teachers to maintain the highest level of performance, they must incorporate substantial shifts in pedagogy.

The responsibility to support teachers in improving the quality of their instruction to meet these increased professional demands and ultimately to improve student achievement lies primarily with building-level administrators (Elmore, 2000; Fullan, 2009; Hatch, 2009; Marshall, 2013; NAESP, 2008). This aspect of a principal’s role is typically referred to as instructional leadership. Many argue that it should be a principal’s primary focus (Elmore, 2000; Fullan, 2009; Hatch, 2009; Marshall, 2013; NAESP, 2008). It is clear that principals who are effective instructional leaders have an effect on student achievement that is second only to the impact of instructional quality (Leithwood et al., 2004). However, the increased demands placed on principals in this time of school reform and rapidly changing student demographics has created a role that some see as unreasonable (Fullan, 2014; Grubb & Flessa, 2006) and difficult to effectively fulfill (Copeland, 2001; Danielson & McGreal, 2000; Fullan, 2014).

One tool at principals’ disposal that is designed to support the role of instructional leadership is a supervision and evaluation system (Marshall, 2013). One of the primary goals of effective supervision and evaluation systems is the improvement of teacher practice and, in turn, increased student achievement (Peterson, 2000). Traditional teacher evaluation systems have fallen short in achieving the desired outcome, however (Peterson, 2000). Recent federal and state educator evaluation reforms mandated by NCLB and RTTT aim to address the shortcomings of traditional evaluations systems with development of systems that are more effective at assessing teacher effectiveness and improving teacher practices (U.S. Department of Education, 2009).
Massachusetts’ new Educator Evaluation System, developed and adopted in compliance with RTTT, is being implemented in phases. Phase I began in 2011, and full implementation is expected in the 2017-2018 school year. The Massachusetts system requires principals to annually judge and rate each teacher’s overall performance, as well as their performance in each of four subcategories called professional standards (Curriculum, Planning and Assessment; Teaching All Students; Family and Community Engagement; Professional Culture). Evidence collected by the evaluator throughout the school year is used to inform the end-of-year formative or summative judgments about teacher performance. The resulting ratings are submitted to the Department of Elementary and Secondary Education. When fully implemented, the system will also include ratings for each teacher based on trends in student growth measures.

Like other standards-based comprehensive evaluation systems, the Massachusetts Educator Evaluation System is designed to fulfill two roles: measurement of teacher performance for accountability purposes and employment decisions and improvement in educator performance.

The implementation of the new evaluation system, which is designed to be an effective tool to support the primary function of instructional leadership, requires a significant shift in practice for most principals. The new system requires a significantly longer time commitment because of more frequent observations of teachers, more detailed written feedback and attention to more than 30 standards, indicators and elements of practice. Implementation of this system is yet another burden on already overwhelmed principals in Massachusetts who must simultaneously manage significant changes in curriculum and assessment because of the adoption (in 2011) of the Common Core Standards for literacy and mathematics, and the pending shift in standardized testing from the Massachusetts Comprehensive Testing System (MCAS) to
MCAS 2.0, a more rigorous state-wide test based on the Partnership for Assessment of Readiness for College and Careers (PARCC) test.

In addition to the challenges of shifting their professional practices to meet the mandated implementation of the Massachusetts Educator Evaluation System, many principals also struggle to find a balance between the accountability function of the evaluation tool and the professional growth function of the tool. To fulfill the accountability expectations of evaluation, the principal must make judgments about the performance of each teacher supported by evidence collected throughout the school year. However, supporting professional growth requires “a more supportive non-judgmental demeanor that most people associate with coaching” (Danielson & McGreal 2000, p.9). The judgmental stance required to fulfill the accountability expectations of summative evaluation can prevent the development of a trusting relationship between a principal and a teacher which is necessary for meaningful collaborative conversations that characterize the effective supervisory role (Tomal et al, 2015). Many researchers have identified these as conflicting roles, further complicating the overwhelming job of school principal (Cogan, 1973; Danielson & McGreal, 2000; Nolan, 1997; Popham, 1988; Sergiovani, 2007).

Evaluation, a process designed to support teacher improvement, should help principals lead, rather than making their jobs more difficult. Developing a deeper understanding of the principals’ perceptions of the effect that implementing the Massachusetts Educator Evaluation System has had on their capacity to support teacher growth and development, particularly as it relates to development and maintenance of trusting relationships with teachers, may allow potential solutions to emerge that will ultimately help principals carry out their primary responsibility to improve student achievement (Ross & Gray, 2006) more effectively. It is important to understand how leaders can most effectively and efficiently perform this role. The
more that is understood about effective instructional leadership and the effective use of evaluation and collaborative structures to improve teacher performance, the greater the potential to improve student outcomes locally, state-wide and across the nation.

**Positionality Statement**

It is important for a researcher to examine his or her positionality with respect to the problem of practice. This examination allows for awareness of potential theories or preconceptions held by the researcher that might influence data collection and analysis (Maxwell, 2005).

Throughout my twenty-three-year career in education, I have consistently been involved with the practice of teacher evaluation, first as a teacher, then as building-level administrator responsible for evaluating teachers, and finally as a central office administrator responsible for oversight of teacher evaluation at the district level. The teacher evaluation processes and procedures I have encountered have varied widely, but from my perspective, the outcomes have been much the same: little change in teaching practice.

In the first five years of my career as a teacher, I worked in three different school districts. In each of those five years, my experience with teacher evaluation was different. My first experience with evaluation was a single announced observation of a partial class period, followed by a written narrative report left in my mailbox with a note to make an appointment with the evaluator if I wanted to discuss anything. The written report did not include a single judgment about the quality of my instruction, nor a single suggestion as to improvement. In my second year of teaching in a new school district, the principal conducted one announced observation of a full class period. The written narrative outlined everything I said and did during the class period and concluded with a statement that described me as an outstanding teacher. A
follow-up discussion with the evaluator did not take place. The next year I remained in the same school, but was evaluated by the assistant principal who met with me prior to the announced observation and encouraged me to try something new when he observed me so that he could provide me with feedback that would improve the lesson. Despite a great deal of apprehension, I did try an instructional strategy that I had not used before when I was observed the following week. The post-conference and observation write-up provided many suggestions for improvement, but also included praise for taking a professional risk to improve my practice. The next two years were spent in a third school, where unannounced observations were required twice a year. In my first year, the principal appeared at my door in early October and sat through my class taking notes. I never received any communication about the observation nor did the principal ever evaluate me again in the two years I worked in the school. I realized that he thought I was a good teacher when I asked for a letter of recommendation at the end of my second year.

I spent the next several years of my career working in a fourth school where I had multiple evaluators due to changes in administrative staffing. With each new evaluator, I experienced yet a different approach to evaluation. Overall in my 13 years as a teacher, the one consistent message I received from evaluators was that I was a good teacher. Despite the fact that my own reflection on my teaching practice and the associated learning outcomes of my students indicated that I had plenty of room for improvement, I received very little evaluative or other feedback that prompted me to consider changes to my professional practice that might take me from good to great.

When I became an assistant principal and evaluator, I vowed that I would work to make evaluation meaningful for my teachers. Although I did try to provide opportunities for
collaborative discussions and meaningful dialogue about lessons that I observed by way of mandated appointments for observation and evaluation, I never felt completely successful in shifting the culture to one in which true collaboration between the evaluator and the teacher resulted in improvements to teaching practice.

My first experience as a central office administrator brought me to a district where a new model for teacher evaluation was under development. There, teachers, building-level administrators and central office administrators were working collaboratively to define a teacher evaluation system that focused on teacher growth and development. In the pilot years, this system involved a somewhat traditional system of observation and written feedback to teachers who were in their first three years of employment, but for those who had earned professional teaching status, many options for professional growth and development were offered. Among the options for professional-status teachers were action research projects, professional development courses, and peer observation. Feedback from teachers and principals during the pilot phase were very positive with respect to the empowerment of teachers to improve their professional practice and to collaborate with their principal. I was encouraged by this model for evaluation and felt it held great promise for shifting the focus of evaluation to continuous improvement. The Massachusetts educator evaluation regulations were enacted shortly after this district’s pilot system had been fully implemented, and because it did not meet the demands of the new regulations, plans were being made to change the process as I left the district to take my current position.

As assistant superintendent of schools I have the primary responsibility for implementing the Massachusetts Educator Evaluation System in my district. When I heard that the regulations were under development and that a team of educators, administrators, consultants, and
researchers were developing recommendations with research-based best practices, I held great hope that the new Massachusetts system would provide a consistent, manageable and meaningful process that would help all educators to continuously improve their professional practice. In year four of implementation, I am seeing many educators struggle with various aspects of the System. I remain hopeful, but concerned, that when teacher evaluation in Massachusetts is fully implemented, it will meet all of its goals.

Given the extensive and varied experiences that I have had with teacher evaluation throughout my career in education, it has been necessary for me to remain mindful of my bias in each stage of the research study. I have taken care to remain impartial by refraining from imposing my viewpoint as I collected and analyzed the data. The data speaks for itself in the presentation of findings of the study.

**Theoretical Framework**

The theoretical framework for this study was the theory of relational trust in schools. This theory was developed by Anthony Byrk and Barbara Schneider (2002) through research they conducted over a 10-year period in the Chicago Public Schools. The theory of relational trust is known as a seminal “specialized trust theory related to schools” (Forsyth, Adams & Hoy, 2011, p.73). In their theory, Byrk and Schneider posited that interpersonal social exchanges in schools are unique and are a “valued outcome in their own right” (Byrk & Schneider, 2002, p.19). Relational trust in schools “requires that the expectations held among members of a social network or organization be regularly validated by actions” (Byrk & Schneider, 2002, p.21). Moreover, it is essential that there is a perception that actions are conducted respectfully and for the right reasons (Byrk & Schneider, 2002).
Byrk and Schneider (2002) conceptualized the theory of relational trust on three levels. On the intrapersonal level, trust is “rooted in a complex cognitive activity of discerning the intention of others” (Byrk & Schneider, 2002, p.22). On the interpersonal level, the determination of these intentions is influenced by the established norms within the school community (Byrk & Schneider, 2002). At the organizational level, relational trust results in “more effective decision making, enhanced social support for innovation, more efficient social control of adults’ work, an expanded moral authority to ‘go the extra mile’ for the children” (Byrk & Schneider, 2002, p.22).

Byrk and Schneider (2002) theorized that four elements — “respect, competence, personal regard for others, and integrity” (p.23) — interact to contribute to trusting relationships. All elements must be present for a trusting relationship to exist.

Respect is characterized by interactions “where individuals intently listen to each other and in some fashion take others’ perspectives into account in future action…this fosters a sense of personal esteem for participants and cements their affiliation with each other and the larger institution” (Byrk & Schneider, 2002, p.23).

The second element, competence, is described by Byrk and Schneider (2002) as related to a discernment of an individual’s ability to achieve the desired outcomes. They noted, however, that competence is difficult to assess accurately in school settings because of the complexity of the roles that educators hold. Byrk and Schneider therefore posited that relational trust can exist in school environments where competence is varied.

“Interpersonal trust deepens as individuals perceive that others care about them and are willing to extend themselves beyond what their role might formally require in any given
situation” (Byrk & Schneider, 2002, p.25). Byrk and Schneider (2002) described this personal regard for others at the third element necessary for relational trust.

One is thought of as having integrity “if there is consistency between what they say and do” (Byrk & Schneider, 2002, p.25). Integrity, therefore, allows one to rely on the consistency of behavior of another individual. Integrity is the final element necessary for relational trust.

Byrk and Schneider (2002) claimed that relational trust plays a significant role in moderating the “sense of uncertainty and vulnerability that individuals confront” (p.33) as they face school reform movements that often demand significant change in a short period of time. Their research indicated that “reform is likely to progress faster in high trust contexts because participants are more able to coalesce around a plan of action” (Byrk & Schneider, p.33), and it is more likely that reform initiatives will diffuse broadly across the school because trust reduces the sense of risk associated with change (Byrk & Schneider, 2003, p.43).

The theory of relational trust in schools developed by Byrk and Schneider is grounded in the theory of social capital developed by James Coleman. Coleman (1988) stated that social capital “comes about through changes in the relations among persons that facilitate action” (p.100) and “the achievement of certain ends that in its absence would not be possible” (p.98). Coleman (1998) further asserted that closure within a social network makes it more likely that social capital will exist. This interconnectedness of individuals allows for the establishment of expectations and obligations between people, which are essential to the establishment of social capital. Coleman (1988) claimed that when one person does something for another, there is an expectation that there will be a reciprocation of some sort in the future. This expectation of reciprocation is thought of as an obligation to repay the other party. Trust is essential to social
capital in the sense that there must be trust that the obligation to reciprocate will be honored by
the receiving party.

Rational choice theory also influenced the development of relational trust theory. The
development of trust according to this theory depends on a decision made by an individual to
interact with someone else involving a factor of risk. Such decisions can be influenced by
previous interactions, known reputation, or common social relationships (Byrk & Schneider,
2002). A third influence on the work of Byrk and Schneider was group theory, which claims that
sustained social interactions are valued by individuals and this value influences the propensity of
an individual to engage in future interactions. A fourth lens considered by Byrk and Schneider is
the moral-ethical aspect of trust, whereby trust is built on a mutual expectation to do what is right
(Byrk & Schneider, 2002).

**Relational trust applied to role of principal.** Principals in Massachusetts are leading
schools in the midst of change. Mandated reform efforts are numerous and student needs are
growing. Principals must incorporate effective leadership strategies in every aspect of their day-
to-day professional practice to ensure that their schools can continuously improve outcomes for
students. The development of trusting relationships is an element of leadership practice that
supports successful implementation of school reform (Byrk and Schneider, 2003). Relational
trust is built in “day-to-day social exchanges within a school community” (Byrk & Schneider,
2003, p.41). A number of sustained social interactions between principals and teachers take
place within the teacher evaluation process. In order to meet the requirements of the
Massachusetts Educator Evaluation System, principals must frequently examine and report on
data collected relative to the performance of each teacher in his or her school. The processes of
data collection and data reporting require regular social exchange between a principal and
teacher. The nature of this social exchange has the potential to affect development or maintenance of relational trust. The essential elements of relational trust, respect, personal regard for others, competence and personal integrity (Byrk and Schneider, 2002) are all at risk in social interactions related to teacher evaluation.

Respect and personal regard for others require principals to “acknowledge the vulnerabilities of others, actively listen to their concerns, and eschew arbitrary actions” (Byrk & Schneider, 2003, p.43). When a principal judges a teacher’s performance and assigns a series of ratings indicating the teacher’s degree of proficiency against specific descriptors of teaching practice as is required by a standards-based evaluation system, it is entirely possible that a teacher might not expose his or her vulnerabilities for fear that knowledge of a teacher’s weakness might cause a principal to assign a lower performance rating (Danielson & McGreal, 2000). Without teachers sharing their vulnerabilities it would be difficult for a principal to demonstrate the ability to acknowledge them. Actively listening to concerns of teachers through the process of evaluation, especially during the initial implementation years, when concerns about the process and ratings are significant could be taxing on a principal’s time. The time required to evaluate teachers with a standards-based system is much greater than with traditional systems. Principals already overwhelmed with jobs that seem “too big” (Fullan, 2014; Grubb & Flessa, 2006), may have difficulty finding the time to actively listen to the concerns of their teachers. The teacher evaluation ratings that principals assign may well be seen as arbitrary actions by their teachers, especially when it has been the norm for many years to not identify low-performing teachers through the evaluation process (Weisberg et al., 2009). Teachers who have been praised in their evaluations for years may now receive feedback that includes
suggestions for improvement. This shift for teachers might be perceived as illogical on the part of the principal, thereby undermining the establishment of trust.

To be perceived as competent, a principal must demonstrate his or her ability to successfully perform a complex set of responsibilities. Competence is likely to be questioned in the process of implementing a standards-based evaluation system for a variety of reasons. Initially, implementation of such a system requires all educators in a school to learn many new processes, guided by specific timelines. The System implementation in Massachusetts is a multi-year process, therefore most principals learn the implementation along with their teachers. It is unlikely that most principals exude confidence in their role as evaluators in the initial implementation years, and many may even make mistakes. Furthermore, it is very likely with implementation of a standards-based evaluation system that a considerable number of teachers will receive ratings that are incongruent with those received in previous years. Most teachers have previously been given feedback through their evaluation cycle that was overwhelmingly positive (Weisberg et al., 2009) and now many may receive ratings and recommendations that indicate areas for improvement. This change is likely to cause some to question the competence of their principal as an evaluator, serving to erode trust.

Personal integrity, the final element necessary for relational trust is also seemingly at risk for principals implementing a new standard-based evaluation system. The expected shift in evaluation ratings for many teachers outlined previously could cause them to question the integrity of their principal, especially if he or she had previously provided feedback that was overwhelmingly positive.

Because of the importance of the effect of relational trust on school improvement, Byrk and Schneider (2002) state, “we constantly need to ask whether any new initiative is likely to
promote relational trust within school communities or undermine it” (p.144). Implementation of the Massachusetts Educator Evaluation System is a new initiative aimed at improving teacher practice and student achievement. Examination of this process through the lens of relational trust will allow insight as to its potential long-term effect on school improvement initiatives and leadership practices.

**Conclusion**

Principals’ perceptions of the implementation of the 2011 Massachusetts Educator Evaluation System is the purpose of this research. This study will examine the perceptions of the effects of the Evaluation System on teacher growth and development, the ability and opportunity for principals as instructional leaders to support teacher growth and development and trusting relationships between principals and teachers. The outcomes of this study may inform state, district and union leaders who seek to improve teacher evaluation processes. In addition, it may inform those responsible for principal professional development.

This chapter included an introduction to, justification of and significance of the research problem, the research questions, the positionality of the researcher and the theoretical framework for the study. The following chapter provides a review of the current literature related to the topic of the study.
Chapter II: Literature Review

Overview

The literature reviewed in this chapter falls into the following categories: teacher evaluation, teacher impact on student learning, effective practices for improvement of teacher performance, trust in schools and the role of the school principal.

The literature review begins with a historical overview of teacher evaluation. The historical overview highlights various schools of thought related to supervision and evaluation that have influenced evaluation practices through the years. The two primary schools of thought related to teacher evaluation are explored in this category, and current trends influenced by research and federal mandates will be explicated as well. The following section will focus on the research related to the impact of effective teaching on student achievement. The literature review then returns to the discussion of supervision and evaluation and examines the research related to the effect of these practices on improvement of teaching. Through this exploration it becomes evident that one element central to effective supervision and evaluation is a trusting relationship between the supervisor and the teacher. As such, the theory of relational trust as well as research related to the impact of trusting relationships on school improvement will be presented. Finally, the role of school principal will be explored because principals are responsible for student achievement and are mandated to use supervision and evaluation practices for accountability purposes and for the improvement of teaching.

Teacher Evaluation

**Historical perspective.** It is important to begin with a brief review of the history of teacher evaluation because there are links between current supervision and evaluation practices
and those used in the past. Marzano, Frontier and Livingston (2011), Nolan and Hoover (2011) and Saundra Tracey (1995) published historical overviews that have guided this brief summary.

In the 1700s, teaching was not considered a profession. Education was locally controlled and teachers were considered community servants. The primary goals for students at the time were the ability to read scripture and to conform to the moral standard of the community. Community leaders, especially the highly educated clergy, were teachers’ supervisors through the early 1800s. Supervision focused primarily on teachers’ fulfillment of their role as determined by community-based priorities. Supervisory practices were inspectional in nature, and supervisors exhibited strict control over instructional practices, materials, curriculum (Tomal, Wilhite, Phillips, Sims & Gibson, 2015), and school facilities (Nolan & Hoover, 2011). Tracy (1995) characterized these early years of teacher evaluation as the community accountability phase.

In the mid-to-late 1800s, teacher evaluation entered the professionalization phase (Tracy, 1995). At this time, city school systems grew and became more complex. State and county superintendents, and eventually principals (principal-teachers) assumed the role of teacher supervision as the community recognized the need for content and educational expertise to foster improved instruction. While the inspection of teacher practice remained the primary mode of evaluation as had been the case in the early part of the 19th century, the quality of instructional practices of teachers became the focus (Blumberg, 1985). At this time however, the prevailing belief system was that improvement of educational quality came primarily from removal of teachers who were inadequate (Nolan & Hoover, 2011).

The work of Frederick Taylor influenced supervisory practices in education in the early 1900s. Tracy (1995) identified this as the scientific phase. Taylor employed a scientific
approach for measurement of specific behaviors for factory workers intended to improve production. Edward Thorndike believed that a more scientific approach to education, based in measurement was warranted and shared this theory in the educational community. In 1916 Ellwood Cubberly used Thorndike’s theory, viewing children as products of an educational system, to develop principles that school administrators should use to measure productivity of teachers. William Wetzel later expanded on Cubberly’s work and included measures of student learning as well as specific instructional strategies to determine teacher effectiveness (Marzano et al., 2011). The goal of such measurements was to study effective and efficient teachers to determine the behaviors that they exhibited, and to measure other teachers’ behaviors against those known to be effective (Nolan & Hoover, 2011). This phase resulted in checklists and rating scales as the common tools for teacher supervision (Nolan & Hoover, 2011; Tracy, 1995), and the practice of teacher evaluation remained primarily as it had been since the early 1800s: that of inspection of teacher practice.

A dramatic shift occurred in the 1930s and lasted through the 1950s. Influenced by development of the social sciences, particularly motivation theory that espoused the positive relationship between attention paid to workers and productivity, the focus for supervision shifted to the improvement of the social and psychological well-being of teachers. Tracy (1995) identified this as the human relations phase. John Dewey also influenced thinking at this time. Dewey’s views of progressive education fostered a sense that the role of teacher supervisor and helpful colleague were conflicting roles (Tomal et al., 2015). “Progressive supervision, while downplaying teacher evaluation, emphasized collaboration, group processes, inquiry and experimentation” (Nolan & Hoover, 2011, p.3). It was thought that assisting teachers would improve instruction and, as such, supervisors concentrated on improving working conditions and
positive relationships with teachers. Often, this meant classroom observations were not conducted as frequently or with rigor for fear that providing such feedback to teachers could negatively impact the supervisor-teacher relationship. Nolan and Hoover (2011) noted however, that although progressive supervision (also called democratic supervision) took hold during this time period, former practices of teacher evaluation consisting of inspection and rating of teacher practice continued to exist and claimed that this time period was the “beginning of an ongoing conflict between the dual role of supervisor as ‘teacher evaluator’ and ‘helpful colleague’” (p.4).

In the late 1950s, a sense of urgency to improve the quality of the educational system in the U.S. developed. Prompted by the launch of Sputnik in 1957, the importance of increasing student achievement in mathematics and science shifted the role of supervisor back to that of inspector through the use of evaluation rating systems that monitored implementation of innovative teaching practices (Nolan & Hoover, 2011) and incorporated student scores on standardized tests (Tracy, 1995). Data collected and reported to teachers through use of these rating systems was intended to promote improvement in teacher practice (Tracey, 1995). Although systems focusing primarily on development of collegial relationships continued to exist, they were not the predominate system for teacher supervision at this time (Nolan & Hoover, 2011).

In the 1960s Morris Cogan from Harvard University developed a clinical supervision model (Tracy, 1995). Cogan’s model, which was further developed by Robert Goldhammer in 1969, encouraged increased teacher-supervisor interaction in pre- and post-observation conferences (Goldhammer, 1969) to promote collaborative analysis of teacher performance for continuous professional learning and teacher performance improvement (Cogan, 1973). Goldhammer’s intent was reflective dialogue that would lead to better instructional strategies.
Cogan and Goldhammer both believed that the role of supervisor as helper of teachers and the role of evaluator as inspector of teachers were incompatible (Nolan & Hoover, 2011). Still, the five-step process that Goldhammer developed to foster the role of supervisor as collaborative colleague soon became a structure used exclusively for evaluation of teachers rather than a specific route to enhanced instruction through professional inquiry (Marzano et al., 2011). As a result, clinical supervision never took hold in practice (Nolan and Hoover, 2011).

The work of Madeline Hunter influenced supervisory practices in the early 1980s. Hunter (1984) developed a seven-step framework for an effective lesson, called mastery teaching. Hunter’s mastery teaching model came to be widely used as a checklist to gauge lesson effectiveness in teacher performance evaluations during this time period (Marzano et al., 2011). Hunter’s emphasis on the importance of identification of research-based effective instructional strategies and professional development to support the continued growth of teaching practice had a significant influence on supervisory practices in the 1980s (Hunter, 1984), further reinforcing the role of supervisor as evaluator of teacher practice.

During the mid-1980s others began to advocate for differentiated supervision and evaluation for educators. Glatthorn focused on supervision linked to a teacher’s career goals and encouraged teacher voice in development of their professional growth (Marzano et al., 2011). Thomas McGreal also advocated for differentiation, but he believed that teacher experience should provide the basis for differentiated approaches, and Carl Glickman emphasized linking organizational goals and teacher needs in the supervisory process (Marzano et al., 2011).

In the late 1980s and early 1990s there was yet another shift in evaluation practices prompted in part by the publication of a report by the National Commission on Excellence in Education. The report, A Nation at Risk: The Imperative for Educational Reform, made several
recommendations related to teacher quality and evaluation. Among the recommendations were that teachers should be highly educated and competent, and that evaluation systems including peer review should be employed to inform employment decision and to reward superior teachers with salary and promotions (National Commission on Excellence in Education, 1983). In addition, a study conducted for the RAND Corporation by Wise, Darling-Hammond, McLaughlin and Bernstein (1984) identified the fact that evaluation systems used in most school districts were not well developed or effectively implemented, and that teacher evaluation on the whole in the U.S. was, “an under-conceptualized and underdeveloped activity” (p.21). One recommendation from that study was that “school districts should hold teachers accountable to standards of practice that compel them to make appropriate instructional decisions on behalf of their students” (p.80). Whereas the major focus up to this point had been evaluating teacher actions, educators were now becoming more interested in the link between teaching practice and student learning.

Charlotte Danielson (1996) developed a supervision and evaluation model in the mid-1990s that delineated 76 elements of essential knowledge, skill, and dispositions for successful teaching. Danielson’s Framework for Teaching described each element at four levels of performance (unsatisfactory, basic, proficient, and distinguished), and as such, addressed many of the shortcomings identified in the RAND report. Danielson recognized the complexity of teaching, the need for a common language about teaching, and the value of self-assessment and reflection. She incorporated these into her model. Evaluation practices under Danielson’s model continued to involve primarily classroom observation of teacher practice.

Supervision and evaluation practices continue to evolve in an effort to find a perfect system, most recently these efforts have been fueled by federal initiatives such as No Child Left
Behind and Race to the Top. Current trends in teacher evaluation encourage looking beyond classroom observation to multiple sources of data and multiple strategies for gathering data are now being developed and implemented across the United States (Darling-Hammond, 2013; Steinberg & Donaldson, 2014).

**Supervision vs. evaluation.** The terms supervision and evaluation have often been used interchangeably as reflected in the preceding historical summary. Still, there is and continues to be an ongoing debate among scholars, researchers and practitioners on the question of the purposes and practices of supervision and evaluation as well as the compatibility of supervision and evaluation processes. Beginning in the 1930s and continuing today, scholars, researchers, and practitioners continue to be divided on this issue. What is universally accepted is that the ultimate goal of both supervision and evaluation is to ensure that schools employ teachers that are effective in their role.

To fully understand the ongoing debate about supervision and evaluation, it is important to explore the concepts of supervision and evaluation within the literature. Some use the terms interchangeably, and others differentiate between them. The most prominent distinction is the differentiation in purpose of evaluation and supervision.

Evaluation is used to make judgments about a teacher’s practice (Nolan & Hoover, 2011; Sergiovanni & Starratt, 2007). Such judgments are used for purposes of accountability, employment decisions, and rewards for top performers. In some cases, evaluative judgments are used to determine various levels of support and oversight in upcoming evaluation cycles (Nolan, 1997; Papay, 2012; Sergiovanni & Starratt, 2007). This type of evaluation is often termed summative evaluation (Sergiovanni & Starratt, 2007). Evaluation is thought of as meeting the
goal of ensuring that schools employ effective teachers by identifying and remediating less effective teachers and by removing ineffective teachers.

Supervision, sometimes also called formative evaluation or formative assessment, is most often described as being used for the “purpose of enhancement of teachers’ pedagogical skills, with the ultimate goal of enhancing student achievement” (Marzano et al., 2011, p.2). Nolan (1997), likewise defined supervision as “an organizational function concerned with promoting teacher growth and leading to improvement in teacher performance and greater student learning” (p.100). Feedback provided to teachers through the formative process is intended to help teachers change their practices, and is not intended to be used to make judgments about overall performance. Tomal et al. (2015) emphasized that effective supervision involves strong professional relationships between supervisors and teachers. The desired outcome of effective supervision as defined by Tomal et al. (2015), is “increased teacher and student learning within a climate of trust and safety” (p.9). Supervision ensures that schools employ effective teachers by helping all teachers engage in professional growth to improve their practices.

The distinction between evaluation and supervision is often confused or overlooked (Darling-Hammond, Wise & Pease, 1986; Hazi, 1984; Sullivan & Glanz, 2000; Zepeda & Ponticell, 1998). Holland and Garman (2001) wrote, “supervision has become indistinguishable from evaluation of teaching in the minds of most teachers, many administrators and virtually all politicians” (p.98). In a 2004 study, Ponticell and Zepda, found that all teachers and the majority of principals defined supervision as evaluation. Principals who equated the two functions saw evaluation as fulfilling a legal mandate and supervision as making sure that teachers complied. The remaining 11 percent of principals, believed supervision should focus more on helping teachers improve instruction, but indicated that they were confused by the distinction between
the two roles. Many factors contribute to this; primary among them are the fact that the same person (often the school principal) performs the role of supervisor and evaluator and that school districts and states often formally link supervision and evaluation processes under one system and mandate implementation in such a way that they cannot be separated.

Many scholars claim that the role of supervisor and the role of evaluator conflict and should not be carried out by the same individual (Cogan, 1973; Nolan, 1997; Popham, 1988). Showers (1985) wrote, “the norms of coaching and evaluation practice are antithetical and should be separated in our thinking as well as in practice” (p.46). Conflicts identified include the fact that while acting in the role of supervisor, one should foster inquiry-based collaborative discussions between a teacher and the supervisor about teaching practices that could or should be improved. Fostering such discussions requires the supervisor to take a non-judgmental stance for the teacher to feel comfortable enough to expose and discuss inadequacies and generate possible action plans to improve open the areas of identified need (Nolan & Hoover, 2011). Some theorists posited that if the supervisor is also the evaluator, teachers are much less likely to share openly and truly collaborate with the supervisor for fear that the supervisor will use information shared as part of the summative evaluation process as well (Danielson & McGreal, 2000, Peterson, 2000). Popham (1988) wrote that formative evaluation will be compromised in this situation because evaluators will not have access to accurate reports of teacher weakness, and therefore will not be able to support teacher improvement. Popham (1998) further claimed that not “under any circumstances” (p.271) should information gained during formative evaluation be used for the purposes of summative evaluation. Showers (1985) maintained that when there is an imbalance of power in a relationship between a supervisor and a teacher, “it is improbable that
the process will create a climate conducive to learning and growing on the part of the teachers” (p.47).

Sergiovanni and Starratt (2007) used supervision as a universal term, but distinguished three types of supervision: supervision for formative evaluation, supervision for summative evaluation and supervision for administrative evaluation and decision. They wrote that whether these roles are performed by one person or by individual members of a school team, there should be a clear and formal distinction between the roles, purposes and procedures that are understood by all stakeholders in the school. Nolan and Hoover (2011) likewise asserted that there are seven fundamental differences between supervision and evaluation, and “that the two processes must be separated,” but that “the same individual can carry out both functions effectively as long as the processes are different” (p.12). Wise et al. (1984) concluded that “a single teacher evaluation process can serve only one goal well … a single process cannot meet the goals of judging and improving all teachers” (p.70). Tschannen-Moran and Tschannen-Moran (2011) recommended that “evaluation and coaching proceed on separate but complimentary tracks” (p.13).

Danielson and McGreal (2000) described evaluation processes as formative and summative, the former for professional growth and the latter for quality assurance, but said that “evaluations are judgments; they are assessment of teaching” (p.9), and that “in interactions focused on learning, the supervisor’s role is more one of coach and mentor, rather than one of judge” (p.9). They acknowledged a conflict between the roles, but claimed that in a well-designed evaluation system, they can be effectively merged. Tomal et al., (2015) concurred that supervision and evaluation are distinct but related functions that merge when “supervisors make judgments on the quality of teaching based on what they have seen in classrooms” (p.11). Like Sergiovanni and Starratt (2007), Tomal et al., (2015) further distinguished formative and
summative functions within the realm of evaluation, with the formative function described as
giving a teacher “directed and meaningful feedback on what was done well and what needs to be
improved” (p.11) whereas the summative function is providing “a performance rating for
purposes of continued employment, remediation, or dismissal” (p.11).

Some claim that supervision and evaluation processes are inextricably linked (Ribas,
2011; Tomal et al., 2015), and although summative evaluations provide a judgment about a
teacher’s overall performance, at the end of a period of time (usually called the evaluation cycle),
the data gathered during the formative evaluation process (Ribas, 2011) or during supervision
(Tomal et al., 2015) should inform the summative judgment. The notion that the accuracy of
summative judgments of teacher performance are more valid when informed by formative data is
claimed that the process of supervision, although intended as non-judgmental, is inseparable
from evaluation because

   Every type of data that an evaluator possesses about a teacher, whether acquired in a
   supervisory, collegial or other way, is part of the total body of data the evaluator uses,
   either consciously or subconsciously, to make his or her final evaluative judgments (p.2).

The literature is mixed with regard to definitions of supervision and evaluation functions
and processes (Nolan & Hoover, 2011). This contributes to educator confusion that causes
highly variable implementation of both functions. Holland and Garman (2001) claimed that that
supervision and evaluation are “in a state of crisis” (p.98) due to educator’s “willingness to live
with the confusion of these often unrelated aspects” (p.98). “The failure to resolve the dilemma
between teacher supervision and evaluation has created major difficulties in the practice of
instructional supervision since the 1920s” (Nolan & Hoover 2011, p.5).
The overarching consistent message, however, is that the primary role that rests on the shoulders of principals is to ensure that the highest possible quality teaching takes place in their schools so that student learning is maximized. Two strategies are used for ensuring teacher quality: 1. building teacher professional capacity by providing feedback about teaching practice and opportunities for collaborative discussion to foster professional growth and development (supervision or formative evaluation), and 2. rating a teacher’s performance for the purposes of making employment decisions, providing accountability data to stakeholders (parents, community members, district offices and state departments of education), and providing remediation for underperforming teachers (evaluation). The primary inconsistency in the literature is the determination of the compatibility of these roles both within one system and for one individual principal.

**Current Trends in Teacher Evaluation**

**Standards-based teacher evaluation.** Standards-based evaluation systems have gained popularity in response to the overwhelming sentiment that teacher evaluation in the United States is not effective (Papay, 2012). Standards-based systems are those designed to measure teacher performance against a set of standards that outline practices for effective instruction (Papay, 2012). Most such systems are developed based on a comprehensive set of research-based standards for performance that are described in detailed rubrics that differentiate levels performance for each standard (Papay, 2012). Data is collected by trained evaluators who observe and interact with teachers throughout the evaluation cycle and then use the data as evidence determine the teacher’s performance rating on each of the standards based on the descriptors in the rubric (Papay, 2012). One such system which has been widely used is Danielson’s (1996) Framework for Teaching, which describes four levels of teaching
performance (unsatisfactory, basic, proficient, and distinguished) on each of 76 research-based elements of teaching quality.

Standards-based systems have faced criticism for their reliance on human judgment. Some claim that observation-based evaluations can be biased, and that such systems lack reliability because evaluators must make judgments based on a limited amount of data (Papay, 2012). It is also possible that different evaluators within the same evaluation system could interpret the rubrics differently and thereby rate teachers who perform similarly with differing ratings (Papay, 2012).

Validity within a standards-based system has also been raised as a concern. Is it valid to assume that when a teacher uses the instructional practices outlined by the evaluation standards that they are actually effectively teaching students? Many researchers have tried to validate standards-based systems by researching the relationship between standards-based teacher ratings and value-added test scores for students. A 2015 study conducted by Steinberg and Garrett found that average teacher performance scores generated through the use of Danielson’s Framework for Teaching were highly correlated with student performance in English language arts and mathematics. These researchers also noted year–to-year variability in teacher’s observation-based scores, suggesting that there are limitations to observation-based measures of teacher practice: “Our results lead us to question the wisdom of solely relying on observational protocols like the FTT [Framework for Teaching] for the purposes of evaluation teachers in a formal system” (Steinberg & Garrett, 2015, p.239). In a subsequent study these researchers found that incoming student performance significantly and substantively affected standards-based measures of teacher performance.
**Value-added teacher evaluation.** The federal RTTT grant program required states to incorporate multiple measures of teacher performance, including student learning outcomes into their evaluation systems (U.S. Department of Education, 2009). This requirement prompted many states to include student achievement data as one measure of student performance (Steinberg & Sartain, 2014). Many states refer to these measures as value-added measures (VAM); however, they are also referred to as student growth percentiles (SGP). Still other states rely on student learning objectives (SLO) for measures of student achievement. Typically, value-added scores are calculated based on student performance on standardized assessments that are administered state-wide (Steinberg & Donaldson, 2014). These standardized assessments do not apply to all content areas, and therefore cannot be used in the evaluation of all teachers. To address the lack of value-added data for approximately of 70 percent of teachers, many states use a combination of VAM, SGP and SLO in the evaluation systems that have been implemented since RTTT (Steinberg & Donaldson, 2014).

In their seminal work examining the relationship between teacher effectiveness and student achievement, Sanders and Horn (1998) concluded that “the effectiveness of the teacher is the major determinant of student academic progress” (p.247), and took the position that their results indicated “that any realistic teacher evaluation process should include as a major component a reliable, valid measure of a teacher’s effect on student academic growth” (p.253). Tucker and Stronge (2005) studied four distinct evaluation systems that incorporated data on instructional practices and student learning and concluded that “student assessment data have the potential to transform teacher evaluation from one of opinions about professional ability to one of factual information on what teachers actually accomplish with students” (p.92).
These systems differ dramatically from standard-based systems because they are focused solely on educational outcomes (Papay, 2012), and therefore can be thought to address the concern that has been raised about validity of standards-based systems (Little, Goe & Bell, 2009). There is generally less concern about bias with value-added measures as well due to the fact that test scores are more objective.

Early studies based on the Tennessee Value-Added Assessment System indicated that “homogeneity and heterogeneity of student ability levels within classes are not major concerns in assessing teacher effectiveness” (Sanders & Horn, 1997, p.66). More recent research conducted on value-added measures used in North Carolina by Rothstein (2010) indicated that there is potential for biased results based on assignments of students and student groups to teachers. McCaffrey, Lockwood, Koretz and Hamilton (2004) claimed that if differences in student populations across schools are not accounted for, value-added estimates of teacher impact can be biased.

Other concerns raised about value-added measures are the ability to differentiate the effect of one individual teacher on a student’s test scores when many students interact with multiple educators each day (Tucker & Stronge, 2005). Also, timing of testing has been raised as a concern; testing takes place prior to the end of the school year, so learning that occurs after the test is not accounted for in the score of the current year’s teacher but may be included in the impact rating for the teacher that the student has in the successive year (Papay, 2012). Furthermore, end-of-year assessment results are unlikely to have significant effect on teacher instructional practices (Garrett & Steinberg, 2015; Steinberg & Sartain, 2014). It has also been shown that on some value-added measures, students can receive no more than a perfect score,
thereby limiting the growth capability of the score attributed to a particular teacher (Papay, 2012).

Reliability within a value-added system is difficult to determine primarily because doing so involves comparing a teacher’s performance from one year to his or her performance in another year. It cannot be assumed that a teacher’s performance from one year to the next is consistent, so differences in performance as measured by a value-added system cannot automatically be attributed to a lack of reliability in the system. Papay (2012) claimed that due to concerns about reliability, most supporters of value-added teacher evaluation systems advocate for use of trends in student achievement data over multiple years, rather than a single measure of student achievement over the course of one school year.

Validity is also called into question with value-added measures. It is possible that an assessment used to measure student learning does not actually measure students’ learning, or that a student or groups of students have actually learned, but have not learned what is measured by the test (Little, Goe & Bell, 2009). McCaffrey et al. (2004) pointed out that achievement tests are limited in their ability to measure all aspects of achievement. In addition, 50 percent or fewer of the nation’s teachers teach subjects that are tested by standardized tests, (Steinberg & Sartain, 2015; Toch & Rothman, 2008). Many of the standardized tests measure low-level skills, and therefore might be useful for identification of failing teachers, but would do little to differentiate those who are high performing (Toch & Rothman, 2008).

McCaffrey et al. (2004) pointed out several additional potential concerns about use of value-added measures. They posited that due to student mobility, there is a strong likelihood that data will be missing for individual or groups of students from year to year, creating potential bias in the data, and that changes to test timing, scales for measuring achievement and weight given
to certain topics could change conclusions about teacher impact. Furthermore, it is possible that teachers and administrators can manipulate test results by cheating or teaching to the test (Papay, 2012).

Chetty, Friedman and Rockoff found that value-added measures “accurately predict teachers’ impacts on test scores” (p.59). However, Little, Goe and Bell (2009) pointed out that variation in teacher value-added scores had not been “consistently and strongly linked to what teachers do in their classrooms” (p.5). Proponents of the value-added approach believe that if teachers’ effect on student achievement could be linked to specific teacher characteristics or behaviors, there is potential to use this information in the selection and training of teachers (McCaffrey et al., 2004).

**Impact of federal mandates on teacher evaluation in U.S. and Massachusetts.** The U.S. Race to the Top Fund was developed from the American Recovery and Reinvestment Act (ARRA) that was signed into law by President Obama in February 2009. ARRA goals included investment in education. Through RTTT, 4.35 billion dollars was allocated to competitive grants for states. The funding was earmarked for education innovation and reform. Qualifying reform efforts were identified as increased student achievement, closing achievement gaps, increased high school graduation rates, and improved college and career readiness, (U.S. Department Education, 2009). States applying for RTTT funding were provided with a list of priority areas, and an associated point value that would contribute to an overall total number of points to be used in assessing each state’s application. One RTTT grant selection criteria was “improving principal and teacher effectiveness based on performance” (U.S. Department of Education, 2009, p.3). Points were gained by states that included use of rating categories that differentiated effectiveness by including data on student growth. Specifically, “RTTT required participating
states to have or develop policies that differentiate educator performance by at least three levels and use student learning and growth as a significant factor in educator evaluation” (Massachusetts Department of Elementary and Secondary Education, 2011, p.9).

Massachusetts was one state successful in its bid to win RTTT funds. Massachusetts was awarded 250 million dollars, which was earmarked for four main priorities: great teachers and leaders, curricular and instructional resources, concentrated support for low-performing schools, and college and career readiness. The focus area related to great teachers and leaders was addressed through development of a new educator evaluation system.

Massachusetts began the work of evaluation system development by convening the Massachusetts Task Force on the Evaluation of Teachers and Administrators. This 40-member task force included teachers, building and district administrators, educational consultants, college and university professors, and members of the Massachusetts Department of Elementary and Secondary Education. They worked together for eight months and developed a series of recommendations that were presented to the Massachusetts Board of Elementary and Secondary Education (MBESE) in 2011. On June 28, 2011 Final Regulations for the Evaluation of Educators, 603 CMR 35.00 were adopted by the MBESE. The regulations were designed to promote growth and development of educators (Massachusetts Department of Elementary and Secondary Education, 2015a) and were used to develop the Massachusetts Model System for Educator Evaluation, published in January, 2012. The regulations and Model System relied heavily on the recommendations from the Massachusetts Task Force on Evaluation of Teachers and Administrators.

Implementation of a mandated evaluation system was a significant change for educators in Massachusetts because prior to June 2011, the state required only that an evaluation system
was in place in each school district. Former regulations for evaluation, adopted in 1995, included a set of Principles for Effective Teaching; “guidelines for districts to use in establishing their own systems of evaluation” (Massachusetts Department of Elementary and Secondary Education, 2011, p.10). Each district school committee was charged with specifying the performance standards that were to be used in the district and collectively bargaining the parameters of the evaluation system with the local teacher’s union. In fact, the report developed by the Massachusetts Task Force on the Evaluation of Teachers and Administrators noted universal agreement that “in its present state, educator evaluation in Massachusetts is not achieving its intended aims: providing educators with adequate feedback for improvement and serving as an important accountability tool to ensure effectiveness that supports student learning and growth” (Massachusetts Department of Elementary and Secondary Education, 2011, p.10).

The claim of inadequacy of teacher evaluation in Massachusetts was supported by research conducted in 2002 by MassPartners for Public Schools. That research indicated that the lack of consistent, shared understanding of effective teacher practices due to the variability of teaching standards used in districts, as well as lack of a common understanding of the effective implementation of teaching standards, inadequate evaluator training, and lack of time for evaluators to effectively implement evaluation practices all negatively affected the effectiveness of teacher evaluation in the state (MassPartners, 2002). The Massachusetts Task Force on the Evaluation of Teachers and Administrators report confirmed a high degree of variability with the quality of educator evaluation across the state. Among the concerns cited were the lack of a system for calibration, inadequate evaluator training, and the prevalence of two-scale rating systems that failed to adequately differentiate educator quality (Massachusetts Department of Elementary and Secondary Education, 2011).
The Massachusetts Educator Evaluation System incorporates aspects of standards-based evaluation systems and value-added systems, and merges the summative evaluation function focused on judgment of teacher practice with the formative (supervision) function focused on teacher growth and development. The standards-based aspect of the Massachusetts system uses a four-level rating system to rate teacher performance against a rubric that describes research-based best practices. The value-added aspect rates the affect that a teacher has had on student performance on three levels. Both ratings contribute to the decision about a teacher’s plan for future evaluation cycles, however the dominant deciding factor is the standards-based rating. Figure 1 illustrates the impact that the standards-based rating (Summative Performance Rating) and the value-added rating (Impact on Student Learning) have on the evaluation plan for each educator.

*Figure 1. Factors determining educator plans for evaluation (Massachusetts Department of Elementary and Secondary Education, 2012, p.29)*
It is important to clarify that Massachusetts’ use of student achievement data is not a strictly value-added system. Massachusetts mandates use of MCAS data for teachers who teach MCAS-tested subjects, but allows districts to develop their own measures of student learning called district determined measures (DDMs) as well. Some DDMs are based on student learning gains and others can be based on student achievement. In the case of educators who do not directly teach students, such as school psychologists and school nurses, DDMs can be indirect measures of student growth. In addition, the Massachusetts system requires that each educator have two measures of student growth per year and that the final impact on student learning rating that is used in the evaluation system be based on a two-year trend (Massachusetts Department of Elementary and Secondary Education, 2012).

Massachusetts school systems were provided with implementation options: They could adopt the Model System, adapt the Model System without violating any of the new regulations, or revise their current evaluation system to meet the requirements of the new regulations (Massachusetts Department of Elementary and Secondary Education, 2015a). The state developed a three-phase implementation timeline for each of the major components of the system. Phase I mandates the use of the standards-based rating system (Unsatisfactory, Needs Improvement, Proficient, or Exemplary) for every educator for each of four performance standards, and overall. In addition, Phase I requires each educator to develop at least one professional practice goal and one student learning goal, along with a plan to achieve each goal. Educators must track their progress toward goal attainment throughout their evaluation cycle. At the end of the evaluation cycle, the evaluator assigns a rating for every educator based on their attainment of their professional practice goal and student learning goal. Goal attainment ratings are Did not meet, Some progress, Significant progress, Met, and Exceeded (Massachusetts
Department of Elementary and Secondary Education, 2015b). Phase II uses feedback from students in the evaluation process. Districts, along with their unions, are free to determine the most appropriate method for collecting student feedback data, as well as how the student feedback data will be used in the evaluation system (Massachusetts Department of Elementary and Secondary Education, 2015a). Phase III adds the value-added component, requiring a rating for each educator based on his or her impact on student learning. The impact on student learning rating is determined by two-year trends in student achievement gains on mandated state tests and/or district-determined measures of student learning.

Some districts were given implementation options and others were mandated to adopt within a specific timeline. Schools designated as level four (annual school rating based on MCAS results) and those districts that chose to pilot the new system implemented Phase I in Fall 2011. Districts that received RTTT grant funds and those that chose to be early adopters implemented Phase I in Fall 2012. All other districts were mandated to implement Phase I in Fall 2013. The timelines for implementation of Phases II and III were modified as statewide implementation proceeded, with final decisions resulting in Phase II implementation in the 2014-2015 school year, and Phase III implementation in the 2015-2016 school year. Districts were mandated by law to negotiate the exact terms of the teacher evaluation system that they would adopt with the teacher union in the district. All regulations had to be adhered to, and the final agreed-upon system had to be sent to the DESE for approval (Massachusetts Department of Elementary and Secondary Education, 2015a).

Mandated in the regulations was a five-step evaluation cycle. Figure 2 illustrates the cycle.
As shown in the evaluation cycle, step four is named Formative Assessment/Evaluation. “Formative Assessments are a means for evaluator and educator to check in on the educator’s progress toward goals and performance on the Standards” (Massachusetts Department of Elementary and Secondary Education, 2011, p.21). “A Formative Evaluation is an evaluation at the end of year one for educators on two-year Self-Directed Growth Plans used to arrive at a rating on progress towards attaining the goals set forth in plans, performance on performance standards, or both” (Massachusetts Department of Elementary and Secondary Education, 2015b, p.41.). It is important to note that the above definition provided by the Massachusetts Department of Elementary and Secondary Education implies that formative data is used to make an evaluative judgment about teacher practice. In fact, the form recommended for use in the Formative Evaluation stage requires the evaluator to rate the educator in exactly the same manner as is required in the Summative Evaluation stage. It is easy to imagine that educators implementing this system that requires reporting of evaluative judgments of teacher practice
within a formative process might confuse the purposes of formative feedback (generally accepted to be teacher growth and development), with the purposes of evaluation (accountability).

In the Model System, as a result of their summative evaluation (standards-based) rating and their impact on student learning (value-added) rating, each educator in Massachusetts is placed on a plan. Four plan types are identified in the Model System: Developing, Self-Directed Growth, Directed Growth, or Improvement. Plans vary in length and intensity of support and oversight.

The Developing Educator Plan applies to teachers who have not yet achieved professional teacher status, a designation typically granted once a teacher has successfully completed three years of teaching in a specific school district. Developing Educator Plans can also be used for teachers in a new assignment within a school district. These one-year plans are designed by the educator and the evaluator, and they include more frequent evaluator/teacher interactions and more frequent formative feedback within the evaluation cycle (Massachusetts Department of Elementary and Secondary Education, 2015a).

Self-Directed Growth Plans are designed by the educator and approved by the evaluator, and they are designated for teachers who have achieved professional status and who have earned an overall summative rating of Proficient or Exemplary. If a teacher has also earned a moderate or high impact on student learning rating, the duration of their plan is two years. If the rating for student learning is low, the Self-Directed Growth Plan is a one-year plan (Massachusetts Department of Elementary and Secondary Education, 2015a).

Directed Growth Plans are one-year plans used for teachers who have received overall summative ratings of Needs Improvement. Directed Growth Plans, like Developing Educator Plans typically involve more frequent evaluator/teacher interactions and formative feedback in
an effort to support improvement in teaching practice (Massachusetts Department of Elementary and Secondary Education, 2015a).

Improvement Plans are used with teachers who are rated Unsatisfactory overall. According to the regulations, Improvement Plans can range from 30 days to a maximum of one year. Improvement Plans are developed by the evaluator and are intended to provide the most intensive levels of support and formative feedback to teachers (Massachusetts Department of Elementary and Secondary Education, 2015a).

Massachusetts requires that school districts annually report teacher ratings on each of the standards of performance, as well as overall ratings. In 2018, Massachusetts will begin requiring annual reporting of impact on student learning ratings for each teacher as well (Comstock et al., 2015). For teachers on Developing Educator, Directed Growth and Improvement Plans, the submitted annual ratings are always summative ratings. For teachers on two-year Self-Directed Growth Plans (the overwhelming majority of teachers), the annual overall rating submitted to the state is a summative rating only every other year; the mid-cycle rating is submitted at the end of the first year in the two-year cycle and is called a formative evaluation rating. Although Massachusetts does not make ratings of individual teachers available to the public, it does annually publish ratings aggregated by school and district (Massachusetts Department of Elementary and Secondary Education, 2015a).

Massachusetts’ system for teacher evaluation includes aspects of standards-based evaluation models and value-added models. In addition, it incorporates multiple data sources, and mechanisms for teacher self-reflection, all of which are recommendations from scholars and researchers who advocate for the most effective possible evaluation system. To date, empirical evidence relating to the Massachusetts Educator Evaluation System is limited to a study
commissioned by the Massachusetts Department of Elementary and Secondary Education, and
two doctoral dissertations, reviewed herein.

The Massachusetts Department of Elementary and Secondary Education commissioned a
study that incorporated principal and teacher perceptions of the Massachusetts Educator Evaluation System (Comstock, Harless, Hsieh, & Mitchell, 2015). Data from that study indicated that 48 percent of school staff and 70 percent of principals perceived the evaluation system as primarily focused on educator development. Similarly, 43 percent of staff and 76 percent of principals perceived the evaluator framework as helping improve student achievement. When specific components of the framework were examined, feedback and self-assessment were rated highest in the perception of both teachers and principals as to the impact on classroom practice, with formative assessment/evaluation and summative evaluations ranked much lower. Eighty-three percent of school staff and 67 percent of principals reported that summative evaluation creates anxiety for teachers. Bullis (2014), who aggregated data on the perceptions of principals in Florida and Massachusetts related to implementation of similar new evaluation systems, found that principals perceived the expansion of teacher rating systems in their states as having the intended consequences of “promoting teacher growth, recognizing teacher excellence, promoting remediation, and supporting the dismissal of ineffective teachers” (p.188). Among the unintended consequences were low teacher morale, interference with teacher growth and production of teacher stress (Bullis, 2014). In a study that examined the perceptions of teachers in districts and school that chose to pilot the Massachusetts Model System, Thomson (2014) found that 39 percent of respondents agreed or strongly agreed that the evaluation process caused positive changes in their teaching practice, whereas 43 percent disagreed or strongly disagreed with that statement. Thompson noted wide variation in
responses within his data, and posited that the evaluation system used previously in each district, variability in implementation procedures as well as the fact that the system had only been used for one year, may have affected responses. However, the majority of respondents (60 percent) thought that the use of SMART (Specific and Strategic, Measureable, Action Oriented, Rigorous, Realistic, Results-focused, Timed and Tracked) goals helped focus efforts on student achievement, and likewise 67 percent agreed or strongly agreed that the self-reflection process caused them to consider their teaching practice. There is a great deal more to learn about the implementation and effect of the Massachusetts Educator Evaluation System. This research provides information related to the unintended consequences suggested in the research conducted by Bullis (2014).

The Link of Teacher Effectiveness to Student Achievement

There is strong empirical evidence to support the fact that student achievement is linked to teacher effectiveness. William J. Sanders, June C. Rivers, Sandra Horn and S. Paul Wright are well known in education for their research in this area. In a series of studies examining student achievement gains as measured by the Tennessee Value-Added Assessment System and teacher effectiveness, those researchers concluded that “the effectiveness of the teacher is the major determinant of student academic progress” (Sanders & Horn, 1998, p.247). Specifically, Wright, Sanders and Horn (1997) found that the average achievement of students who had spent three successive years in the classrooms of highly effective teachers was 52 percentile points higher than the average achievement of students who had spent three successive years with low-performing teachers. The findings further indicated that there is a wide variation in effectiveness of teachers, and that effective teachers improved the achievement of all students, regardless of the degree of heterogeneity in the classroom. In addition, the impact of effective and ineffective
teachers was longstanding. Wright, Sanders and Horn (1997) reported that learning gains after a year with an effective teacher were maintained, but the effect of a low-performing teacher persisted when students were subsequently placed with a highly effective teacher. Studies conducted by Mendro (1998) showed similar results:

> across three or four years, students with ineffective teachers can perform at a level fifty percentile points lower than students with effective teachers … the effects of a teacher in the bottom third of effectiveness last through three years of teachers in the top third of effectiveness (p.262).

Subsequent research conducted by Nye, Konstantopoulos and Hedges (2004) suggested that a difference of 0.35 standard deviation in reading achievement gains and of 0.48 standard deviation in mathematics achievement gains could be found between students who had a 25th percentile teacher and a 75th percentile teacher. Similar results (0.33 standard deviation difference in reading gains and 0.46 standard deviation difference in mathematics gains) were found between students who had a 50th percentile teacher and those who had a 90th percentile teacher. Goldhaber (2002) likewise found that variations in teacher quality account for 8.5 percent of variation in student achievement outcomes. These results are similar to Rivkin, Haushek and Kain (2005), who estimated a minimum impact of 7.5 percent variation in student achievement due to teacher quality.

There is little disagreement among educators that the quality of a teacher significantly influences student outcomes. From this, one can draw the conclusion that improvement in student achievement can be accomplished by improving teacher effectiveness (Sanders & Horn, 1997). Researchers, scholars, and practitioners are engaged in ongoing debate as to the most effective strategies to promote teacher growth and development. Some claim that the
accountability aspect of and judgment-based feedback provided by evaluation offers information that allows teachers to understand shortcomings and act to improve upon areas of weakness (Goldsberry, 1997; Sanders & Horn, 1998). Others claim that a supervisory approach that involves collaboration with the principal and encourages teacher growth and development through an inquiry process that involves open discussion of non-judgmental feedback and problem solving to improve teaching practice is more effective in producing gains in teacher effectiveness. A third camp believes that a system that merges the accountability aspects of evaluation with the collaborative approaches of supervision will allow for greatest increases in teacher effectiveness.

**Effectiveness of Teacher Evaluation for Improvement of Teaching**

There is ongoing debate and confusion about the functions of and processes for effective supervision and evaluation, but the common goal of these evaluations is ensuring the highest possible quality education for students. Given that fact, it is necessary to examine the research relating to the effectiveness of supervision and evaluation in practice. Empirical research on teacher evaluation has primarily focused on the effectiveness of evaluation based on validity and reliability of teacher performance data.

In a review of the research on teacher evaluation in 2000, Peterson concluded evaluation practices had no impact on improving teacher practice, and in fact did not even accurately represent classroom practice. In a later review of the same type, Peterson (2004) identified numerous concerns related to use of evaluative judgments of teacher performance. Among the concerns cited were inadequate reliability and validity of judgments due to unreliable numbers of samples, principal bias, and changes in classrooms due to the presence of the evaluator. Peterson (2004) pointed out that that evaluation systems that include use of student achievement data have
been found to be problematic in assessment of teacher quality due to the absence of reliable and valid assessment data for all teachers. Weisburg et al. (2009), after studying 12 school districts in four states with 15,000 teachers and 1,300 administrators concluded that evaluation systems fail to “provide credible and accurate information about individual teachers’ instructional performance” (p.4). Their research found failure to measure and identify variation in teacher quality, which in turn prevented teachers from receiving specific and meaningful feedback. They further identified the absence of agreed-upon standards for teaching practice, lack of evaluator training, infrequent observations of teachers, and cultures where teachers expect high performance ratings and resist suggestions for improvement as contributing factors. Toch and Rothman (2008) in a report exploring the “causes and consequences of the crisis in teacher evaluation” (p.1), cited many of the same concerns as Weisburg et al., calling infrequent, unfocused observations guided by capricious rating systems “drive-bys,” that overwhelmingly result in high teacher ratings.

In contrast, Milanowski, Kimball and White (2004) found a relationship between higher teacher ratings and greater gains in student test scores in Cincinnati, Las Vegas and Los Angeles, which all use rubrics like Danielson’s to rate teacher performance. Robert Pianta found similar results in a 2005 study of another rubric-based evaluation model (Classroom Assessment Scoring System) in Virginia. Rockoff and Speroni (2010) studied the extent to which subjective (rubric-based) and objective (student test scores) evaluation of new teachers predict their future effect on student achievement. Study results indicated that both measures predict future success. This research supports the validity of such ratings, but does not address whether the evaluation system had a positive impact on teacher practice.
Ensuring reliability and validity of evaluative judgments made by evaluators is essential if the judgments are to be used by teachers to improve upon areas of identified weakness. It is important as well to consider empirical evidence related to teacher improvement in response to evaluative feedback. To date, there has been very little research directly addressing this question. Taylor and Tyler (2011) conducted a study with mid-career elementary and middle school teachers in Cincinnati Public Schools and found that teachers were more effective in increasing student achievement during their evaluation year than they had been previously, and that their effectiveness continued to increase in the year after their evaluation. The improvements in the year after evaluation were most significant for those teachers who were weakest prior to evaluation. It is important to note that the system used in Cincinnati typically includes four classroom observations, three by a peer and one by the principal. During each observation (the first of which is announced) teachers are rated on a rubric based on Danielson’s Framework for Teaching. Written feedback in addition to the ratings is provided to teachers after each observation, and at least one meeting is held to discuss results. A final summative score is calculated at the end of the year. Taylor and Tyler concluded that their evidence supports the notion that “subjective evaluation can improve employee performance, even after the evaluation period ends” (p.84). More recently, Steinberg and Sartain (2015) found that schools in Chicago that were randomly assigned to participate in the pilot program for use of Danielson’s Framework for Teaching as an evaluation tool showed significantly greater gains in reading performance, and positive, but not significant gains in math performance compared to schools that did not use the Framework for Teaching. It is important to note that both systems studied include aspects of collaboration that are typically associated with the supervisory function.
In studying the IMPACT teacher evaluation system implemented in the Washington, DC public schools, Dee and Wycoff (2015) found that teacher practice improved, particularly for low-performing teachers who remained in the district under the threat of dismissal based on performance, as well as for high-performing teachers. The IMPACT system, unlike the systems studied by Taylor and Tyler, and Steinberg and Sartain, employs the use of teacher observation and student test performance. IMPACT is also unique in its incorporation of compensation-based rewards linked to teacher performance. These studies, though limited in scope, provide the first promising empirical evidence that systems for evaluation can be used to positively impact teacher quality.

Impact of Collaboration on Teacher Practice and Student Learning

Although there is limited empirical evidence to support the idea that teacher evaluation processes improve teacher practice, and in turn, student achievement, it has been a longstanding belief that effective evaluation systems will accomplish this goal (Darling-Hammond, 2013). Recent studies involving evaluation systems that incorporate a component of supervision have shown that there is a positive effect on teacher performance (Steinberg & Sartain, 2015; Taylor & Tyler, 2011). However, there is a significant body of empirical evidence that supports the link between teacher collaboration and collaborative cultures in schools, and changes in teacher practice as well as improved student learning, leading to overall school improvement (Dufour and Eaker, 1998).

Collaboration with principals has also been shown to impact changes in teacher practice. Principal-teacher relationships that allow for ongoing dialogue about teaching practice and student achievement are integral to teacher professional growth (Drago-Severson, 2007 Danielson, 2002; Glickman, 2002).
As discussed previously, supervision is considered a collaborative process when it involves non-judgmental, inquiry-based discussions related to improved teacher practice or student learning. Research supports the idea that collaboration within supervision does affect classroom instruction. Blasé and Blasé (1998) conducted a study of 800 teachers in the United States seeking to explore the question of what influenced teachers’ classroom instruction. Among the positive influences identified by teachers were principals’ “purposeful, appropriate, nonthreatening suggestions,” as well as principals’ attempts to provide a collaborative environment, and to be non-evaluative and non-judgmental in providing post observation feedback (p.30). Nolan, Hawkes and Francis (1993) reviewed six case studies of clinical supervision that were successful in their impact on changes in teacher thinking found that one factor that contributed to the success of the process was “the development of a collegial relationship in which the teacher feels safe and respected” (p.54).

Trust in Schools


Tschannen-Moran, Hoy, Byrk and Schneider with colleagues have conducted much of the contemporary research on trust in schools. Their conceptualizations of trust are closely related. The influence of the research conducted by Tschannen-Moran, Hoy and their partners
has had a significant influence on what is known about the foundations and results of trusting relationships in schools.

Tschannen-Moran (2004) indicated that trust is only necessary when “parties are dependent upon one another for something that they care about or need” (p.20). According to Tschannen-Moran (2004), “Trust is one’s willingness to be vulnerable to another based on the confidence that the other is benevolent, honest, open, reliable and competent” (p.19). The five factors contributing to trust are defined as follows:

- Benevolence is “confidence that one’s well-being will be protected…and not harmed by the person in whom one has placed one’s trust” (p.21).
- Reliability is “the sense that one can depend on another consistently” (p.33)
- “Competence is the ability to perform a task as expected, according to appropriate standards” (p.34)
- “Honesty concerns a person’s character, integrity, and authenticity” (p.25)
- “Openness means the disclosure of facts, alternatives, judgments, and feelings” (p.28)

Byrk and Schneider (2002) posited that “trust relations can have a profound effect in the processes of school reform” (Byrk & Schneider, 2002, p.33), and extended the work of Tschannen-Moran and Hoy in development of the theory of relational trust in schools. This theory will underpin the proposed research study. The theory of relational trust focuses on three levels of reciprocal social interaction in schools: intrapersonal, interpersonal, and organizational. At the intrapersonal level “trust is rooted in a complex cognitive activity discerning the intentions of others” (Byrk & Schneider, 2002, p.22). At the interpersonal level, the discernment of intentions of others relate directly to role relations and expectations established within the culture and traditions of the unique school community. At the organizational level, “more
effective decision making, enhanced social support for innovation, more efficient social control of adults’ work and an expanded oral authority to ‘go the extra mile’ for the children” (Byrk & Schneider, 2002, p.22), both result from and enhance relational trust.

Byrk and Schneider (2002) theorized that respect, competence, personal regard for others and integrity are the four essential elements that contribute to trusting relationships. “Respect is built when “individuals intently listen to each other and in some fashion take others’ perspectives into account in future action” (Byrk & Schneider, 2002, p.23). The confidence that an individual can effectively perform his or her formal role and achieve desired outcomes is described as the discernment of competence by Byrk and Schneider (2002). Personal regard for others is the perception that the one person cares about another person, a concept closely related to benevolence as described by Tschannen-Moran (2004). The discernment of integrity involves the confidence that one will be consistent in what they say and do, and that a “moral-ethical perspective guides one’s work” (Byrk & Schneider, 2002, p.26). Byrk and Schneider (2002), like Tschannen-Moran (2004) emphasized that trust is necessary in schools due to the reciprocal vulnerabilities inherent in the interdependent relationships that exist among all school roles.

**Trust and school improvement.** “One of the most consistent findings of the research on school improvement is the critical role that collaborative cultures play in the improvement process” (Dufour and Eaker, 1998, p.196). “Collective decision making with broad teacher buy-in, a crucial ingredient for reform, occurs more readily in schools with strong relational trust” (Byrk & Schneider, 2003, p.42). Trust produces positive attitudes, cooperation, risk-taking and better team processes (Dirks & Ferin, 2001), and “helps to focus collective energy on what is important” (Forsyth, Adams & Hoy, 2011, p.157). These factors enable participants to coalesce
around a plan of action” (Byrk and Schneider, 2002, p.33) and as such, school reform efforts are likely to progress at a faster rate in contexts of high trust (Byrk and Schneider, 2002).

Danielson (2015) claimed, “the first, and some would argue the most important, characteristic of a school making progress toward improved student learning is that the leader has established an atmosphere of trust: trust among teachers and between teachers and administrators” (p.19). Research conducted in 2003 by Bryk and Schneider confirmed that “elementary schools with high relational trust were much more likely to demonstrate marked improvements in student learning” (p.43)

School reform efforts necessarily focus on improved student learning. Improvements to student learning are directly linked to improvements in teacher effectiveness. Implementation of new teaching practices often results in a period of decreased performance known as an implementation dip. The implementation dip occurs due to the fact that it takes time for teachers to perfect new instructional strategies. Trust “moderates the sense of uncertainty and vulnerability” (Byrk and Schneider, 2002, p.33), that teachers feel in this situation and increases the likelihood that teachers will discuss the implementation challenges (Tschannen-Moran, 2004) and implement the necessary change even when there is a risk of temporary unskillful performance (Handford & Leithwood, 2013).

**Role of the School Principal**

After an extensive review of research related to leadership and student achievement, Leithwood, Louis, Anderson and Wahlstrom (2004) concluded that “of all the factors that contribute to what students learn at school … leadership is second in strength only to classroom instruction” (p.70). Principals bear the primary responsibility for ensuring the quality of classroom instruction as they work to guarantee that all students achieve at high levels. This
responsibility is a recent and significant shift for principals, many of whom find their role “too complex and highly stressful” (Metlife, 2013, p.3), and claim that among other responsibilities, “evaluating teacher effectiveness and providing guidance and opportunities for teachers to build their competence and skills” is “challenging or very challenging” (Metlife, 2013, p.3).

Despite recent emphasis on the implementation of new of systems for teacher evaluation in the U.S. to improve teaching effectiveness (Danielson & McGreal, 2000; Papay 2012; Peterson, 2000), there exists little evidence to support the direct impact of evaluation on teacher improvement (Peterson, 2004; Tomal et al., 2015). Nevertheless, principals must use a system of evaluation which, in most states includes mandated components (Tomal et al., 2015). In Massachusetts, the mandated Educator Evaluation System includes components that address what is typically referred to in the literature as supervision, and components that address evaluation (Massachusetts Department of Elementary and Secondary Education, 2015b). The purposes for supervision are teacher growth and development (Marzano et al., 2011; Nolan, 1997), and supervisory practices typically resemble practices used by coaches (Sergiovanni & Starratt, 2007; Tomal et al., 2015). Interactions in supervision are non-judgmental (Tomal et al., 2015), and foster inquiry-based discussions related to improvement of teacher performance (Sergiovanni & Starratt, 2007). In contrast, the purpose for evaluation is to measure teacher performance and to make a judgment about the teacher’s proficiency (Nolan & Hoover, 2011; Sergiovanni & Starratt, 2007). The outcomes of the evaluation function are often referred to as “high-stakes” because they are used to make employment decisions (Sergiovanni & Starratt, 2007), and are often submitted to state offices for accountability reasons. The debate about whether evaluation and supervision functions can be successfully merged is a moot point in Massachusetts due to the adoption of regulations that require schools to implement teacher
evaluation systems that include both supervision (formative) and evaluation (summative) functions (Massachusetts Department of Elementary and Secondary Education, 2015b). Because principals must be primarily concerned with student achievement, they must employ the mandated requirements of new evaluation systems in such a way that the systems will be most effective in supporting the improvement of teaching practices (Darling-Hammond, 2013; Marzano et al., 2011; Sergiovanni & Starratt, 2007; Tomal et al., 2015).

The development of trusting relationships between principals and teachers that allow for collaborative discussion of teaching practices have a positive effect on teacher improvement (Tomal et al., 2015), school reform efforts (Byrk & Schneider, 2002) and student achievement outcomes (Byrk & Schneider, 2003). “In hierarchical organizations such as schools, it is the obligation of the individual with the greater power to assume the responsibility of building and sustaining trusting relationships” (Danielson, 2016, p.20). Principals must be ever-cognizant of leadership practices that have positive and negative influences on the development and maintenance of trusting relationships.

Supportive principal leadership behaviors have been linked to increased faculty trust (Gimbel, 2003; Hoy, Tarter & Witkosie, 1992; Tartar, Sabo & Hoy, 1985). The four essential elements of trust as outlined by Byrk and Schneider’s (2002) theory of relational trust create the framework for considering specific principal actions:

**Respect.** Relational trust is grounded in “respectful exchanges…marked by genuinely listening to what each person has to say and by taking these views into account in subsequent actions” (Brewster & Railsback, 2003, p.42). Respect is built when principals are interpersonally and professionally open (Blumberg, Greenfield & Nason, 1978) and involve teachers in decision making, thereby allowing and truly considering the contributions of those
with dissenting viewpoints (Blasé and Blasé, 2001; Blumberg, Greenfield & Nason, 1978).

Within the practice of supervision, principals can build respect by engaging in open and honest discussions that allow teachers to share and discuss their perspectives without fear of judgment.

**Personal regard for others.** “Principals establish both respect and personal regard when they acknowledge the vulnerabilities of others, actively listen to their concerns, and eschew arbitrary actions” (Byrk & Schneider, 2003, p.43). Actions taken to show consideration and sensitivity for the needs and personal wellbeing of others (Blumberg et al., 1978; Sebring & Byrk, 2000), as well as to reduce others’ vulnerability further demonstrate personal regard (Byrk & Schneider, 2002). Maintaining confidentiality, protecting employees’ rights and “refusing to exploit others for personal gain” (Tschannen-Moran, 2004, p .22), are acts of benevolence that build trust.

**Integrity.** Clear and accurate communication, honesty, fairness, consistency, predictability and commitment to follow through on promises are perceived as acts of integrity (Blumberg et al., 1978; Byrk & Schneider, 2002; Costa & Garmston, 2002; Handford & Leithwood, 2013 Tschannen-Moran, 2001), as are modeling of attitudes behaviors and commitments that one expects of others (Dufour and Eaker, 1998, Handford & Leithwood, 2013). “Teachers associate consistency with predictable patterns of action by principals, timely feedback about classrooms and instructional activities, availability of classroom materials and supplies … routines related to discipline, and regular demonstrations of involvement with children and/or their families unrelated to discipline” (p.204).

**Competence.** Overall, the ability of principal to fairly, effectively and efficiently perform his or her professional responsibilities is perceived by teachers as competence (Costa & Garmston, 2002; Dufour & Eaker, 1998). More specifically, Handford and Leithwood (2013)
found that “functional, work-related skills” are most significant for teachers in perceptions of principal competence. Among the skills identified were “being visible, especially in the classroom; and principals’ formal and informal engagement in classroom observations and other classroom based activities such as instructional planning with teachers and providing teachers with feedback about their instruction” (p. 202).

**Trust in Supervision and Evaluation**

Trust can be built and destroyed within the teacher supervision and evaluation role (Nolan & Hoover, 2011; Sergiovanni & Starratt, 2007). Without a preexisting trusting relationship, it is impossible for supervision or evaluation to have the desired impact of improved teacher effectiveness (Nolan & Hoover, 2011).

“The nature of trusting relationship can be altered instantaneously with a simple comment, a betrayal of confidence, or a decision that violates the sense of care on has expected of another” (Tschannen-Moran and Hoy (1998, p. 335). Principals, charged with the primary responsibility for improvement of teacher performance through the establishment and maintenance of trusting relationships with their teachers must always guard against actions that will erode or destroy trust (Darling-Hammond, 2013; Sergiovanni & Starratt, 2007; Tomal et al., 2015).

Those scholars and researchers who believe that the role of supervisor and the role of evaluator are incompatible, as well as those who believe the roles can be performed by the same individual but not at the same time, point to supervision and evaluation processes as having the potential to destroy trusting relationships. “When schools fail to separate supervision from evaluation…one of the resulting problems is that teachers respond to observation… with mistrust and apprehension” (Nolan & Hoover, 2011, p.23). A true supervisory coaching relationship that
supports a teacher learning and growth cannot occur when there is an imbalance of power between the participants (Showers, 1985). The imbalance of power, negates the interdependent nature of the relationship and does not result in the condition of mutual vulnerability, necessary for trust (Byrk & Schneider, 2002, Tschannen-Moran, 2004). In the relationship between evaluator and teacher, “the evaluator has all the power and the teacher has all the vulnerability” (Nolan & Hoover, 2011, p.9). This makes establishing trust highly challenging, if not impossible (Nolan & Hoover, 2001). Tschannen-Moran and Tschannen-Moran (2011) conceptualized an imbalance of power as bureaucratic (evaluative) versus professional (focused on growth) and claimed that “when the balance of power tips too far to the bureaucratic elements … rules replace trust, communications become constrained, people hide problems, management becomes intrusive and cooperation is withheld” (p. 12).

The necessity of developing trust with teachers in order for the practice of supervision to be successful, paired with the responsibility to make judgments about teacher practice required for evaluation, is viewed as a conflict in responsibilities by some scholars (Cogan, 1973; Nolan, 1997; Nolan & Hoover, 2011; Peterson, 2000; Popham, 1988; Showers, 1985). Theorists on both sides of the argument claim that each function (supervision and evaluation) has the potential to improve teacher practice and student outcomes. The Massachusetts Department of Elementary and Secondary Education and the Massachusetts Board of Education have merged both functions in the Massachusetts Educator Evaluation System (Massachusetts Department of Elementary and Secondary Education, 2015b). Principals must use leadership practices in implementation of the Educator Evaluation System that allow for the maximum impact on teacher growth and development both through evaluation and supervision (Darling-Hammond, 2013; Tomal et al., 2015).
Conclusion

This literature review was intended to explore what is known about the impact of teacher evaluation practices on school leadership, teacher improvement and student achievement. An overview of teacher evaluation from the historical perspective outlined the development of a long-standing conflict between supervision and evaluation. Further evidence supporting more contemporary perspectives on this conflict indicates that researchers, scholars and practitioners remain at odds over whether the role of supervisor and evaluator can effectively be carried out by the same person. Current trends in evaluation practices influenced by federal education mandates underpinning elements the Massachusetts Educator Evaluation System were explored. As the system that is the subject of this research study, the policies and procedures related to the Massachusetts Educator Evaluation System have been outlined in detail.

The research pertaining to the positive relationship between teacher quality and student achievement outlined herein is compelling, and reinforces the importance of the role of the school principal in ensuring that teachers are continually developing and performing at high levels. Studies on the effectiveness of teacher evaluation systems, in terms of accuracy of ratings and alignment of ratings with student achievement levels are variable. There has been very limited research conducted that investigates the widely accepted premise that effective evaluation will lead to improved teacher and student performance. Recently, a promising body of research shows that teacher evaluation systems that include a formative component do improve teacher practice. There is however, a substantial body of research suggesting that schools where teachers and administrators have developed trusting relationships that allow for open and honest discussions of teaching and learning challenges and generation of collaborative solutions to such challenges see improvements in teacher practices and student outcomes.
In review of the scholarly literature related to the effect of evaluation and supervision practices, it is found that principals should focus their efforts on building trust and collaboration in their role as supervisors and evaluators. The very limited research related to perceptions of teachers and principals about the Massachusetts Educator Evaluation System raises questions as to whether the System is achieving the goal of improving educator effectiveness, and further whether unintended consequences of the System could be impeding the development of trusting relationships between principals and teachers.

Knowing that effective school leadership has a substantial impact on student achievement in a school, Leithwood, Louis, Anderson and Wahlstrom (2004) suggested that efforts to understand how successful leaders make sense of and respond to external mandates and to use that information to improve the practices of educational leaders will be a highly cost effective way to improve school performance. The Massachusetts Educator Evaluation System, one such external mandate, is named for the evaluation function, yet is designed to accomplish both judgment of teacher performance (evaluation) and, teacher growth and development (supervision).

This proposed study explored how principals perceive the effects of this System on teacher growth and development, their ability and opportunity as instructional leaders to support teacher growth and development and the relationships between principals and teachers.
Chapter III: Research Design

Introduction

Standards-based teacher evaluation systems are designed to meet two goals: measurement and improvement of teacher effectiveness (Danielson, 2010; Hinchey, 2010). Measuring teacher effectiveness in a standards-based evaluation system requires a principal to use multiple sources of evidence to determine an overall (summative) rating of teacher performance. Using a standards-based evaluation system to improve teacher effectiveness requires an evaluator to provide ongoing formative feedback. Teachers and leaders collaboratively discuss and reflect on that feedback and use it to determine recommended changes in teaching practice (Danieleson & McGreal, 2000). Principals must take a judgmental stance when measuring teacher effectiveness, yet to create trusting the relationships that are necessary for true collaboration (which are known to improve teacher practice), principals must also take a non-judgmental stance (Danieleson & McGreal, 2000). Thus, principals who implement a standards-based teacher evaluation system attempt to fulfill conflicting roles, both intended to improve teaching practice, and in-turn, student achievement (Hinchey, 2010; Marzano, 2012).

This study examined principals’ perceptions of the Massachusetts Educator Evaluation System to better understand how the system effects teacher growth and development, the ability and opportunity for principals to support teacher growth and development and the relationships between principals and teachers. To capture these perceptions, the following research questions were used:

1. How do principals perceive the Massachusetts Educator Evaluation System as supporting teacher growth and development, in theory and as they have experienced it?
2. How do principals perceive the effect of the Massachusetts Educator Evaluation System on their ability, capacity and opportunity to effectively support their teachers’ development?

3. How do principals perceive the effect of the Massachusetts Educator Evaluation System on the development and maintenance of trusting relationships between principals and teachers, as they have experienced it?

**Purpose Statement**

The purpose of this qualitative study was to explore principals’ experiences with the Massachusetts Educator Evaluation System, and to understand their perceptions of the effect that the system has on teacher growth and development, the ability of principals as instructional leaders to support teacher growth and development and the relationships between principals and teachers. Implementation of the Massachusetts Educator Evaluation System is generally defined as the actions that principals take to enact the teacher evaluation system that is required by the Massachusetts Department of Elementary and Secondary Education.

**Research Paradigm**

The constructivist paradigm aligns with the goals of the research because it focuses on understanding the “lived experience (Erlebnis) from the point of view of those who live it day to day” (Ponterotto, 2005, p.129). The constructivist researcher’s intent “is to make sense of (or interpret) the meanings others have about the world” (Creswell, 2009, p.8), assuming that “reality is socially constructed, that is, there is no single, observable reality” (Merriam, 2009, p.8). Implementation of a new and different leadership responsibility such as teacher evaluation requires a significant shift in principals’ day-to-day practices. The study sought to interpret the
perceptions of principals as they experience the implementation of the Massachusetts Educator Evaluation System as teacher evaluators.

**Research Design**

A qualitative approach was used in this study. The researcher’s objective was to develop “a complex, holistic picture” (Creswell, 1998, p.15), of Massachusetts principals’ lived experiences in the real settings of their schools (Hatch 2002). Denzin and Lincoln (2005) state that qualitative researchers “study things in their natural settings; researchers attempt to make sense of, or interpret, phenomena in terms of the meanings people bring to them” (p.3). The qualitative approach was consistent with the stated goals of this study: understanding principals’ experiences as they implement the Massachusetts Educator Evaluation Regulations. Qualitative studies allow for an emergent research design (Creswell, 2009). Because the specific experience of implementing the Massachusetts Educator Evaluation Regulations is new for principals, and research related to implementation is limited, it might have been necessary to change the research process, questions and/or data collection strategies as the research progressed. Qualitative studies allow for such emergent research design (Creswell, 2009). Although it is common for qualitative studies to involve multiple data sources (Creswell, 2009), in this case data was collected through individual interviews. This methodology is consistent with essential characteristics of qualitative studies because the researcher is the primary data collection and analysis instrument, and the in-person interactions between researcher and participant take place in a natural setting. Qualitative research is interpretive (Creswell, 2009). As such, researchers recognize that their “interpretations cannot be separated from their own backgrounds, history, contexts and prior understandings” (Creswell, 2009, p.176), yet they must guard against the
influence of biases in data collection and analysis. The measures that the researcher took to monitor researcher bias are discussed in greater detail later in Chapter III.

**Research Tradition**

This qualitative study used individual interviews to ascertain the perspectives of elementary school principals who are the sole teacher evaluators in their schools. Fraenkel and Wallen (2009) state, “qualitative researchers are … concerned with understanding situations and events from the viewpoint of the participants” (p.15). The use of interviews allowed the researcher to elicit an understanding of principals’ perceptions of the effect that the Massachusetts Educator Evaluation System has on the growth and development of their teachers, their capacity and opportunity as educational leaders to support the growth and development of their teachers, and their relationships with teachers in their school buildings. These perspectives are not accessible through observation or artifact analysis (Patton, 2015).

The theoretical framework of relational trust was used to examine more deeply the perceived effect that the Massachusetts Educator Evaluation System has on the development and maintenance of trusting relationships between principals and teachers. Byrk and Schneider (2002) theorize that “respect, competence, personal regard for others and integrity” (p.23) are necessary for trusting relationships, and that school “reform is likely to progress faster in high-trust contexts” (p.33).

**Participants**

The study employed a homogeneous, purposeful sampling strategy. Purposeful sampling is defined as the intentional selection of individuals “to learn or understand the central phenomenon” (Creswell, 2012 p. 206). Homogeneous sampling involves the selection of participants based on “membership in a subgroup that has defining characteristics” (Creswell,
Participants in the study included elementary school principals who are the sole evaluators of teachers in their buildings, and who have completed at least one full evaluation cycle within the Massachusetts Educator Evaluation System. The rationale for this criterion is twofold. First, principals who are the sole administrator in a school building shoulder full responsibility for all aspects of school-level educational leadership. This may not be the case in schools with teams of administrators and/or instructional and curriculum leaders. Teams of administrators can, and often do, divide responsibility for different aspects of school leadership. In schools with leadership teams, the principal may not have the experience of balancing all leadership responsibilities in the same way that a principal who is the lone administrator in his or her building does. Principals of secondary-level schools (middle and high schools) were not chosen as participants because they are rarely the only administrator in their buildings. Secondly, it is important for the study participants to have full knowledge of the central phenomenon of the proposed study. The Massachusetts Educator Evaluation process includes several stages that culminate in a summative evaluation for each teacher. At the summative evaluation stage, a principal must rate the teacher’s performance overall and in each of four specific standards. Principals who have not worked with teachers though the entire evaluation cycle at least once do not have sufficient experience with the phenomenon to draw responses to interview questions. Further, including those with limited experience with the evaluation cycle in the participant pool would reduce the homogenous nature of the sample.

Creswell (2012) writes, “it is typical in qualitative research to study a few individuals” (p. 209). Patton (2015) writes, “in-depth information from a small number of people can be very valuable, especially if the cases are information-rich” (p. 311). For this study a sample size of 11 was used.
Recruitment and Access

Study participants were solicited by email using contact information accessed from the public school administration database maintained by the Massachusetts DESE. The DESE uses this to contact principals for official business, and is therefore constantly updated and accurate. The emails included relevant information such as the background of the researcher, the purpose of the proposed study, the participant criteria, how the results would be used, how the identity of participants would be protected (Creswell, 2012). Interested principals were asked to respond by email or telephone. Upon confirmation that those who responded met the criteria for participation, consent forms were emailed for review by participants. The consent forms complied with Northeastern University Institutional Review Board guidelines and included relevant information such as the purpose of the study, an outline of the data collection procedures, measures that would be taken to protect the anonymity of participants and the potential benefits to participation in the study. The consent form also outlined the voluntary nature of participation in the study and clearly stated the participant’s right to withdraw from the study.

Protection of Human Subjects

The process for research study approval by an Institutional Review Board is designed to “ensure that the research meets ethical guidelines and does not in any way impinge on the rights of the individuals being studied or harm them in any way” (Butin, 2010, p.103). All requirements set forth by the Northeastern University Institutional Review Board for the protection of human subjects were met in this study. In addition, the anonymity of all participants was maintained by using pseudonyms. The researcher chose female and male names that were different from the names of actual participants and used these names in place of the
participant names from the time of transcription through publication of the research study. As a result, there are no known risks to individuals as a result of their participation in this study.

Data Collection

One-to-one, semi-structured interviews were conducted with each study participant. The semi-structured interview strategy was employed because it enabled the researcher to “elicit detailed stories, thoughts and feelings from the participant” (Smith et al., 2009, p.57) and “to enter into the other person’s perspective” (Patton, 2002, p.341), while also allowing “the researcher to respond to the situation at hand, to the emerging worldview of the respondent, and to new ideas on the topic” (Merriam, 2009, p.90).

Each participant was offered the choice of an in-person interview at a location and time of his or her choice, or a telephone interview at a time that was convenient for him or her. Interviews were a maximum of 60 minutes in length. The decision to allow study participants the choice of an in-person interview or a telephone interview was made in order to maximize the number of principals willing to participate in the proposed study. All interviews were digitally recorded.

Prior to beginning the interview, each participant was reminded of the consent form and was asked if he or she has any questions about consent. Each participant who engaged in a telephone interview was sent an unsigned consent form and as such was asked if he or she consented to participate in the study before any interview questions were asked. Participants who chose in-person interviews were likewise reminded of the consent form and were asked to sign it before the interview started.

The interviews were guided by pre-planned sets of questions (see Appendix D and Appendix E). The researcher designed the interview questions after careful consideration of
what the interview might cover, as well as consideration of difficulties that might arise in the interview process. The sequence of questions was carefully developed to ensure that more general questions were asked initially so participants could build on their views later in the interview process as key questions are posed (Merriam, 2009). The pre-planned questions served as a guide for the semi-structured interviews and were written so as to be open-ended to allow for meaningful narrative responses (Butin, 2010) and for follow-up questions or elaboration.

**Data Storage**

Digital recordings of interviews were saved on a digital recording device. Immediately following the conclusion of each interview, the digital file was transferred from the digital recording device to the password-protected hard drive on the researcher’s personal computer. Upon confirmation that the interview file was successfully saved, the file was deleted from the recording device. Interviews were transcribed in full by the transcription service Rev.com. In consideration of protecting the anonymity of the participants, all data files, transcripts and researcher notes were identified only by the assigned pseudonym and are kept in a locked, fire-proof box that is permanently stored at the home of the researcher. The document containing the participants name and assigned pseudonyms was destroyed. Signed consent documents will be stored for three years following the end of the study. These are stored in a locked, fire-proof box that is permanently located at the home of the researcher and, at the end of the three-year period, will be returned to the participant or destroyed upon his or her request in accordance with Northeastern University Institutional Review Board guidelines.
Data Analysis

Coding is a qualitative data analysis method that requires successive cycles of analysis to facilitate the organization of raw data into themes (Hahn, 2008). Two coding cycles were employed in the analysis of the focus group interview data collected from principals.

The researcher employed In Vivo coding as the first-cycle coding strategy. This strategy required the researcher to carefully read the entire transcribed text to identify a particularly significant or salient “word or short phrase from the actual language found in the qualitative data record” (Saladaña, 2009, “In Vivo Coding,” para. 1). The use of In Vivo coding served to honor principals’ voices and “preserve[d] the participants’ meanings of their views and actions” (Charmaz as cited in Saladaña, 2009, “In Vivo Coding,” para. 8). Quotes that seem to represent similar ideas were grouped together to prepare for the second round of coding. As successive interview transcripts were coded, the established organizational structure was modified to better reflect the overall data set.

The second cycle of coding for the purpose of establishing themes employed pattern-coding. The researcher reviewed the grouping of quotes that resulted from first cycle of coding and merged those categories that had commonalities based on the research questions to establish more expansive categories (Saladaña, 2009). Themes related to each research question developed from analysis of the pattern codes developed in second-cycle coding (King & Horrocks, 2010).

Trustworthiness

Maxwell (2005) states, “Validity is a goal rather than a product; it is never something that can be proven or taken for granted” (p. 105). Merriam (2009) further notes that “the validity and reliability of a study depend upon the ethics of the investigator” (p.228), and “it is ultimately up to the individual researcher to proceed in ethically a manner as possible” (p.230).
Respondent validation was employed in this study as one measure to ensure trustworthiness. Once the initial analysis of data was complete, study participants were contacted and asked to review the preliminary interpretation to confirm that the researcher had captured the true meaning of what they had said. Maxwell (2009) suggests that this strategy safeguards against researcher bias and misinterpretation.

This study specifically excluded from participation any principal employed by the Quabbin Regional School District (the school district that employs the researcher), thereby eliminating any bias due to “power differentials” (Mears, 2009, p.100) or “the researcher’s relationship to participants” (Merriam, 2009, p.230).

Minimizing researcher bias is essential to ensuring trustworthiness. Two approaches to minimizing bias were employed in this study. The researcher described her experiences with the phenomenon to identify for herself views that she needed to set aside prior to engaging in the analysis of interview data. This process is called bracketing, and it allowed the researcher to focus solely on the experience of the study participants (Creswell, 2007). In addition to bracketing, it is advisable to inform the reader of possible researcher bias due to past experiences that may have affected the design of the study or data analysis process. In this study, the direct involvement of the researcher with teacher evaluation processes in various schools throughout her career as mentioned previously, has influenced her perception of teacher evaluation in general. Furthermore, the researcher has ongoing direct involvement with the specific phenomenon of this study. As the assistant superintendent in her school district, she is primarily responsible for the implementation of Massachusetts Educator Evaluation System. The researcher was the lead facilitator of the implementation planning team and the co-facilitator of the joint labor-management negotiating committee that negotiated the terms of the evaluation
system. One of her primary responsibilities as assistant superintendent is oversight of the teacher evaluation process. This involves partial management of the online system for teacher evaluation, reminding teachers and principals of specific upcoming deadlines in the teacher evaluation process, and ensuring that all evaluators are fully trained. In addition, she schedules and participates in calibration observation sessions and low-performing teacher coaching sessions. During calibration observation sessions, groups of evaluators from across the district observe teachers together and compare their observation findings to ensure that there is consistency in interpretation of teaching practice between and among evaluators. Bi-monthly low-performing teacher coaching sessions involve collaborative discussions that focus primarily on how to most effectively support teacher improvement for teachers on Directed Growth Plans or Improvement Plans. These sessions also frequently include discussion of challenging situations that evaluators currently face. Furthermore, she is an evaluator in the district. She uses the Massachusetts Educator Evaluation System to evaluate three instructional coaches and two academic coordinators. Finally, her professional evaluation is governed by these same regulations. Due to these first-hand experiences, the researcher has developed opinions about the Massachusetts Educator Evaluation System Regulations. She took care to guard against allowing her personal experiences and viewpoints to influence the data collection and analysis process.

Fully transcribed interview data are considered rich data (Maxwell, 2005); such data contribute to the transferability of research findings (Creswell, 1998). In this study, all participant interviews were fully transcribed and used in the data analysis process to ensure that interpretations of results did not exclude from consideration any data that might have been missed by using a note-taking data collection strategy. Further, the rich descriptions will allow
readers to assess shared characteristics that may allow transferability of findings (Creswell, 1998).

The data analysis process employed in this study involved identification of discrepant data to assess whether this data was cause to modify the conclusions. “Identifying and analyzing discrepant data and negative cases is a key part of the logic of validity testing in qualitative research” (Maxwell, 2005, p.112). Merriam (2009) states, that “the question then is … whether the results are consistent with the data collected” (p. 221).

**Limitations**

It is important to recognize in qualitative research, particularly when the data are generated from participant interviews, that “access to experience is always dependent on what participants tell us about that experience, and that the researcher then needs to interpret that account from the participant in order to understand their experience” (Smith et al., 2009, p.3). With this in mind, several limitations to this study have been identified:

1. This study is specific to Massachusetts. Although there are other states that have recently implemented standard-based teacher evaluation systems, the results of this study are not generalizable to states other than Massachusetts; the regulations and implementation policies and procedures in other states may differ from those in Massachusetts.

2. This study involves only elementary school principals who work alone to evaluate teachers in their schools. It cannot be assumed that the results are generalizable to schools where someone other than the principal shares or takes full responsibility for teacher evaluation.
3. The research questions required study participants to rely on memory for recollection of events and perceptions. Memories can be inaccurate, and therefore the perceptions reported by participants may be inaccurate depictions of reality. However, the nature of qualitative research is to seek to understand the meaning that participants make of a phenomenon (Merriam, 2009) and although this is identified as a limitation, it does not invalidate the findings (Maxwell, 2005).

4. “In qualitative research a single case or small, nonrandom, purposeful sample is selected precisely because the researcher wishes to understand the particular in depth, not to find out what is generally true of many” (Merriam, 2009, p.224). The small sample size and voluntary nature of the participant selection process does not allow one to assume that results are representative of the perceptions of all Massachusetts public school principals.
Chapter IV: Research Findings

The purpose of this study was to explore principals’ perceptions of the Massachusetts Educator Evaluation System enacted in 2011 and the degree to which they felt it effected teacher growth and development and their ability to support their teachers’ development and growth. Given the purpose of the study, the following research questions were developed:

1. How do principals perceive the Massachusetts Educator Evaluation System as supporting teacher growth and development, in theory and as they have experienced it?

2. How do principals perceive the effect of the Massachusetts Educator Evaluation System on their ability, capacity and opportunity to effectively support their teachers’ development?

3. How do principals perceive the effect of the Massachusetts Educator Evaluation System on the development and maintenance of trusting relationships between principals and teachers as they have experienced it?

This chapter begins with a summary of the study participants and data collected, and is followed by the presentation and discussion of themes that emerged in the analysis of data in relationship to each research question.

Summary of Study Participants and Data Collected

The participants in this study were elementary level public school principals in Massachusetts. One criterion for participation was experience using the Massachusetts Educator Evaluation System for at least two years as the sole evaluator of teachers in the school building. Eleven principals, each from a different school district, were interviewed. Six participants chose
to be interviewed over the telephone and five chose in-person interviews. Names, gender, number of years as a principal, number of teachers evaluating are included in Table 1.

Table 1

*Name, Gender, Number of Years Evaluating Teachers, Number of Years in Present Position, Number of Years as a Principal, Number of Years Using the Massachusetts Educator Evaluation System, Number of Teachers Evaluating*

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>Number of Years Evaluating Teachers</th>
<th>Number of Years in Present Position</th>
<th>Number of Years as a Principal</th>
<th>Number of Years Using the Massachusetts Educator Evaluation System</th>
<th>Total Number of Teachers Evaluating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbara</td>
<td>Female</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Christine</td>
<td>Female</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Mitchell</td>
<td>Male</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>Darlene</td>
<td>Female</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>Carl</td>
<td>Male</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>25-30</td>
</tr>
<tr>
<td>Jocelyn</td>
<td>Female</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Amy</td>
<td>Female</td>
<td>16</td>
<td>11</td>
<td>16</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Janet</td>
<td>Female</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>June</td>
<td>Female</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>37</td>
</tr>
<tr>
<td>Carol</td>
<td>Female</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Leah</td>
<td>Female</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>32</td>
</tr>
</tbody>
</table>

Data were collected through semi-structured interviews. Participants were given the choice of in-person interviews or telephone interviews. The semi-structured interview protocol (Appendices D and E) was designed to offer participants the opportunity to respond to open-ended questions that would allow the researcher to develop “thick-rich descriptions” of participants’ responses (Creswell, 2009). Following each interview, the audio recording was
transcribed and the researcher began the analysis of the transcript. All transcripts were reviewed and themes were developed through an iterative review and coding of each.

**Research Question 1: How do principals perceive the Massachusetts Educator Evaluation System as supporting teacher growth and development, in theory and as they have experienced it?**

The first research question focuses on two different foci in relationship to the participants’ perspectives. The first is how principals perceived the Massachusetts Educator Evaluation System as expected to support teacher growth and development, in theory; the second being how principals perceived the Massachusetts Educator Evaluation System as actually supporting teacher growth and development as they have experienced it.

The initial theme, identified in Table 2, speaks to how principals perceived the intent of the Massachusetts Educator Evaluation System as supporting teacher growth and development in theory. The subsequent themes, presented in Table 3, speak directly to these same principals’ perceptions of how the Massachusetts Educator Evaluation System actually supported teacher growth and development as the principals experienced it.

Table 2

*Themes emerging from the question, How do principals perceive the Massachusetts Educator Evaluation System as supporting teacher growth and development in theory?*

<table>
<thead>
<tr>
<th>Principal perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principals perceive that the Massachusetts Educator Evaluation System is designed to ensure that effective teachers are in classrooms and that they are continually improving.</td>
</tr>
</tbody>
</table>

Principals who participated in this study had all engaged in professional development as the
Massachusetts Educator Evaluation System was implemented. In addition, some principals reported that they were members of the team that made the decision to adopt the Model System, adapt the Model System or to develop a different system aligned to the regulations for their districts. These experiences likely contributed to the fact that all principals felt that the System was designed to ensure that effective teachers were working with students in Massachusetts public schools. Carol captured this view in her statement, “I think one of the big goals is to make sure that highly trained and qualified teachers are in front of kids every day.”

Many principals spoke of growth mindset and helping teachers grow throughout their interviews for this study, indicating that they believed the System was intended to foster improvements to teaching practices. Christine characterized this position when she said, “I believe [the System was developed] to grow teachers with that continuous model of improvement and reflection . . . when you read the research and you look at how it was developed and methods of it, it was intended to be a growing system.”

Table 3

Themes emerging from the question, How do principals perceive the Massachusetts Educator Evaluation System as supporting teacher growth and development as they have experienced it?

<table>
<thead>
<tr>
<th>The performance ratings that are required in the Massachusetts Educator Evaluation System cause many teachers to view the teacher Evaluation System as a grading system, rather than an opportunity to get feedback to help educators grow.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principals perceive the rubric used in the Massachusetts Educator Evaluation System provides clarity of expectations which supports teacher growth.</td>
</tr>
<tr>
<td>Principals believe the use of specific evidence to support a judgment helps teachers to more clearly understand their performance levels.</td>
</tr>
<tr>
<td>The setting and tracking of progress toward goals is perceived by principals to support teacher growth and development.</td>
</tr>
</tbody>
</table>
The performance ratings that are required in the Massachusetts Education Evaluation System cause many teachers to view the teacher Evaluation System as a grading system, rather than an opportunity to get feedback to help educators grow. When school districts in Massachusetts began to implement the Massachusetts Educator Evaluation regulations, they were provided with the following three implementation options: adopt the Model System that had been developed by the DESE, adapt the Model System, or develop their own system that was aligned to the new regulations (Massachusetts Department of Elementary and Secondary Education, Final Regulations on Evaluation of Educators, 2011). All principals participating in this study reported that their school district had adopted the Model System. The Model System requires that teachers are rated annually overall, and on each of the four performance standards: Curriculum, Planning and Assessment, Instruction, Teaching All Students, Family and Community Engagement, and Professional Culture. There are four levels of teacher performance ratings: Exemplary, Proficient, Needs Improvement and Unsatisfactory.

Principals in this study expressed concerns about the effect of these ratings on their teachers’ growth and development. The ratings, principals feel, create the sense that performance is static and de-incentivizes teachers to change. Principals believe that all teachers should continually improve, no matter what their current performance level, and they try to provide all teachers with feedback that will help them take steps to get better. They are concerned however that a rating of Proficient or Exemplary is perceived by teachers as a “passing” grade and that if they pass, they do not need to improve. Three principals felt as though their teachers viewed their evaluation as a grading system. One principal said that her teachers “really believe it as Exemplary is an A, Proficient is a B, Needs Improvement is a C,
and Unsatisfactory is a D.” Christine, who also referenced her teacher’s perception of ratings as report card grades said,

I have a group of teachers that are taking a course right now, and they get mad if a professor marks them off for something. It’s legitimate. It’s feedback, because, again, we are so locked into being perfect … as soon as somebody says “You’re not perfect” then they get offended. It’s like they can’t take that as just good feedback.

Two other principals shared that their teachers simply wanted to “pass,” and that they were happy when they “get an A.” In reference to the ratings, one principal said, “I believe that there’s the thought, they get Proficient, they’re all set.”

Jocelyn described that her teachers expressed a sense of relief when they felt that she had completed a sufficient number of observations of them. She shared that her teachers’ typical reaction was:

“Oh, are you going to write that up?” They want to know if that’s their observation.

“Are they done?” There’s a lot of those feelings around, “Oh. Now I’m done, right?

That was two. That was two observations, so I’m done, right?”

She questioned if her teachers found the System to be helpful, and if they valued it as a professional growth tool based on these reactions.

Darlene discussed her many efforts to help her teachers understand that evaluation was about improvement only to find out that some of her teachers had actually developed lessons that they could immediately begin to teach if she were to walk in to their classroom to evaluate them. She expressed disappointment and surprise at the realization that her teachers did not embrace the observation as an opportunity to get feedback that would help them get better, rather that they
were fearful that they would not receive a good rating. She said, “That hurt my feelings because I was like, ‘Really? That’s where we are right now?’”

**Principals perceive the rubric used in the Massachusetts Educator Evaluation System provides clarity of expectations which supports teacher growth.** When principals rate teacher performance using the Massachusetts Educator Evaluation System, they use detailed rubrics to guide their decisions. The rubrics are broken down into sections based on the four performance standards, Curriculum, Planning and Assessment, Teaching All Students, Family and Community Engagement, and Professional Culture. Under each of the four performance standards, the rubrics are further delineated into subsections called elements and indicators (see Appendix F). For each subsection, the rubrics describe teacher practices along a performance rating continuum of Unsatisfactory, Needs Improvement, Proficient and Exemplary.

Previous experiences that the majority of principals in the study had with teacher evaluation, either as a principal who was evaluating or as a teacher who was being evaluated, did not include use of a rubric. The specific descriptions of teacher practice outlined in the rubric that is used in the Massachusetts Educator Evaluation System are helpful to teachers because they define exactly what good teaching practices look like and very clearly outline exactly what a teacher must do to achieve a particular performance level. Principals feel that if teachers use the language of the rubric to guide changes in their practices, it will help them improve in areas where they are currently performing below the Proficient or Exemplary levels. One principal stated, “I think it’s made it very clear to the teachers what my expectations are,” while another principal said, “You know what you have to do to get the high score. Just do it,” as she compared the use of rubrics with children in the classroom to use of rubrics with teachers in the Evaluation System.
The rubric sets high expectations for teacher performance, and principals felt that rigorous nature of the rubric descriptors helped teachers understand what excellent teaching practice looks like and provided them with particular practices that they could use as they strive to improve. A principal said “I think the rubrics are really rigorous, and I think it has challenged our thinking about practice, and it has pushed our practice along.” Another principal referenced the Exemplary rating as one to which her teachers aspire, while a third principal reinforced the idea that the rubric, “keep[s] people thinking about the standards, understand[ing] their performance standards, and know[ing] what it looks like to be Exemplary…. [a teacher] can really aspire to [a high performance level] and know how to get to it.”

The rubrics also help teachers and principals come to a common understanding of the difference between performance levels. Many principals spoke about teachers who initially disagreed with a rating of Proficient that had been assigned to them, because they viewed their performance level as Exemplary, but who eventually agreed with the rating after reviewing the rubric descriptors of Exemplary practice. Amy shared how she uses the rubric in when a teacher questions her rating of them.

Having a rubric helps people, absolutely. We also had to flesh out … the whole notion between Proficient and Exemplary and what makes a difference between the two, because I’ll rate somebody one way, yet a teacher might rate themselves a different way, so I pull out the rubric and I say, “If you feel you’re this, I want you to show me the evidence that supports that.” That helps. I think having a rubric helps, because then there’s no more, “You feel this way, but my old principal felt that way,” or, “I feel this way, and I’m not seeing why you feel that way.” It’s really black and white.
Principals believe the use of specific evidence to support a judgment helps teachers to more clearly understand their own performance levels and this understanding prompts teachers to make changes to their practice when necessary. The use of specific evidence to support judgment about teacher practice is required in the Massachusetts Educator Evaluation System (Massachusetts Department of Elementary and Secondary Education, 2015b). Within the Evaluation System, evidence is gathered by principals and by teachers. There are many ways that principals gather evidence about teaching practices. Among the most common methods are observing in classrooms, meeting with teachers to discuss teaching practices and to review artifacts such as assessments and student work, and observing teachers in meetings with colleagues and parents. Teachers also gather evidence about their teaching practices and their progress in achievement of their goals at least twice in each evaluation cycle. The evidence collected by teachers is submitted to the principal and the principal considers it when making his or her final judgment about a teacher’s performance.

Principals participating in this study felt that the use of specific evidence gathered either by the principal or the teachers themselves helped them develop a realistic view of their own performance levels. Once they had the understanding of their performance levels and where they were exhibiting deficits, teachers were able to work toward improvement in the areas of need.

Principals stated that when they had to assign a rating that was less than Proficient, teachers were often upset. In many cases the initial reaction of the teacher was to “blame” the principal or to disbelieve the principal’s judgment of his or her practice. In these situations, when the principal provided the teachers with exactly what he or she had observed in the classroom, the teacher was less likely to dispute the judgment of the principal, and more likely to
accept the feedback that the principal provided to help him or her improve. The following scenario shared by one principal is illustrative of this point:

[When a teacher comes to me after I have observed them to dispute my judgment] … it gives me opportunities to present my evidence and say, “Actually, what about this [that I saw] and is this still an area of need?” I think when you do it that way, and you’re fair with it, I think they can see, “Yeah, I get it. It doesn’t feel like someone got me or dinged me on something but really, she showed me that I had this example, this example and this example, three times when she was in my classroom.” I think that provides an opportunity to see, “Okay, that’s something I can work on. I can grow.”

When teachers are not aware of problems in their own classrooms, it is impossible for them to address the problems. Evidence collected during classroom observations can make a teacher aware of things that he or she has not previously noticed or change his or her perception of something that has happened. Jocelyn described an interaction that she had with a teacher after an observation, wherein the evidence that Jocelyn provided enlightened a teacher as to an inaccurate perception that she had about what had been going on in her classroom.

When I went in to watch her small group instruction, one of the things I noticed during her small group instruction was that her focus on one student was really good, but the rest of the group was off task. I talked about that. I said, “Okay. What were the other students doing while you were working with [the one student]?” She said, “Oh. I think they were reading.” I said, “Let me tell you what I saw that they were doing. They were all off task.” I said. We just worked there. “What are some ideas that maybe would work to keep them busy?” . . . She tried [the ideas], and [they] worked.
Using evidence in this manner allows a teacher to come to a new or different understanding of an area for which improvement is needed. Often, awareness of a problem is all that is necessary for the teacher to work toward improvement.

During the Educator Evaluation Cycle, there are opportunities for each teacher to provide evidence to their evaluator that will be considered as the final annual judgment about his or her performance is made. Principals have found that having teachers engaged in the process of gathering and evaluating evidence of their practice on the performance standards has helped teachers identify areas where they may need to work to improve. For example, when a teacher is trying to find evidence that they have communicated well with parents and they are not able to do so, it causes the teacher to become aware of an area of deficit and can prompt them to repair the deficit without a great deal of intervention on the part of the principal.

**The setting of and tracking of progress toward goals is perceived by principals to support teacher growth and development.** At the beginning of every Evaluation Cycle, each teacher is required to develop at least two SMART (Specific and Strategic, Measureable, Action Oriented, Rigorous, Realistic, Results-focused, Timed and Tracked) goals (Massachusetts Department of Elementary and Secondary Education, 2015b). At least one SMART goal must focus on improvement of “learning, growth and achievement of the students under the educator’s responsibility” (Massachusetts Department of Elementary and Secondary Education, 2015b, p.24), and the other SMART goal must focus on the improvement of the teacher’s professional practice, and must be linked to at least one of the Evaluation System’s four performance standards (Massachusetts Department of Elementary and Secondary Education, 2015b). Throughout the evaluation cycle, each teacher must track his or her progress toward goal
achievement using the measures that have been identified in their goal statement. The principal must approve the SMART goals for each teacher that he or she evaluates.

The participants in this study viewed the goal development and tracking process as supportive of teacher growth. Amy shared that she liked the fact that there was a provision for evaluator approval of goals because she felt that it allowed her to work with each teacher to ensure that the goals were focused in areas where she felt the teacher needed to improve. The process of setting goals and monitoring progress on goals caused teachers to take responsibility for improving their practice, and for reflecting on their progress. June felt that the goals not only helped focus the work of her teachers during the school year, but helped at the end of the year as well because they caused teachers to reflect. As she said,

You have your professional goal and your student learning goal, and I think that’s really important to have as you walk in as you start the new year, but then at the end of the year, you reflect on those two goals. You think back to the year. What was successful? What were my challenges? What were my struggles? Then, what am I going to do next year that’s going to be different? If every teacher just wrote that, I think we’d have a lot more pathways to conversations about teaching and learning.

When professional practice goals and student learning goals were connected with school and district goals, this alignment provided teachers with clear direction to areas of focus for improvement, and created a sense that educators were working together toward common goals. At the same time, because the goals can differ by teacher, there is a degree of autonomy possible for teachers as they develop their specific goals. These two factors help teachers buy into the use of goals within the System and foster opportunities for teachers to collaborate to support
achievement of their goals. Christine, when discussing the value that she sees in the use of goals to support teacher growth said,

[Growth is] all I’ve seen, seriously. Just giving them the freedom to pick their goals, obviously within the parameters of what needs to be done, but they know what needs to be done, and having them track their progress. They’re coming to me and saying, “Look, this is what’s going on. This is what I want to do next.” That’s so powerful. They’re eagerly embracing it … and they’re taking risks…. They’re holding themselves accountable. It has developed a system where, yeah, you have to work on that, but don’t work on something totally outside of what you need to do anyway. Pick something that you’re doing. I’ve seen some really great opportunities for growth.

**Research Question 2: How do principals perceive the effect of the Massachusetts Educator Evaluation System on their ability, capacity and opportunity to effectively support their teachers’ development?**

The second research question focuses on principals’ perceptions about the effect of the Evaluation System on their ability, capacity and opportunity to effectively support teacher development. The themes developed through an iterative coding and analysis of the transcripts in relationship to research Question 2 are provided in Table 4. A discussion of the themes follows the table.
Table 4

Themes emerging from the question: How do principals perceive the effect of the Massachusetts Educator Evaluation System on their ability, capacity and opportunity to effectively support their teachers’ development?

The mandate of frequent, unannounced classroom observations allows principals to develop a clear understanding of a teacher’s practice, and contributes to principals’ ability to provide high-quality and relevant feedback.

Principals find that discussing teaching practice with their teachers is a high-leverage strategy to improve teaching practice.

Principals perceive that written feedback has minimal impact on teacher improvement.

Principals are frustrated that they do not have time to adequately implement all aspects of the Massachusetts Educator Evaluation System because they feel it is detrimental to their ability to effectively support teacher growth.

Principals’ concerns about the detrimental effects of Needs Improvement and Unsatisfactory performance ratings make it difficult for them to assign such ratings.

The mandate of frequent unannounced classroom observations allows principals to develop a clear understanding of a teacher’s practice and contributes to principals’ ability to provide high-quality and relevant feedback. The accountability for the frequent observations that is built into the System is perceived by principals to support the growth and development of their teachers. Principals reported that frequent observations deepened their understanding of the practice of an individual teacher. As Amy described,

I think with the teacher eval system and the walkthroughs that were embedded that you get a better picture of what’s going on in the classroom, not just sitting there once or twice a year. Sometimes I do a walk-through and I don’t write anything up, but I have those notes in my mind, and I connect it to the next time I go in. I’m always in the
classroom. I think anytime you walk in the classroom, you’re evaluating, even if it’s not an official time.

Many principals compared the unannounced observations that are mandated in the current System with the announced observations that were typical prior to the implementation of the Massachusetts Educator Evaluation System, referring to those used previously as a “dog-and-pony show.” In many districts, evaluation involved four very specific steps prior to the 2011 Massachusetts Educator Evaluation System. First, a teacher and a principal would meet to discuss the lesson that the principal was going to observe in the near future. In many cases the teacher would share his or her lesson plan with the principal in this meeting, and together they would discuss each component of the upcoming lesson. Then, on the day of the observation, the principal would arrive in the teacher’s room at the designated time and write scripted notes of everything that he or she saw during the lesson. The principal would remain in the teacher’s classroom for the entire lesson if possible. Later that day or during the next school day, the principal and teacher would meet to discuss the lesson. During this meeting the principal would typically ask questions of the teacher to clarify anything that he or she found confusing, and often the teacher was given an opportunity to share how he or she felt the lesson had gone. Finally, within a few days, the principal would summarize his or her observation of the lesson and provide some commendations and/or recommendations to the teacher in a written document that would be sent to the teacher.

Because these observations were scheduled well in advance and usually on a date and time chosen by the teacher, principals felt that it gave them nothing more than a sense of what a teacher could do on his or her best day. It did not provide a principal with information about the
day-to-day classroom practices of their teachers, thus, the characterization of it as a “dog-and-pony show.” One principal described the contrast:

I think that everybody realizes that the dog-and-pony show was not a true reflection of what actually goes on in the classroom. [The Massachusetts Educator Evaluation System] gives a flexibility of administrators to evaluate on what happens every single day as many times as we want.

The understanding of teacher practice that comes from observations that are more authentic because they are unplanned, and more frequent because they do not have to span the entire class period helps principals understand what routinely takes place in a particular classroom and what the everyday practices of a teacher are. This knowledge is helpful to principals because it helps them make better decisions as to the most effective feedback to provide a teacher to improve his or her practice. For example, if on a particular day, a teacher did not refer to his or her learning objective while the principal was in the classroom, the principal might decide to provide the teacher with feedback telling him that he should regularly communicate the learning objective to the students, and might make some suggestions about how the teacher could remember to do so every day. If on the other hand, if the principal had been in the teacher’s classroom multiple times and there had been no mention of the learning objective, the principal might learn that the teacher did not know the value of developing and sharing learning objectives with his students, or that he did not understand how to write a learning objective. In this case, the feedback provided to the teacher would focus on how to write a learning objective and why it is an important practice to improve student learning. In the first case, the feedback would not have helped the teacher correct the problem, but because the principal had a better understanding of
the problem in the second case, the feedback provided was much more likely to result in improvements to his practice.

**Principals are frustrated that they do not have time to adequately implement all aspects of the Massachusetts Educator Evaluation System because they feel it is detrimental to their ability to effectively support teacher growth.** Principals all spoke about the limited time that was available to them because of the many responsibilities that they have as building leaders. They emphasized that implementation of the Massachusetts Educator Evaluation System taxed their time, and they felt frustrated that they did not have enough time to complete each aspect of the evaluation process as effectively as possible.

Primary among their concerns was the time necessary to write high-quality, evidence-based formative, summative and observation reports for teachers. One principal was annoyed that her that it took her “an hour to write up” a report after a 15-minute observation of a teacher, and another principal reported each of the 24 summative or formative reports that she needed to do at the end of the year took her at least 40 minutes, and that she felt that the quality of her reports suffered due to lack of time to complete them well. She said, “It’s daunting, and it takes a lot of energy and discipline to get them done, and that’s really what you’re trying to do is simply get them done.”

Although many principals noted that they did not believe that the written feedback was as important to helping teachers grow as the verbal feedback that they could have during a conversation with a teacher, they did want to provide high-quality written feedback, and many felt that their ability to do so was compromised by a lack of time. One principal talked about the time necessary for her to ensure that the feedback she was giving was an accurate depiction of what took place during a classroom observation:
Yeah, time is crazy. It’s hard because you’re evaluating someone’s practice. It’s difficult. You don’t want to do it wrong. I don’t want to say something that isn’t fair, or say something that I may or may not have seen, or they may or may not have done.

She also shared the care that she tries to take in writing something that the teacher will find helpful: “I don’t want the write-ups to be meaningless. They’re hard to write,” and wondered if other principals found the writing as time consuming. She said, “I don’t know, maybe every principal takes hours to do them [as I do] … crafting what you want to say, because, what do you want them to hear? You have to be careful with how you say it. It’s a lot.”

Many principals reported that they were not able to devote adequate time to effectively work with all of their teachers within the Massachusetts Educator Evaluation System. Because of lack of time, principals find themselves making choices about which teachers will receive their time and attention. One principal said, “I find myself picking and choosing who I want to dedicate time to…. Say I have five teachers who really needed support and guidance, I just don’t have the time or manpower to do it.” All principals who spoke of this reported opting to spend most of their available time focused on the lower-performing teachers, thereby leaving Proficient and Exemplary teachers with less feedback to improve their practice. Although these principals reported focusing on lower-performing teachers, some noted that they were concerned about the amount of time and energy that would have to be devoted to support a teacher on a Directed Growth or Improvement Plan and questioned whether a principal would be willing to take that on given all of their other responsibilities. These two ideas taken together suggest that teachers who perform below Proficient in some areas, but not poorly enough to warrant a Directed Growth or Improvement Plan might receive the bulk of the support time that principals have available, and those with more significant needs may not be
put on plans that require more support and oversight due to the other demands on principal time. This further reinforces the concern identified by principals that uneven support is being provided to teachers, with very low and very high performing teachers having the lowest levels of time and attention from principals. Those who perform in the low-to-moderate range seemingly receive the most support.

**Principals find that discussing teaching practice with their teachers is a high-leverage strategy to improve teaching practice.** Principals are encouraged that the Massachusetts Educator Evaluation System requires them to be in classrooms frequently, and that it requires them to collect specific evidence, because they feel that the discussions they have with their teachers as a result are high quality and make a difference in teacher performance.

Principals were careful to acknowledge that the short observations typical in the Massachusetts Educator Evaluation System did not always provide them with an adequate understanding of what was happening in the classroom, and they used post-observation conferences to gain additional pertinent information from their teachers. Carol described it this way:

If there is something that I’m making a recommendation for, I always ask the question.

We can never make assumptions. When you are in the classroom and you’re only there for 15 or 20 minutes, you can’t assume to know why someone did something.

After having a conversation to follow-up shorter classroom visits, principals felt more confident that the feedback that they provided to the teacher was more accurate than it would have been if the conversation had not occurred. Principals felt the accuracy of the feedback made it more helpful.
Principals discussed the need to help teachers to continually improve, not just when a principal provided them with feedback. They felt that helping teachers be more reflective through the conversations they had with teachers was an effective way to achieve this goal. Principals described purposefully withholding judgmental statements when discussing evidence collected during an observation of a teacher, opting to engage the teacher in thinking about the lesson as one strategy they employed to foster reflection. One reported asking questions such as,

Well, what made you do it that way? When you think about how the students responded, what do you think about that? Did you get the result that you wanted? Did you think the lesson met your objective at the outset? Were the kids able to do what you thought they should be able to do?

These strategies allow teachers to purposefully think about their practice, and to make decisions about what steps they might take to improve and makes them less likely to rely only on principal feedback to drive improvement.

Carl described that his strategies for helping them become more reflective sometimes made his teachers frustrated with him. He told of a conversation that he had when a teacher just wanted him to tell her what to do, and explained his thinking about why he was unwilling to do so:

This isn’t about there’s a right way and a wrong way, and we see that with teachers, too. I go through a classroom, and I hear, “Was that right?” like there’s a right way to teach and a wrong way to teach. I don’t know if that’s personalities or whatever, but there’s this idea out there that I know what the right way is and I’m withholding it from them, and we’ve heard teachers say, “Carl, would you just tell us what you want us to do?” I wish it were that easy. I said, “Do you understand? I tell you all the time teaching is the
most complex job there is, and I respect that, and I respect how difficult your job is, and
if there was a magic bullet to say, ‘Here’s what we need to see’ . . . It’s not that easy.” …
The high-performing people are so reflective, I can ask one question and we just talk, and
those other ones, you can tell that it’s been a while since they’ve done that or have
actually thought about why they do what they do because maybe they’ve just been in it so
long or they’ve never been asked those questions. To me, our job is to get to a point
where everyone does that.

**Principals find that written feedback has minimal impact on teacher improvement**

**compared to feedback that occurs during conversations with teachers.** As described
previously, principals are required to provide specific, evidence-based written feedback to
teachers in the formative and summative reports (twice per evacuation cycle) and after each
observation of a teacher. Many principals perceive the written feedback, especially what is
provided in the formative and summative reports, has little impact on teacher improvement
compared to the effect that conversations between a teacher and a principal have. Leah shared
her thoughts about the written reports saying,

What I don’t find meaningful is, especially for myself, the quantity of the reports I have
to do. I cut and paste the rubric and then I have…. I put something in and it’s not as
effective as I would like them to be. I get more out of meeting with my teachers,
conferencing, than all this bullshit writing.

One principal did not feel that the written feedback was important in helping her teachers at all,
this was evidenced by her reflection,

As far as the written evals, the summative or the formative, it has never happened,
where somebody’s like, “Oh yeah, that’s really good feedback,” or “I’m going to
change something up because of something that you wrote about in those evaluations.”

It’s not those pieces that are changing it.

The requirement to provide written feedback to support the ratings in each of the four performance standards and overall at the mid-point of the school year for teachers who are on one-year evaluation cycles and at the end of the year for all teachers seems unnecessary and redundant to principals because the written feedback provided at these stages of the process is typically written feedback that has previously been provided to the teacher merged together into one document.

June reported feeling like her teachers are not “getting enough bang out of their buck from me writing out the summative or formative evaluations.” She went on the express her frustration that that the writing was redundant. Her teachers receive feedback from her on a regular basis, she did not feel that the written document at the end of the year provided them with anything other than a recounting of feedback that they had already received.

**Principals’ concerns about the detrimental effects of Needs Improvement and Unsatisfactory performance ratings** performance ratings make them hesitant to assign such ratings. There were many varied concerns related to performance ratings raised by principals interviewed for this study.

As described previously, at least twice in the evaluation cycle for each teacher, principals must rate the teacher’s performance on each of the four performance standards, and overall. These ratings are used to determine the evaluation plan that the educator will be placed on for the following year. In addition, for teachers in their first three years in a school district, a rating of Needs Improvement or Unsatisfactory in one or more standards or overall can result in termination of employment. The same holds true for teachers who are working under a Directed
Growth or Improvement Plan which has been previously put in place to rigorously support the teacher’s improvement in areas of need. All performance ratings for teachers are submitted to the Massachusetts Department of Elementary and Secondary Education at the conclusion of the school year. The results are aggregated by school and made available to the public through the DESE website. The ratings are often published in newspapers and noted on the major New England television news broadcasts at the time of their release each year.

Many principals expressed concern that the rating titles of Needs Improvement and Unsatisfactory set a negative tone for teachers. They did not feel that these ratings were consistent with the intention of the Massachusetts Educator Evaluation System to support teacher growth and development. One principal described it this way:

Again, we do that with kids. We don’t say kids need to improve. We say you’re not there yet, or you’re working toward, and this year we’re all about the growth mindset and using that language. This is fixed mindset language. You’re Unsatisfactory, you’re Needs Improvement, you’re Proficient, and I think a lot of staff feel, “That’s where I am and that’s where I stay,” but that’s not what the intent is.

Several principals suggested changing the Needs Improvement rating to something that is more positive and growth-oriented. Barbara, in reflecting on her concern related to the negative connotation of the ratings said,

I don’t like the ratings anyway. I think we need a “I’m growing” rating or “developing” or something. I think the Unsatisfactory rating … is very negative. If you read what it says, it’s super negative. I’d like a category where they could say, “I’m doing a good job, but this is where I know I could do better.”
One principal described that the Needs Improvement and Unsatisfactory ratings felt punitive. He worried because “when [someone] is struggling, the worst thing you can do is start beating them up. You’ve got to do things to build them up.”

Because of concern about the negative tone of the ratings, many principals find it difficult to assign ratings based on their judgments of teacher practice. Carl, in his first year in a new school shared,

I found myself as an administrator really erring on the side of caution because [I wondered], is [ranking a teacher Needs Improvement or Unsatisfactory] going to get me what I want? And really, I want to improve people. If a rating is going to derail that process, then I’m not [going to give them one of those ratings]. But at the same time, I want to give valid feedback, and ethically I have to tell you what I’m seeing and thinking, and so it’s really tricky.

The public nature of the evaluation ratings was also identified as a concern by principals. Although the public ratings are not associated with the names of teachers, they are associated with the names of schools and school districts, and some principals expressed concern about this aspect of the use of ratings. Barbara said, “I think the concern of making the data public is something that I’m not sure leads to professional growth for teachers.” Principals worry that members of the public who probably don’t understand the intention of the Evaluation System to support teacher improvement could form inaccurate perceptions of their teachers and their schools based on the published ratings. Carol shared her feeling that performance ratings of teachers could be misinterpreted by parents, who have little knowledge of the Evaluation System. “When parents look at those ratings they may not realize that just because a teacher’s
rated Needs Improvement, next year that teacher is probably going to work [his or her] way up to being Proficient.” She also shared a related concern,

One thing that I worry about is that when these ratings come out, and with school choice as it is in Massachusetts, that people are going to look at ratings and they’re going to say, “Well … only 10 percent of the teachers at this school are Exemplary and 20 percent Needs Improvement.” Do you know what I mean? I feel like they’re going to be able to shop, so to speak with school choice. I think there’s pressure on schools to not give their teachers appropriate ratings. If I start to think that my school choice numbers are going to suffer, or people aren’t going to want to bring their kids in [this town] to my school because certain percent of my teachers may be Needs Improvement, then that might hinder me from giving them an appropriate rating. I’m not saying that about myself, but I think it’s out there.

Christine also shared a concern about public perception saying “if they see a school district reports that they have 92 percent Needs Improvement, they’re going to be like, ‘Well fire all those teachers’….That was never the intent of the Needs Improvement label.”

Carol said, “For schools like me who are very, very small, it’s not hard to figure out who that Needs Improvement teacher is or how many Needs Improvement teachers you might have.” She fears that parents, community members and even educators in her building and in the school district could make assumptions about the identity of someone who has received a rating that is below Proficient, potentially causing damage to the professional reputation of a teacher.

Darlene, discussing the impact of the ratings and the cautious approach that she takes, revealed that she will ignore an observation of a teacher if it does not go well in order to allow the teacher to have another opportunity for a positive observation. She said of the ratings,
It’s crushing to them, but I will be honest, this year, anything negative that I’ve done, I’ve always . . . I’ve met with them ahead of time and I’ve said, “I think I want to do a redo on this,” just because I feel like I need to build those emotional deposits right now and we’re on a good momentum. I can’t afford for there to be something negative.

Some principals mentioned that they worked with other administrators in their districts to calibrate their expectations of teacher performance to the four ratings (Unsatisfactory, Needs Improvement, Proficient and Exemplary), but principals worry that there is not universal agreement about the ratings and that the lack of calibration is not fair to their teachers. Carl expressed his concern saying, “I think we’re all struggling in the district to clearly define what is Proficient, what is Needs Improvement.”

As described previously, based on the Massachusetts Educator Evaluation Regulations when a professional status teacher receives an overall rating of Needs Improvement or Unsatisfactory, it causes the plan on which their evaluation is based in the following year to be modified from a Self-Directed Plan to a Directed Growth Plan (if the rating is Needs Improvement) or an Improvement Plan (if the rating is Unsatisfactory). For teachers who are in their first two years of employment in a school district, the consequence of a Needs Improvement or Unsatisfactory rating may be dismissal. Such a rating will cause teachers in their third years to lose their jobs (Massachusetts Department of Elementary and Secondary Education, Final Regulations on Evaluation of Educators, 2011). Directed Growth and Improvement Plans require principals to provide more intensive and specific support to teachers, and to increase the frequency of observations and feedback to teachers in order to help the teacher improve to the Proficient level within a maximum one-year time frame. Directed Growth and Improvement Plans also require principals to painstakingly document every step in the process so that if the
teacher does not improve and if termination is recommended, the school district has all the necessary documentation to be used as supporting evidence should the teacher fight his or her dismissal. The concerns about rating a teacher Needs Improvement or Unsatisfactory in the case where such a rating would trigger a Directed Growth or Improvement Plan were two-fold. Some principals were worried about the amount of work that was involved. June expressed her concern that principals may hesitate to assign an overall rating that would cause a teacher to be placed on a Directed Growth or Improvement Plan saying,

I’ll be very honest with you. I think principals are fearful of giving [a rating of Needs Improvement] because of the amount of work that’s required to document everything. It’s an astronomical amount of work…. In the end, is it worth it? Yes, but you always have to second-think, “Oh do I really want to put myself through this?”

The second concern raised by principals related to Needs Improvement and Unsatisfactory ratings was that because these ratings result in changes that are often perceived as negative by teachers, it is commonplace for the union to become involved to help a teacher dispute the rating. Two principals described avoiding such ratings if they did not have an abundance of evidence, even if they felt it was deserved, in order to avoid union involvement.

**Research Question 3: How do principals perceive the effect of the Massachusetts Educator Evaluation System on the development and maintenance of trusting relationships between principals and teachers, in theory and as they have experienced it?**

The final research question is concerned with principals’ perceptions of how the new Massachusetts Educator Evaluation System impacted the relationship between principal and his or her teacher. Table 5 presents themes that emerged related to principals’ experiences with the effect on trusting relationships.
Principals find that it is difficult to establish an appropriate balance between accountability and support that is necessary to build trust with their teachers.

In most cases where the Massachusetts Educator Evaluation System has hindered a relationship:

- A negative rating initially caused the relationship to breakdown, but the relationship was repaired over time as the principal and teacher worked closely together to improve the teacher’s practice.
- In instances where the relationship was permanently damaged, principals pointed to the personality of the person to whom the recommendations for improvement were directed, rather than the Massachusetts Educator Evaluation System as the cause.

Principals build trusting relationships by trying to be as fair, clear and transparent as possible to ensure that teachers do not feel surprised about performance ratings.

Principals approach the Massachusetts Educator Evaluation System with a growth mindset and focus on providing help to their teachers what will allow them to accomplish improved teaching practice.

Principals perceive that trusting relationships are important to allow open and honest conversations about how to improve teaching practice, but recognize that these relationships require a great deal of time and effort to develop.

**Principals find that it is difficult to establish an appropriate balance between accountability and support that is necessary to build trust with their teachers.** Principals try to balance holding teachers accountable for the achievement of their goals and improvement in teaching practice with providing them with the help to make the necessary changes as they implement the Massachusetts Educator Evaluation System. Based on the comments of the participating principals, this balance seems difficult because they realize that in order for
teachers to improve, they often need help. In schools where the principal is the only leader, the principal is often the source of the help, as well as the person who ultimately will judge whether the teacher has been successful in his or her efforts to improve. Darlene characterized her struggle to achieve this goal in the following way:

I’m still grappling with that, finding a good balance between just principal, “Here I am, what can I do to help you?” and principal, “Here I am, make sure you’re doing your best, I’m here for your observation,” kind of thing. I’m still working that out myself.

In most cases where the Massachusetts Educator Evaluation System has hindered a relationship:

- A negative rating initially caused the relationship to breakdown, but the relationship was repaired over time as the principal and teacher worked closely together to improve the teacher’s practice.

- In instances where the relationship was permanently damaged, principals pointed to the personality of the person to whom the recommendations for improvement were directed, rather than the Massachusetts Educator Evaluation System as the cause.

In every case when a principal had to deliver difficult a message about a teacher’s performance, the teacher was initially upset. However, in many cases, as the principal and teacher worked together, the teacher began to believe that the principal’s goal was his or her improvement, and when improvements actually occurred as a result of the work that they were doing together, the relationship between the principal and the teacher improved. Amy described one such situation:

In the beginning, it was like, “I got you and I want to get rid of you.” But in reality, what I tell people is, it’s all what you make of it. So if you’re going to come in with that
attitude, that’s going to be the struggle. But if you come in like I’m trying to help you because I want my [teachers] to be at this level, and I need all my teachers to be there, then it’s like, let’s work together to get you there. That’s the bottom line. It’s not to get rid of you. It’s I need you to be at that level. That’s hard. That’s really, really hard, and it’s hard to hear. Tears and boxes of tissues off my desk, absolutely, but once we got past that, it helped. It took a long time … I think just seeing the support from the principal … Again, it was a struggle in the beginning because at first their defenses are up; “It’s not me, it’s the kids.” It’s always somebody else, but continuing just to break that down … It’s multiple conversations, and they don’t want to have them, but you have to have them.

Some principals recalled relationships that were permanently damaged as a result of the Evaluation System. In all of these cases, the teacher did not show improvement despite efforts on the part of the principal to support him or her. In some cases, the teacher was not able to accept the judgment of the principal and was not willing to listen to the feedback that the principal was providing. Barbara described the defensiveness she faced when she was trying to help a special education teacher:

Her kids weren’t making progress. We began having the conversation where I would say, “Okay, what kind of modifications are you making to their daily work?” She would say, “Well, the other teacher made these changes so I used this form.” I said, “Well, okay, this is the sheet you asked the kids to work on. What are the identified needs through the IEP [Individualized Education Plan] for this child? What have you identified as specific skill-based needs, and how does this form tie to that?” [The teacher would say] “Well, we’ve never done that before.” She was very used to doing complete pull-out model where she just went a lot slower with the same material. When I pushed on her practice, you know,
I wasn’t yelling at her, but I’m like, “Okay, well, how does this correlate with what you gave them to do? You as the special educator need to be making these changes to the work, to make them successful with grade-level work.” She was very old-school; she was not having it. So we had numerous conversations like that. She would get very heated, very heated, because I was targeting her. This wasn’t her role. I was confusing her. She didn’t know what I was asking her to do. I gave her examples; I worked through certain ones with her, to say, this is how I would change this number grid to be more beneficial to this child, knowing what we know about him and what he can do. She would use that but then a month later, when I would say, “Okay, now, show me what you’ve done with him,” none of that would happen. When I put that pressure on her about her practice, I never even had to get to the point where I had to write her up or move her up. She retired.

Christine found it difficult to work with a teacher who she felt needed to improve her practice and attributed the difficulty to the teacher’s insecurity. She recounted her efforts and the teacher’s resistance this way:

Feedback is huge, for some people you can give all the feedback you want and they don’t listen to it because they don’t take feedback. That’s challenging. [I tried] closing the door and having the honest conversation, being upfront, being honest. I do all those things and I still haven’t gotten through. I don’t know that I can. [It’s her] insecurity. I don’t have to think; I know what it is. It’s insecurity … It’s the person.

**Principals build trusting relationships by trying to be as fair, clear and transparent as possible to ensure that teachers do not feel surprised about performance ratings.** Many principals noted that they wanted their teachers to feel that they were not trying to catch them
doing something wrong and used the term “not a gotcha” when they described their attempts to be open and honest in their work with teachers. Principals described making it a priority to have a conversation with a teacher after a classroom observation and felt that the discussion that took place helped teachers know exactly what the principal was thinking. Carl described his efforts as follows:

When I observe [a teacher], we have a meeting. I always want to meet with [the teacher], talk with [the teacher] and come to some conclusions about next steps…I’m trying to eliminate surprises in their formatives and summatives, and in their observations. I don’t want anyone to get something and go, “I didn’t know that you felt that way.”

Principals felt that the no-surprises strategy helped build trust for the process and improved their relationships.

Principals also spoke about the need to be clear and transparent with expectations. Some principals reported that their school districts had developed focus standards or had specifically delineated priority areas that principals would focus on during evaluations. These principals felt that this practice provided clarity and improved transparency. Jen described the process used in her district to choose only seven indicators across all four performance standards for collection of evidence during evaluation. She explained the reasoning behind the decision saying, “We didn’t want to overwhelm anybody and we didn’t want [the teachers] to feel like we were trying to catch them.” Carl, a new principal wanted to ensure that his teachers understood the System and his expectations as he began to work with them. He said,

I really wanted to make sure [what] I did with my staff that [first] year was spend a lot of time clarifying everything, and what is the process. We did a lot of work around that, how they were going to be evaluated.
Fairness was also mentioned as a necessary component for strong relationships that support teacher growth. As described previously, one principal discussed allowing a “re-do” when an observation did not go well. Another principal said,

They also know I’m going to be fair, I’m going to care for them, and then that can translate into when I’m really talking about their practice, which is very personal to many of them, that we can have an honest conversation that nobody gets slammed, nobody gets yelled at, nobody feels diminished, but the expectation being, we’re always going to get better. We’re just going to keep getting better.

Additionally, several principals talked about fairness in the context of needing to hear from a teacher before making decision about what they had observed, to ensure that their interpretation of an event or teaching practice was accurate.

If there’s something that I’m making a recommendation for, a suggestion for, I always ask the question. We can never make assumptions. When you’re in a classroom and you’re only there for 15 or 20 minutes, you can’t assume to know why someone did something. You have to allow them to explain what they were thinking when they implemented, or they chose to do, or not to do something. Why was your objective on the board or why wasn’t your objective on the board? We have to find out where they’re coming from … I want to make sure that I’m being fair and I’m being honest and I’m doing everything I can to help them improve.

Principals approach the Massachusetts Educator Evaluation System with a growth mindset and focus on providing help to their teachers what will allow them to accomplish improved teaching practice. Throughout the interviews, principals referenced the supportive nature of the principal teacher relationship. All principals noted multiple times that they
consistently reinforced with their teachers the idea that evaluation was about improvement; that they believed that teachers would improve and that the principals felt it was their “job to provide them with the resources to help them get there.” This is especially significant given that all of the principals participating in this study work alone to lead their buildings. Providing supports for teachers can be very time consuming. A principal working alone is already incredibly busy. To engage in additional work to help a teacher or teachers improve demonstrates that these principals believe that their efforts will pay off in improved teacher practice and student achievement. One principal illustrated this in her comment:

The door is always open. It is hard. It’s really important, but I’m their support system. I’m the person that they go to to get the help because they work with their colleagues and now they need to bounce it off of somebody else, or get information. That’s really important. In order to do the teacher’s job well, they need to be supported.

In describing the supports that they provided for teachers, principals tried to find just the right way to help each individual. In some cases, this meant that the principal and teacher sat down and developed a plan together. In other cases, the principal developed the plan after considering what he or she knew about the teacher’s deficits and the way the teacher would respond to the support offered. Diane described supporting a teacher by providing her with very “specific, explicit feedback, because [she] knew that’s what she needed … She needed it spelled out.” Some principals provided their teachers with professional development that was targeted to their areas of need. One principal described working in the classroom with a teacher to demonstrate exactly what he wanted her to do. This same principal established a weekly meeting with himself, a reading specialist, a special education teacher and a teacher in need of improvement so they could collaboratively plan the next week’s lessons together. For another
teacher this same principal started a book club with other teachers the building and chose a book for the group that targeted a particular teacher’s need. He chose the book club route because he knew that the teacher who needed the help would benefit from collaborating with colleagues, but would not be likely to reach out to her colleagues for support on her own.

Principals felt that the feedback that they provided to teachers, both orally and in writing was intended to support improvement. They took care when providing feedback to say or write things in a manner that would allow the teacher to be receptive. Christine’s comment captures the sentiment shared by many of the principals:

I want to give the teachers information…to support their growth. You have to word things so that people can be receptive, so you’ve got to be careful. You’ve got to know the person to give them the feedback that they’re going to hear, that they’re going to listen to.

One principal shared that because her teachers know that she will provide support and resources for them “it’s made [them] be more open to trying new things.” Another shared that knowing support was available allowed her teachers to feel that they “could ask for help.”

Interestingly, some principals reported knowing that other principals did not view the Evaluation System as a tool to support teacher growth, and they found this troubling. Jocelyn’s comments characterize the views of these principals.

One of the principals that was in the course [that I was taking] spoke so negatively about [his] teachers and about evaluations, and I thought, he doesn’t get what the tool is for. He doesn’t understand how to use it to your benefit. I think he uses it as a punitive thing. Just listening to him, I had the feeling like, because he said things like, “Well, I say to
them, you just wait and get your evaluation.” I thought, “Oh. Why would you ever use this as a negative?”

Principalst perceive that trusting relationships are important to allow open and honest conversations about how to improve teaching practice, but recognize that these relationships require a great deal of time and effort to develop. All principals talked about the need to be cognizant of the relationships with their teachers and acknowledged that such relationships developed over time. Interestingly, principals who had led a school for more than four years were more likely to feel that they had strong, trusting relationships with their teachers prior to the implementation of the Massachusetts Educator Evaluation System and did not feel that the System had affected those relationships in any way. Amy, a longstanding principal shared, “I’m feeling like my staff is in a good spot, but again, that doesn’t happen overnight. … That’s something that’s been developed and grown and developed over time. That’s 11 years’ worth of work.”

Principals who had been working in their schools for fewer than four years were more likely to express concerns about the relationships that they had with their teachers. One principal in his first year in a school said,

I’ve been really careful this year about having an open mind, being available to teachers, even those ones I’ve heard are really tough and won’t change, really going slow at them and really trying to be empathetic and get them to take small baby steps toward improvement…. Again, it’s baby steps, it’s trust building. I’ve seen it improve over a year here, and I would hope to think that it’ll continue, but I want to be really careful. That’s why I’m trying to be really careful about these evaluations right now. I really feel strongly that one check box could negate the whole thing.
One indicator of a strong relationship mentioned by principals was whether or not their teachers felt comfortable coming to them for help or inviting them into their classroom to observe. June, a principal with only two years in her position, shared that she felt her teachers were receptive to her feedback saying, “I think they’ve demonstrated that by calling me down with something, ‘Oh you’ve got to see this. This is exactly what we were talking about.’ I will run down there to see what they’re doing.” Two other new principals both noted that they did not feel this was the case yet in their buildings. Mitchell lamented that his teachers do not yet feel comfortable coming to him and sharing their challenges saying, “I think that’s the difficult part. That’s the piece that takes a long time.” Carl shared a similar sentiment:

I want [my teachers] to ask me, “Hey, Carl, can you come watch this? I’d like to get your input.” We’re not really at that point where it’s comfortable just because there is that at the end of the day, I do have to evaluate you.

Regardless of whether or not they were working to establish strong trusting relationships with their teachers or whether or not they believed that the relationships existed in their schools already, principals acknowledged the significance and positive impact a trusting relationship between the principal and his or her teachers has. One principal said, “I work hard to build a relationship of trust where people can come in and be honest and open and know that I’m here to support them.”

**Summary of Findings**

The findings of this study were collected through semi-structured interviews conducted over the telephone or in-person with 11 elementary level public school principals, each from a different school district in Massachusetts. Principals participating in this study had all used the Massachusetts Educator Evaluation System for at least two years to evaluate teachers, and each was the only evaluator of teachers in their school building.
Regarding the perceived effect on teacher growth and development:

- Principals perceive that the rigor of the rubric, the use of specific evidence as part of the feedback that they provide to teachers, and the development and progress monitoring of student learning goals and professional practice goals components of the Massachusetts Educator Evaluation System support teacher growth and development.

- Principals perceive that the accountability aspect of the Massachusetts Educator Evaluation System, which requires an annual rating (Unsatisfactory, Needs Improvement, Proficient and Exemplary) of teacher performance, causes teachers to focus more on the achievement of a rating than on professional growth.

Regarding the perceived effect on principal capacity, ability and opportunity to support teacher growth and principal teacher relationships:

- Principals perceive the requirement for frequent and short classroom observations helps them provide teachers with high-quality feedback that supports their growth.

- Principals perceive that the written feedback required in the System is very time consuming and in some cases redundant, and they do not perceive that it has a high impact on teacher growth and development.

- Principals worry about the impact of performance ratings on teacher growth and development, and they spend a great deal of time thinking about how to use the performance ratings in a way that accurately communicates a teacher’s performance level without damaging the principal-teacher relationship or de-incentivizing teachers to improve.
• Principals value the collaborative conversations that they have with their teachers during which they discuss feedback because they feel that these discussions foster teacher reflection and support strong relationships.

• Principals are committed to helping teachers improve and are willing to provide varied supports targeted to individual teacher needs.

In summary, principals felt that the Massachusetts Evaluation System has positively impacted teacher development in their buildings, however they are concerned that some of the requirements of the System may impede teacher growth and development. Principals seem frustrated that some of the less effective parts are very time consuming, which limits their ability to spend adequate time with the components that they perceive to be more valuable. They specifically noted collaborative conversations with teachers as very beneficial to teacher growth and development, and strong relationships.
Chapter V: Discussion of the Findings

Chapter V begins with a review of the problem of practice and methodology, presentation and discussion of the key findings, and presentations of the findings in relation to the theoretical framework and literature review. The final sections of the chapter include a conclusion and significance of the study, limitations of the study, and recommendations for future studies.

Revisiting the Problem of Practice

One of the primary goals of effective supervision and evaluation systems is the improvement of teacher practice and, in turn, increased student achievement (Peterson, 2000). However, with the exception of three very recent studies, research on teacher evaluation strategies have not shown that evaluation had caused teachers to improve their practice. The Massachusetts Educator Evaluation System adopted in 2011 and implemented in phases through 2017 has two stated purposes: to promote student learning, growth, and achievement by providing educators with feedback for improvement, enhanced opportunities for professional growth, and clear structures for accountability, and to provide a record of facts and assessments for personnel decisions (Massachusetts Department of Elementary and Secondary Education, Final Regulations on Evaluation of Educators, 2011)

This System incorporates supervision (also called formative assessment) with evaluation (called summative assessment). The purpose of supervision is to provide non-judgmental feedback to teachers to help them improve their practice. When engaging in supervision, the role of the supervisor is similar to that of a coach. Evaluation has a different purpose; it is designed to assess performance for accountability and employment decisions. Evaluators must judge teacher practice to determine the level of performance that is most appropriate. Principals who are the only evaluators of teachers in their schools are required to perform both functions as they use the Massachusetts Educator Evaluation System. The implementation of a new evaluation
system may present challenges for principals, not the least of which is the fulfillment of two roles with dichotomous purposes: supervision and evaluation. This study was designed to examine the Massachusetts Educator Evaluation System in its early stages of implementation and to determine the principals’ perceptions of its effect on the achievement of the goal to support teacher development as well as what effects it was perceived to be having on principals’ ability to support teacher growth and development and the principal-teacher relationships.

**Review of the Methodology**

This qualitative study explored the perceptions of public school elementary level principals in Massachusetts relating to the Massachusetts Educator Evaluation System. Principals participating in the study had at least two years’ experience using the System, and were the sole evaluator of teachers in their school building.

The study was designed to address the following research questions:

1. How do principals perceive the Massachusetts Educator Evaluation System as supporting teacher growth and development, in theory and as they have experienced it?

2. How do principals perceive the effect of the Massachusetts Educator Evaluation System on their ability, capacity and opportunity to effectively support their teachers’ development?

3. How do principals perceive the effect of the Massachusetts Educator Evaluation System on the development and maintenance of trusting relationships between principals and teachers as they have experienced it?

The research questions were addressed through semi-structured interviews. Each interview was transcribed immediately after it concluded, and the transcript was reviewed by the
researcher for emergent themes as soon as possible in order for the researcher to consider possible revisions to the interview protocol that could enhance the responses from future participants. In Vivo coding was employed as the first round coding strategy in order to capture the language of the participants (Saladaña, 2009). Pattern coding followed as the second round coding strategy which allowed the researcher to develop broad categories that revealed themes (Saladaña, 2009).

**Presentation and Discussion of Key Findings**

A thorough review of the data collected from interviews produced multiple themes that support the following key findings:

1. Principals see virtue in a formal system to support teacher growth because it clearly defines desired teaching practices, and because it requires them to spend time in classrooms seeing authentic teaching. The quality of written and verbal feedback that they provide to their teachers is improved due to better understanding of the practice of each teacher.

2. Principals find that the performance ratings used in the Massachusetts Educator Evaluation System impede teachers’ recognition and use of it for the purpose of professional growth and development.
   - Rating titles of Exemplary, Proficient, Needs Improvement and Unsatisfactory imply that performance is fixed rather than constantly improving.
   - The public nature of the accountability ratings orients teachers to focus on achievement of a particular rating, rather than focusing on using evaluative feedback to improve their practice.
3. Principals find that there are components of the Massachusetts Educator Evaluation System that are very challenging to implement well, and those components are not helpful to them in supporting teacher growth and development.

- The written feedback necessary after classroom observations and for the formative and summative stages of the evaluation process is very time consuming and is not as valuable as other components of the System.
- Their concern about the impact of performance ratings on teachers and their school causes principals to be very strategic in their assignment of ratings, and to devote a great deal of time and effort to minimizing the negative impact of performance ratings in order to maintain trusting relationships with their teachers.

4. Principals wish they had time to increase the frequency and number of collaborative discussions that they have with teachers as they share their feedback because these discussions help to clarify teaching practices, build trusting relationships, and foster reflective practice for teachers.

**Key finding 1: Principals see virtue in a formal system to support teacher growth because it clearly defines desired teaching practices, and because it requires them to spend time in classrooms seeing authentic teaching. The quality of written and verbal feedback that they provide to their teachers is improved due to a better understanding of the practice of each teacher.** The new Massachusetts Educator Evaluation System was repeatedly compared to the dog-and-pony show by principals, a reference to previous evaluation systems used in Massachusetts which required scheduled observations that seemed more like performances than authentic teaching. Although it is difficult to find the time, principals appreciate that they are required under the new System to be in classrooms frequently for short
periods of time, and that they no longer need to make an appointment with a teacher to conduct an observation for evaluation. This flexibility allows principals to see what happens regularly in a classroom, and enables them to glean an accurate perception of the usual practice of a teacher. With this information, they can craft feedback that accurately highlights a teacher’s strengths and targets specific areas for improvement. The feedback that they provide teachers is, therefore, more precisely crafted for each teacher.

The accuracy of principals’ knowledge of teachers’ practice also influences their ability to make correct judgments of teacher performance. They rely on the rubric to guide these decisions. Principals spoke positively about the descriptors of teaching practice for each level of performance (Unsatisfactory, Needs Improvement, Proficient and Exemplary) in the rubric for this use. Principals also praised the descriptors of effective teaching practice in the rubric because when teachers and principals both use them to guide evidence collection, decisions about performance levels, and evidence-based feedback discussions, the rating descriptors help develop a common understanding of effective teaching practice within their schools. The rigor of the rubric was said to have given teachers something to which they can aspire. One principal in describing an interaction with a teacher who asked her how to reach the Exemplary level, referenced the rubric, telling the teacher that to be rated Exemplary she had to model the practices for her peers. The conversation motivated the teacher to offer an after-school workshop for other teachers in the district. The motivation to take her practices to the next level was the impetus for this teacher to teach her peers, but as the principal noted, the teacher gained confidence and validation as well, “It’s when they see the sign-up sheet and 15 people have signed up to come and see them, that’s huge, on a Wednesday afternoon after school.”
There was consensus among principals about the goal of the System to foster continual improvement, and principals shared many examples of teachers who had made changes to their practices as a result of feedback provided through the System. They expressed genuine excitement when they described these scenarios and hoped for a time when all teachers would embrace feedback as an opportunity for growth and development.

Key finding 2: Principals find that the performance ratings used in the Massachusetts Educator Evaluation System impede teachers’ recognition and use of it for the purpose of professional growth and development. Principals believe that the primary purpose for evaluation is the improvement of teaching, but they do not feel that the rating system is aligned with that purpose, and they wish that there would be changes to the names associated with each rating, and to the ways in which the ratings are used within the System. One principal, expressing her frustration with the fact that ratings were necessary for the accountability aspect of evaluation and doubting that changes would be forthcoming said, “[The ratings are] the biggest part of this whole process, right? [We] report to [the DESE] if we [have] good teachers or not, so how else do you do that without a rating?” Principals understand that the rating system allows the State of Massachusetts to hold districts and administrators responsible for completion of evaluations and that it provides data to the DESE about the quality of educators in school and districts. However, they are troubled by the impact that it has on the attitude of their teachers toward the evaluation process.

Rating titles of Exemplary, Proficient, Needs Improvement and Unsatisfactory imply that performance is fixed rather than constantly improving. Comments related to this finding show that principals feel that the ratings of Needs Improvement and Unsatisfactory are harsh and negative. Principals are bothered by this and think that changing the names associated with each
performance level would help shift teachers’ thinking about the ratings. One principal insisted that she was trying to “ingrain the mindset of ‘We can just always get better,’” and believed if that was possible, the goal of continuous improvement would be realized for all teachers. All of the names associated with the performance levels send the message that performance is static according to principals, and this interferes with their attempts to focus teachers on improvement rather than rating levels. Several principals suggested more growth-oriented rating names, such as “developing” or “growing.” One principal even suggested “a category where [teachers] could say, ‘I’m doing a good job, but this is where I know I could do better,’” to validate good practice but orient teachers to make improvements as well. This principal goes so far as to tell her teachers to rate themselves Proficient as they complete their self-assessment, but to include a statement about what their next step will be for each element.

*The public nature of the accountability ratings orients teachers to focus on achievement of a particular rating, rather than focusing on using evaluative feedback to improve their practice.* Currently, the goal for many teachers, according to principals, is to achieve a rating rather than to receive productive feedback. One principal reported that her teachers counted the number of times she had been in their classrooms and when she reached the minimum number of observations required, they expressed relief that they were “done.” This implies that her classroom visits seemed more like a burden than a help, and that the process was being done to them, rather than done with them. Another principal described asking a teacher if she minded that her summative evaluation was not going to be completed on time, the teacher said, “No, actually, put me last. I don’t even care.” The principal in this case implied that many on her staff would have reacted similarly because they feel that they have achieved proficiency
and are satisfied with that. Principals wish for a system where the feedback received was valued more than the rating.

The publication of performance ratings is perceived by principals to contribute to the perception that teachers and schools are being graded by the Evaluation System. They worry that the ratings cause individual teachers, schools and school district to be inaccurately judged by parents, community members and colleagues. The publication of performance ratings, coupled with public lack of understanding that the primary purpose of the System is continuous improvement, reinforces the concept that the ratings are levels of achievement, and make it more difficult for teachers to accept them as opportunities for growth and development.

Key Finding 3: Principals find that there are components of the Massachusetts Educator Evaluation System that are very challenging to implement well, and those components are not helpful to them in supporting teacher growth and development. All principals communicated their strong desire and commitment to improve teaching and learning in their schools and expressed frustration that some requirements of the Evaluation System were not useful to them in accomplishing this goal. They were primarily concerned about the requirements for written feedback and ratings.

The written feedback necessary after classroom observations and for the formative and summative stages of the evaluation process is very time consuming and is not as valuable as other components of the System. Principals, especially those who lead buildings by themselves, have numerous demands placed on their time, many of which are out of their control. The Massachusetts Educator Evaluation System has placed additional demands on a principal’s available time since its implementation in 2011. Principals reported that they spend more time in classrooms and discussing teaching with teachers, which they value, but they also spend much
more time writing reports that contain feedback. The model contract language developed by the DESE to align with the regulations and adopted by all districts involved in this study specifically requires written feedback to teachers within three-to-five days of a classroom observation, and at the formative assessment, formative evaluation and summative evaluation stages of the process (Massachusetts Department of Elementary and Secondary Education (2012b). Principals commented that the written observation reports often felt redundant and unnecessary because written observation reports are often a reiteration of information that they have previously discussed with a teacher, and the formative and summative reports are a summation of evidence and feedback previously provided verbally and in writing.

Because principals are concerned that the written feedback is not only accurate but helpful, they spend a great deal of time considering exactly what they will say and how they will say it so as not to damage the relationship or “deflate” a teacher. After all the time they spend on the written feedback, they do not feel that it has a great impact in supporting teacher growth. One principal summed up her frustration with lack of adequate time as she spoke about trying to be patient in conversations with teachers to provide enough time for an in-depth discussion:

[You need to have] patience [and] when we’re working under the time pressures we work under, sometimes [that] is really hard to maintain. That’s the kind of bugger in the whole thing… I think the System has great power and great potential to build a community of learners that will always want to get better, but if we don’t have the time to take to build the relationships, to have the conversations, to think about as administrators what we’re going to say and not [say], I don’t think it’s going to work.

*Their concern about the impact of performance ratings on teachers and their school causes principals to be very strategic in their assignment of ratings, and to devote a great deal*
of time and effort to minimizing the negative impact of performance ratings in order to maintain trusting relationships with their teachers. Because principals feel that the ratings imply static performance and do not reinforce the concept of continual growth, and that the Needs Improvement and Unsatisfactory ratings seem harsh, they tread lightly when assigning performance ratings below Proficient to teachers. One principal reported that it was very difficult to put a negative rating down in writing because she “felt bad about it” and worried that a rating would “make [the teacher] feel defeated … shut down or … angry.” Another principal who was new to the building asked the superintendent of schools for permission to “give [all teachers] the rating that they’ve had [in the past with the previous principal]” so as not to upset them the first year under new leadership. The superintendent supported this strategy. Two principals reported avoiding negative ratings by allowing feedback resulting from an observation that was not positive to be ignored and revisiting the same classroom in future hoping for a better result. One principal summed up the feelings of many when he said “The feedback trumps the rating.” This likely why principals feel somewhat reluctant to assign ratings of Needs Improvement or Unsatisfactory to teachers.

**Key finding 4: Principals wish they had time to increase the frequency and number of collaborative discussions that they have with teachers as they share their feedback because these discussions help clarify teaching practice, build trusting relationships, and foster reflective practice for teachers.** All principals saw the relationships that they build with their teachers when they talk with them about their teaching practice as one of the most beneficial aspects of the Massachusetts Educator Evaluation System. Interestingly, the regulations and model contract language adopted by all districts included in this study do not require principals and teachers to meet to discuss evaluative feedback, except at the time of
completion of the formative assessment, formative evaluation, or summative evaluation documents in the following limited circumstances: if a meeting is requested by an evaluator or teacher or when a teacher has been rated Unsatisfactory or Needs Improvement overall (Massachusetts Department of Elementary and Secondary Education (2012b). However, the principals in this study were adamant that these discussions were crucial, and that they do not have enough time to have them as often as they wish they could.

The discussions with teachers help principals better understand what they have witnessed during a classroom observation, and how their teachers make decisions. Knowing more about what they observed and how the teacher’s thought process during the observation influenced what he or she did helps principals make more accurate judgments and provide better feedback. In describing this, one principal said “Sometimes I have questions [about what I saw in a classroom and] once I get clarity … it gives me an understanding of what I was seeing, and that could take [my feedback] in a different direction.”

The trusting relationship that forms between a principal and teacher when they have discussions about teaching was cited as an additional benefit. One principal described that even when an observation had gone bad, the conversation that she had with her teachers was an opportunity for her to reinforce that she was there to support them,

They…know I’m going to be fair, I’m going to care for them…that we can have an honest conversation that nobody gets slammed, nobody gets yelled at, nobody feels diminished, but that [the] expectation [is], we’re always going to get better…even if things don’t go well, it’s going to be a fair and reasonable conversation and that we’re going to talk about next steps.
Conversations with teachers provide an opportunity for principals to engage their teachers in reflective practice. This was mentioned by principals as one of the most significant benefits to the discussions that take place after an observation. As characterized by one principal, “I really get them to try to reflect through questions, to get them to what I might see as a possible improvement … I think I shouldn’t always have to be there to encourage their thinking.” Many principals spoke about the need for teachers to become reflective practitioners who constantly independently reflect on their own practice and use that reflection to make decisions about changes in curriculum, instruction and assessment.

**Discussion of Findings in Relation to the Theoretical Framework**

The theory of relational trust developed by Byrk and Schneider was chosen as the lens for this study because the theorists claim that relational trust plays a role in moderating the “sense of uncertainly and vulnerability that individuals confront” (p.33) as they face school-reform movements that often demand significant change in a short period of time. “Trust reduces the sense of risk associated with change” (Byrk & Schneider, 2003, p.43).

In this time of educational reform, teachers and principals are immersed in change. Use of the new Massachusetts Educator Evaluation System, in and of itself represents a significant change for many principals and teachers, and if it meets its stated goals, teachers and principals will be engaged in making regular and repeated changes to the practices they use in instruction, assessment and curriculum. When relational trust exists, in theory, these changes will be easier to enact and maintain in the school setting.

**Theory of relational trust.** There are four components to relational trust: Respect, Personal Regard for Others, Integrity and Competence. Each aspect must be developed in order for relational trust to exist between two or more people (Byrk & Schneider, 2002).
**Respect.** Relational trust is grounded in “respectful exchanges … marked by genuinely listening to what each person has to say and by taking these views into account in subsequent actions” (Brewster & Railsback, 2003, p.42). Respect is built when principals are interpersonally and professionally open (Blumberg, Greenfield & Nason, 1978) and involve teachers in decision making, thereby allowing and truly considering the contributions of those with dissenting viewpoints (Blasé & Blasé, 2001; Blumberg, Greenfield & Nason, 1978). Within the practice of supervision, principals can build respect by engaging in open and honest discussions that allow teachers to share and discuss their perspectives without fear of judgment.

Principals in this study valued the opportunity to have open and honest discussions with their teachers about teaching practices. They specifically noted the importance of speaking with a teacher after a classroom observation to gain a better understanding of what they had witnessed as an observer to ensure that judgments they made and feedback they provided were accurate and fair. This demonstrates an effort on the part of principals to engage in respectful interchanges with their teachers, and to listen to and consider what their teachers tell them prior to making up their mind about what took place. One principal described her respect for teachers this way: “I think I’ve always been really open and honest with my teachers and I have a really good track record with them of listening to them and giving them the benefit of the doubt.”

**Personal regard for others.** “Principals establish both respect and personal regard when they acknowledge the vulnerabilities of others, actively listen to their concerns, and eschew arbitrary actions” (Byrk & Schneider, 2003, p.43). Actions taken to show consideration and sensitivity for the needs and personal wellbeing of others (Blumberg et al., 1978; Sebring & Byrk, 2000), as well as to reduce others’ vulnerability further demonstrate personal regard (Byrk & Schneider, 2002).
It was evident in the data collected from interviews that principals valued the well-being of their teachers, and they made ensuring that their teachers’ needs were met a high priority in their work. All principals spoke passionately about finding ways to support what their teachers needed to make necessary improvement to their practice. They indicated that they felt it was their obligation to do whatever it took to help a teacher improve and described spending great amounts time and effort to customize supports so that they would be most helpful to the individual.

One principal described his thoughts when another principal from his district expressed lack of care and concern for a teacher who was underperforming,

[I thought] have you asked them? Are they intentionally [not teaching well]? I find that hard to believe. Sometimes you just need to be people too and take the human side of it. Let’s be a little empathetic toward people who are underperforming. Let’s find out why. Let’s not assume that they’re just doing it because they don’t want to be better. I don’t think they know how to do better, and if they did, wouldn’t you think they would?

Evidenced in the comments from another principal was the importance that she placed on letting her teachers know that they were cared for.

This is a job. [The people in this school are] a second family. We will do what we can to support one another … I will move the earth if somebody needs to get out of here because there’s a family issue. You know what I mean? We’re just going to take care of each other that way … You have to show them that you’re going to be consistent with that belief. That you are here to support them.

She continued, saying when her teachers knew that she cared for them, the personal conversations about their teaching were much easier to have.
**Integrity.** Clear and accurate communication, honesty, fairness, consistency, predictability, and commitment to follow through on promises are perceived as acts of integrity (Blumberg et al., 1978; Byrk & Schneider, 2002; Costa & Garmston, 2002; Handford & Leithwood, 2013; Tschannen-Moran, 2001), as are modeling of attitudes behaviors and commitments that one expects of others (Dufour & Eaker, 1998; Handford & Leithwood, 2013). “Teachers associate consistency with predictable patterns of action by principals, timely feedback about classrooms and instructional activities, availability of classroom materials and supplies … routines related to discipline, and regular demonstrations of involvement with children and/or their families unrelated to discipline” (p.204).

Fairness was paramount in the minds of principals as they discussed evaluation. They discussed wanting to talk with a teacher to ensure that their conclusions after an observation were on point, allowing an observed lesson that did not go well to be ignored and carefully comparing their evidence to the descriptors in the rubric to ensure that their judgments were accurate, all in an effort to ensure that they were unbiased as they used the Evaluation System.

Communication of feedback was an area of concern for principals in this study. They loved the opportunity to provide verbal feedback during face-to-face conversations with teachers, but felt uncomfortable with the written feedback and worried that if it wasn’t stated perfectly it might send the wrong message. The concern about the written feedback caused principals to spend a great deal of time crafting written documents that would be clear and truthful, but would encourage teachers to grow. As a group, they did not seem confident that they were accomplishing this goal.

**Competence.** Overall, the ability of principals to fairly, effectively and efficiently perform professional responsibilities is perceived by teachers as competence (Costa and
Garmston, 2002; Dufour & Eaker, 1998). More specifically, Handford and Leithwod (2013) found that “functional, work-related skills” are most significant for teachers in perceptions of principal competence. Among the skills identified were “being visible, especially in the classroom; and principals’ formal and informal engagement in classroom observations and other classroom based activities such as instructional planning with teachers and providing teachers with feedback about their instruction” (p. 202).

Demonstration of professional competence was important to principals participating in this study. First and foremost, principals felt that their professional obligation was to ensure that they did what was right for the children they serve. This goal was what caused them to focus time and attention on the improvement of teaching. One principal commented that in order to build trust, he knew what he needed to do when he said, “You have to know what’s going on in your classrooms, and you have to be visible, and you have to be the instructional leader. But to be that, they need to know that you understand what’s going on.” Another principal worried that she might not be competent to give appropriate feedback to high-performing teachers, knowing that she had not been in a classroom in a long time.

What am I going to tell them? That sometimes worries me because you want your feedback to be valuable to everyone, no matter what level they are at … sometimes I wonder, do I have the expertise to give them the feedback they need?”

Discussion of Findings in Relation to Literature Review

Chapter II included a comprehensive review of the literature related to teacher evaluation, specifically examining the distinction between supervision and evaluation, the impact of federal mandates on teacher evaluation and current trends in evaluation. Additionally, the link between teacher effectiveness and student achievement was examined, as was that effectiveness of
evaluation and collaboration in improvement of teaching and learning. The sections that follow will describe links between study findings and literature reviewed.

**Supervision vs. evaluation.** Supervision is intended to provide educators with information that helps them improve (Marzano et al., 2011). When carrying out supervision, principals and teachers engage in inquiry-based collaborative discussions that help teachers understand their strengths and areas for improvement, and help them make decisions about changes that they might make to improve their teaching. Discussions such as these require the principal to take a non-judgmental stance that encourages teachers to discuss challenges and openly explore possible actions for improvement (Nolan & Hoover, 2011). The role of supervisor is said by some to resemble to role of coach (Danielson & McGreal, 2000; Showers, 1985). The primary purpose of evaluation on the other hand, is the judgment of teacher practice (Nolan and Hoover, 2011; Sergiovanni & Starratt, 2007) for accountability purposes and employment decisions (Nolan, 1997; Papay, 2012; Sergiovani & Starratt, 2007).

There is ongoing disagreement in the literature as to the compatibility of the roles of supervisors and evaluators of teaching. Some feel that the roles are inextricable (Ribas, 2011; Tomal et al., 2015), others believe they can be successfully carried out by one individual if they are clearly delineated (Sergiovanni & Starratt, 2007) and still others believe that they conflict and should never be handled by the same individual (Cogan, 1973; Nolan, 1997; Popham, 1998).

The name of the Massachusetts Educator Evaluation System connotes the accountability purpose, yet as it was designed, it incorporates both the accountability function and the supervision function. The regulations state

The specific purposes of evaluation under M.G.L. c.71, §38 and 603 CMR 35.00 are:
(a) to promote student learning, growth, and achievement by providing educators with feedback for improvement, enhanced opportunities for professional growth, and clear structures for accountability, and
(b) to provide a record of facts and assessments for personnel decisions. (Massachusetts Department of Elementary and Secondary Education, Final Regulations on Evaluation of Educators, 2011)

Study outcomes reveal that principals struggle primarily with the accountability role. They reported feeling uncomfortable with the rating system that was used to judge teacher performance and felt that it did not align with the intended purpose of promoting the continuous improvement of teaching practice. They called for changes to the names of performance levels to ratings that were more indicative of a progression of teacher learning than the current ratings, which they believe imply fixed levels of performance. They further questioned the merit of making public the aggregated performance ratings by school and district, and indicated that they felt this aspect of the System was flawed due to lack of calibration across the state and even within districts. They perceive the rating system to have many detrimental effects on educators, schools and districts.

Principals embraced the role of supervisor, some called it “instructional leader”, and although they found it challenging primarily due to time constraints, they felt that many teachers responded positively to verbal feedback and had made changes to their practice as a result of discussions with their principal. Over and over again during interviews, principals’ comments revolved around the benefits they perceived as a result of the face-to-face supervisory discussions with teachers about effective teaching practices. It is evident that the conflict between supervision and evaluation is real for the principals who participated in this study.
Many implied that they were working to reconcile the roles, but none expressed that he or she had successfully done so.

**Impact of federal mandates on teacher evaluation in U.S. and Massachusetts.** The Massachusetts Educator Evaluation System was developed and enacted in response to a requirement of the award of 250 million dollars in federal funds from the Race to the Top grant. Among the priorities areas identified in the Massachusetts grant was a commitment to great teachers and leaders, which was addressed through the development of a new Educator Evaluation System. The new System was designed to address identified inadequacies in the existing systems used across the state, and to align with research related to effective evaluation systems (Massachusetts Department of Elementary and Secondary Education, 2011).

A lack of common understanding of effective teaching practices (MassPartners, 2002) and the prevalence of a two-scale rating system that failed to adequately differentiate educator quality (Massachusetts Department of Elementary and Secondary Education, 2011) were both cited as inadequacies in teacher evaluation in Massachusetts as the new System was under development. The results of this study indicate that use of the rubric implemented with the 2011 System is helping to build a common understanding of effective teaching practices, and the four performance levels delineated in the rubric made it clear to both principals and their teachers what the appropriate rating of performance should be. The descriptors in the rubric were explained as valuable to both teachers and principals in explicating the practices that effective teachers implement. Common understandings of effective teaching practice help principals provide more effective feedback that is targeted to discrete elements of teaching practice. When suggested changes to teaching practice are specific, they are likely to feel more manageable for teachers. Principals said that the differentiation between the levels as described in the rubric
helped reduce conflict between evaluators and teachers caused by disagreements about performance judgments.

Inadequate evaluator training (Mass Partners, 2002; Massachusetts Department of Elementary and Secondary Education, 2011) was another problem identified prior to the implementation of the 2011 System. All principals involved in this study reported that they had been trained to implement the new System. However, the trainings they described were highly variable. Some principals described being trained by someone in their district who used training modules developed by the DESE. Other principals described training they received from outside consultants who had been brought into their districts to work with all evaluators. One principal described being part of the committee that oversaw implementation in her district. She felt as though she had intimate knowledge of the System from examining it in detail during the committee work. A principal who had worked as a teacher in a different district when the System was initially implemented described being trained by someone from the DESE. When he became principal in the same district a year later, he did not receive any training. He used the DESE training modules available online to train himself initially. Then he moved to a different school district as a principal and he did not receive any district-based training to support his understanding of how the System was being used there. Although principals did not raise concerns about training, the discrepancies in their experiences across just 11 school districts is worthy of note. The variability of training experiences raised several questions as to whether the new System has overcome this identified inadequacy in the former system. Is the training that is taking place adequate? Is it high quality? Do new administrators get trained? What is the consistency of training across districts?
A final complaint about teacher evaluation in Massachusetts prior to the 2011 System was a lack of a system for calibration (Massachusetts Department of Elementary and Secondary Education, 2011). Although the use of a rubric should help with calibration, it did not appear that principals perceived it as doing so. Interview data showed that lack of calibration existed in some districts, and that in other districts principals were not sure whether it existed, but were worried about it. Many principals cited lack of calibration between districts, especially when they reflected on their concerns about the publication of teacher performance levels.

**Current trends in teacher evaluation.** Teacher evaluation in the U.S. has come under a great deal of criticism for its ineffectiveness, both in measuring the competence of teachers as well as improving the quality of teaching. Peterson, in a review of the literature on teacher evaluation in 2000 concluded that evaluation practices had no impact on improving teacher practice, and did not accurately represent classroom practice. A seminal study, conducted by Weisberg et al., in 2009 concluded that evaluation systems fail to “provide credible and accurate information about individual teachers’ instructional performance” (p.4). These and many other studies contributed to general acceptance that educator evaluation in the United States was in need of reform. Standards-based teacher evaluation and value-added teacher evaluation systems are two prominent reforms that are being implemented across the country.

**Standards-based teacher evaluation.** The Massachusetts Educator Evaluation System is a standard-based evaluation system. The use of standards-based systems is increasing across the United States is an attempt to repair teacher evaluation practices nationally that were widely accepted as ineffective (Papay, 2012). Standards-based systems define teaching practice in research-based categories called standards, and they describe levels of competence related to each standard in a rubric (Papay, 2012). In Massachusetts, the rubric that principals use to guide
their decisions about the practice of teachers (see Appendix F) is organized into four standards: Curriculum, Planning and Assessment; Teaching all Students; Family and Community Engagement; and Professional Culture. Within each of the four broad standards there are several sub-standards called indicators that describe the components of teacher practice encompassed by the standard. Each indicator is further sub-divided into elements that describe with more specificity the practices that align to each indicator (Massachusetts Department of Elementary and Secondary Education, 2012a). For each of the thirty-three elements that comprise the teacher rubric, there are descriptions of teacher behaviors at each of the four performance levels: Unsatisfactory, Needs Improvement, Proficient and Exemplary. These descriptors are intended to be used to “as the basis for identifying the level of teaching … performance” (Massachusetts Department of Elementary and Secondary Education, 2012a, p. ii).

The literature review revealed concern about reliability of standards-based systems because they depend on human judgment, involve the use of limited data, and are carried out by different individuals who may have different interpretations of rubric descriptors (Papay, 2012). All of these concerns were raised by principals in this study as they relate to the use of ratings to judge teacher performance. Many spoke of reliability within their own district, wondering if other evaluators were making judgments consistent with theirs. One principal referenced knowing that there were calibration problems within her district and spoke of the commitment that her administrative team had made to work on the obvious inequities during the upcoming school year. She shared that when they reviewed their data district-wide, one principal had rated three percent of her teachers Exemplary, another had thirty percent Exemplary teachers and the remaining four schools were somewhere in between. She said
Some of us are more apt to give an Exemplary rating than others and I think we’re just not on the same page. I think for the teachers we all need some more calibration on that because I think that it depends on who your principal is as to what your rating could be. These were all cited as reasons that principals were uncomfortable with the use the rating system.

The Massachusetts Department of Education describes the rubric as a tool designed to help educators and evaluators (1) develop a consistent, shared understanding of what proficient practice looks like in practice, (2) develop a common terminology and structure to organize evidence, and (3) make informed professional judgments about formative and summative performance ratings on each standard and overall (Massachusetts Department of Elementary and Secondary Education, 2012a).

Despite their disdain for the performance ratings that are based on the rubric descriptors, principals’ spoke very positively about the impact of the rubric descriptors of teaching practice. The value that they see in the rubric aligns perfectly with the DESE’s intended use. One principal described using the rubric in a discussion with a teacher who disagreed with the rating that she had assigned, and felt that after urging the teacher to review the rubric descriptors with her that he understood why she had rated him as such. Another principal speaking about a common understanding among all educators that had developed as a result of the use of the rubric said, “When we’re going through the rubrics, we can all have the same knowledge and expectations of what we are going to see in classrooms, and what we’re truly looking for.” Still another stated that she felt the rubric helped her judgments to be more objective, saying “I like when I have a rubric in front of me and I’m marking from that. It really helps me define what I saw and what I want to see or why it’s great that I saw that.” Others referenced that teachers had
a better understanding of their own strengths and areas for improvement as they used the descriptors in the rubric to gather evidence to submit to their evaluator.

**Value-added evaluation systems.** Value-added evaluation systems are also gaining popularity in the United States in response to the call for reform of a failed educator evaluation system (Papay, 2012). These systems judge teacher performance based on student learning outcomes. Some states and systems refer to the student learning outcomes as value-added measures (VAM), others refer to student growth percentiles (SGP). Another measure used in some states is student learning objectives (SLO). VAMs are typically derived from student pre- and post-test scores on standardized state-wide assessment. Because most state-wide assessments do not apply to all content areas, states have had to come up with ways to measure student growth for all educators, and SGP and SLO have been developed to address this data deficit.

Value-added systems are seen by some as addressing the reliability concerns associated with standard-based systems due to the objectivity of test scores (Papay, 2012). However, recent research indicated that there is potential for bias in value-added measures due to assignment of students and student groups to teachers (Rothstein, 2010). There have also been questions raised about whether it is appropriate to judge the impact of a teacher based on the performance of a student or students who interact with many educators during a school day and school year (Tucker & Strong, 2005), as well as a concern about whether student growth that takes place after an assessment taken prior to the end of a school year is attributed to the teacher with whom the student is placed in the successive school year (Papay, 2012).

Among the many validity issues related to value-added systems in the literature are concerns about whether the assessment used actually measure the desired student learning
(Little, Goe & Bell, 2009). In addition, research has not shown a consistent and strong link between value-added scores and teaching practices (Little, Goe & Bell, 2009).

Because the RTTT grant that Massachusetts received required states to “have or develop policies that…use student learning and growth as a significant factor in educator evaluation” (Massachusetts Department of Elementary and Secondary Education, 2001, p.9), the Massachusetts Educator Evaluation System calls for the use of District Determined Measures (DDMs) of student learning. Each educator in the state must have two measures of student learning (DDMs) per year. After two years of data collection, a single educator will have four measures of student learning and the data from these measures will be used to determine the educator’s impact on student learning, which will be reported to the state as low, moderate or high. The Student Impact Rating will be reported to the DESE independently of the teacher’s performance rating, and will only have an impact on a teacher’s evaluation plan if it is determined that the teacher has both a low impact on student learning and an overall performance rating of Needs Improvement or Unsatisfactory. (Massachusetts Department of Elementary and Secondary Education, 2012c).

In Massachusetts, districts were allowed to choose the DDMs that will be used for teachers with the exception of teachers whose students in grades 4 through 8 take the state mandated MCAS assessments for mathematics and English language arts. Those teachers must use MCAS growth scores as one measure of their impact on student learning each year. For educators who teach academic subjects direct measures such as pre- and post- tests are recommended. For educators who do not interact with students in a classroom setting, or for those who support student learning in academic classes, such as special educators, indirect measures such as student attendance or development of study skills are allowed. Most districts
in the state are using teacher developed direct and indirect measures for the majority of teacher DDMs. Because the state of Massachusetts has allowed a great deal of local control over the DDMs used, it is likely that there is a high degree of variability among the assessments used from district to district in the state-wide Evaluation System. Further, since the determination of an educator’s impact on student learning (low, moderate or high) is made at the district level, and independently for each DDM that exists in the district, there is a likely to be very little calibration of impact ratings across the state.

Interestingly, only four principals in this study made mention of the district determined measures (DDM) component of the Massachusetts Educator Evaluation System. This is likely due to the fact that for most districts, this element of the System is currently in the first year of a two-year implementation plan. Those principals who did refer to DDMs had many reservations about their use. One principal raised concern about the validity of the data since teacher-made assessments were the basis for many DDMs across the state. She questioned whether assessments developed by teachers were valid measures of student learning. Another principal was worried about calibration and fairness, noting that from district to district, grade level to grade level, or content area to content area, assessments with varying degrees of rigor could have been chosen or developed. She said, “I think this process with the DDMs of people picking whatever they want to pick is a problem. It’s so willy-nilly…it’s opened up all this uncertainty and distrust.” She went on to explain that her district had very high expectations for student achievement on a particular commercially available reading assessment, and that a neighboring district was using the same assessment but had chosen less rigorous goals for their students. She felt it was unfair that a rating of moderate growth for her teachers would not equate to a rating of moderate growth for a teacher in the district next-door. Two principals spoke about their fear
that use of assessment data in evaluation will cause teachers to focus on achievement of the growth score, rather than the examination of student work to inform changes in curriculum and instruction. This concern seems valid because it is very much like what they have experienced already with teachers who focus on achievement of performance ratings rather than the feedback they receive from their principal to inform improvement of teaching practice.

Although there were few principals who referenced DDMs during their interviews, it is important to state their concerns here. Currently in Massachusetts, concerns about the use of DDMs as required by the Final Regulations on Evaluation of Educators, 603 CMR 35.00, by educators and administrators has resulted in legislative action that calls for elimination of the right of the DESE to mandate use of student performance data in teacher evaluation. It remains to be seen what the final outcome and future impact of this legislative action will have on the use of DDMs in the Educator Evaluation System. If DDMs continue to exist, given the concerns expressed by principals who did mention them and the outcry from educators across the state, it will be important to study this further to assess their actual effect.

The link of teacher effectiveness to student achievement. Empirical evidence is strong in support of the fact that student achievement is linked to teacher effectiveness. In a series of studies that examined student achievement gains on the Tennessee Value-Added Assessment System, Sanders, Rivers, Horn and Wright showed repeatedly that “the effectiveness of the teacher is the major determinant of student academic progress” (Sanders & Horn, 1998, p.247).

Studies conducted by Mendro (1998) showed that “over three or four years, student with ineffective teachers can perform at a level fifty percentile points lower than students with effective teachers” (p.262), and Nye, Konstantopoulos and Hedges (2004) showed difference in math and reading achievement gains between students with a low performing (25th percentile)
teacher and a higher performing (75th percentile) teacher. Goldhaber (2002) and Hanushek and Kain (2005) also showed student achievement differences to be related to teacher quality.

Although this research study did not address changes in student learning as a result of the Massachusetts Educator Evaluation System, principals’ responses implied their understanding that improvements in teaching would result in improvement of student learning. When principals described their interactions supporting the growth and development of teachers, many spoke of wanting their teachers to improve because they wanted the best for their students. As one principal described the way she reconciled feeling bad about her decision to terminate a teacher who worked hard to improve but ultimately did not, “We’re in this job [to do] what’s best for kids…I always make decisions based on what’s best for the little people, not big people. We have to make sure that’s right.”

**Effectiveness of teacher evaluation for improvement of teaching.** Research on the effectiveness of teacher evaluation is mixed. Studies that examine the relationship between teacher evaluation scores and student achievement scores have shown positive correlations (Milanowski et al., 2004; Pianta, 2005). However, there has been very little research that links teacher evaluation to improvements in teacher practice or student achievement. In a 2004 review of the literature on teacher evaluation, Peterson concluded that evaluation practices had no impact on improving teaching practice. More recently however, a study conducted by Taylor and Tyler (2011) showed that teachers were more effective in improving student achievement during their evaluation year than they had been previously, and that their effectiveness continued to increase in the year after their evaluation, suggesting that the evaluation process influenced teacher practice and student learning. In a recent study Steinberg and Sartain (2015) found greater gains in reading and math performance among students for teachers who were evaluated
with Danielson’s Framework for Teaching than for those who were evaluated with another system, again suggesting that a system of evaluation could influence student outcomes.

Evidence that evaluation prompts changes in educator practice has been shown in the research of Dee and Wycoff (2015) who found that teacher performance improved, particularly for low-performing teachers who remained in the system under the threat of dismissal as well as for high-performing teachers when the IMPACT teacher evaluation system was used. Although the existence of only three empirical studies is very limiting, it is hopeful, considering the paucity of research that precedes it.

The data collected in this study did not in any way relate the Massachusetts Educator Evaluation System to changes in student achievement. However, principals were asked about changes in teacher practice, as this is the precursor to improved student achievement. Principals all described teachers who had made changes to their practice as a result of feedback from their principal. Most felt that the verbal feedback and related discussions were the most impactful component of the System in this regard. When they saw evidence of changed practice, the use of the System was validated in the minds of principals.

There were some teachers, however, who were very resistant to feedback and principals perceived them as unwilling or unable to make the changes that principals recommended. Principals describing those teachers expressed frustration that their efforts to help a teacher had not had the desired impact. Most described that their relationship with teachers who were not receptive to feedback was harmed by the use of the Evaluation System. Many of these teachers had retired or left the school and principals felt that their decisions to leave were a direct result of the Evaluation System. Principals expressed relief that teachers had left their school on their own because it meant that they did not have to go through the process of termination. However,
some felt sorry that teachers who were very nice and worked very hard did not make changes that would have allowed them to continue on in their schools.

It is important to revisit principals’ perceptions relating to the rating system in the context of teacher improvement. Although principals felt strongly that the evidence-based feedback provided to teachers supported teacher growth and development, they did not think that the assignment of performance ratings was helpful at all. It seemed as though the assignment of performance ratings caused teachers to focus on achievement of a “grade” rather than how they could get better. If they achieved a proficiency level that was acceptable to them, they felt as though they had “passed.” If the proficiency level was one with which they disagreed, the conversation with the principal revolved primarily around why the principal had arrived at the judgment, rather than around the changes that the teacher should make to achieve their desired performance level. Principals hoped for schools full of teachers who would seek them out for feedback. One principal said that he wished his teachers would seek him out for feedback, but that he had come to realize in his two years in the principal role that “[Feedback] is uncomfortable for people.” This idea was foreign to him as a teacher and hard to accept in his role as principal.

For me as a teacher, I always wanted it…I was the one who wanted my principal to come in and tell me. Don’t tell me I’m doing great, because I know I’m going good, but I want you to tell me what I can improve. I guess that’s unique.

**Impact of collaboration on teacher practice and student learning.** The supervision component of evaluation, in which the focus is on improvement of teacher practice rather than judgment of teacher practice fosters strong collaborative relationships between the evaluator and teacher. Safe and respectful collegial relationships were identified as contributing to changes in
teacher thinking in Nolan, Hawkes and Francis’ review of six research studies of clinical supervision. Blasé and Blasé (1998) also found that “purposeful, appropriate and nonthreatening suggestions” provided to teachers during supervision were found to influence teachers’ classroom instruction (p.30). Recent studies of evaluation systems that incorporate the supervisory role as the Massachusetts Educator Evaluation System does, have shown positive impact on teacher performance (Steinberg & Saratian, 2015; Taylor & Tyler, 2011). The importance of collaborative conversations and relationships was paramount in discussions with principals interviewed for this study. Each and every principal discussed the powerful effect of collaborative discussion about teaching practice with their teachers. One said,

[Going into classrooms is] the best part of the job. Doing those ten-to-twenty minute observations…and having a conversation afterwards that’s on-the-spot … There is good feedback that can be had … having [teachers] think about their lesson reflectively … and [having] a good conversation with [their] building principal who’s a sounding board, listener, who can offer an opportunity … or just [give] them some good clear instruction. That to me is meaningful.

Many principals wished they had a great deal more time to engage in such conversations and would choose to let go of the components of the Evaluation System that they perceive to be less meaningful, if that were possible, in order to engage in more discussions about teaching with their faculty members.

**Conclusion and Significance of the Study**

Implementation of a new educator evaluation system represents a significant change in practice for both principals and teachers. In Massachusetts, this change took place alongside many other educational reforms implemented simultaneously. New standards for English language arts and mathematics based on the rigorous Common Core Standards were adopted the
same year as the Final Regulations for the Evaluation of Educators were enacted. Since then, and during the multi-year phased-in implementation of the Educator Evaluation System, many other changes have been required in Massachusetts schools. Among them are changes to certification for all teachers to better address the needs of students with learning disabilities and students for whom English is not their first language; changes to laws related to suspension of students from school, making it more likely that students who would have previously been suspended for violations of the code of conduct remain in classrooms; and new standards for science education. These changes and more, coupled with demographic shifts in the student population that make the work of teachers and principals more challenging and require new and more effective instructional strategies, have immersed Massachusetts Educators in a sea of change since 2011. The Educator Evaluation System that Massachusetts implemented is designed to help educators continually improve their practices. It seems as though the System would be perceived as beneficial to teachers and principals in addressing the increased and ever changing demands that they are facing.

The results of this study point to some benefits that the Massachusetts Educator Evaluation System brings to principals and teachers. The System is perceived to provide clear and high expectations for teaching practice that support both teachers and principals in developing common understandings about effective teaching. It also provides principals with opportunities to clearly understand the current state of teaching practice in each classroom and the in the school as a whole, given that they are in classrooms frequently and without appointments. The common understandings of effective teaching and evidence of current teaching practice collected from classroom observations fosters rich and productive discussions between teachers and principals that encourage teacher reflection and promote teacher
improvement. Principals report many instances of teachers who have embraced the feedback that has been provided to them through the Massachusetts Educator Evaluation System and who have made appropriate and lasting changes to their practices because of it.

There are many challenges related to the Massachusetts Educator Evaluation System that emerged in this study as well. The rating system was found to be a significant problem. Principals do not feel it is helpful due to the negativity of some of the names associated with performance levels and also because they feel the ratings are perceived by teachers and others as achievement levels that are unchanging. This perception does not align with the plan of principals to use the System as intended to promote continuous growth and development for all teachers.

Time constraints due to the overwhelming number of responsibilities that principals have do not enable principals to effectively implement all aspects of the Evaluation System. Principals tend to spend less time than they feel they should be crafting written feedback after observations and at the formative and summative stages of the evaluation process. This is in part because they do not have time, and in part because they feel that the written feedback is less effective than verbal feedback in helping teachers to grow. Principals when faced with limited time will likely choose to talk with a teacher over completion of a required written report.

Findings from this study may impact the actions of principals as they work to ensure that the supervision and evaluation practices that they use with the Massachusetts Educator Evaluation System are as effective as possible in supporting continuous teacher growth and development. Furthermore, the study findings may affect district-level discussions between administration and union leaders through the work of the district evaluation committees. Many districts established these standing committees to negotiate the terms of the Massachusetts
Educator Evaluation System for their districts and to plan and monitor implementation of new phases as required by the Massachusetts Department of Elementary and Secondary Education.

The Massachusetts DESE continues to study the implementation and effectiveness of the Educator Evaluation System. The findings related to the requirements for the names of performance ratings cannot result in changes at the school or district level since they are included in the regulations. However, the concerns about the rating names raised among the participants in this study are so significant that the DESE should consider whether the ratings are in fact having a negative impact state-wide, and if so, should make changes to them.

**Limitations**

It is not the intent of qualitative study to generalize beyond the study context (Creswell, 2009) and as such this qualitative study has several limitations.

The small sample size and purposeful participant selection in this research are factors that limit the generalizability of the findings. The participants were principals of public elementary schools in Massachusetts, all of whom are the sole evaluators of teachers in their buildings. The perceptions of the participants may not be representative of elementary principals who share evaluation responsibilities with an assistant principal or other professional, or of secondary level principals. Furthermore, this study was limited to participants who had used the Massachusetts Educator Evaluation System for at least two years. Those who have used it for less time than that may also have different perceptions. Finally, principals from other states, although they may be using an evaluation system similar to that used in Massachusetts may not perceive such systems in the same way that the study participants did.

Data collection in this study involved semi-structured interviews that asked principals to recall experiences that they had using the Massachusetts Educator Evaluation System, as such, the participant responses were based on memory. It cannot be assumed that the memories
expressed were accurate. Because qualitative research focuses on understanding the meaning that a person makes of a phenomenon (Merriam, 2009), and although the nature of the data collection process makes this a limitation of this study, it does not invalidate the findings (Maxwell, 2005).

Further, as described previously, the Massachusetts Educator Evaluation System is currently in the initial implementation phase, with full implementation required by October 2017 when DDM data will be submitted to the Massachusetts DESE for all educators. The perceptions expressed by participating principals may not persist as all educators become more experienced in using it.

**Validity and Reliability**

The researcher took care to attend to research-based guidelines related to validity, reliability and trustworthiness for this study. Creswell (2009) recommends use of multiple strategies to enhance validity, and identifies eight specific ways to go about enhancing the validity of study findings. Maxwell (2005) similarly outlines several strategies for reducing validity threats in qualitative studies. Several strategies outlined below were employed in this research.

Study participants were offered the opportunity to review the transcripts for accuracy (Mears, 2009), as well as preliminary analysis of data to ensure that their perspectives were not misinterpreted. This strategy, known as member checking or respondent validation, allowed the researcher to gain additional confirmation that her interpretations were accurate (Creswell, 2009, Maxwell, 2005) prior to publication of study findings.

The researcher reflected on her possible sources of bias prior to engaging in the research study and disclosed these in Chapter I in order to allow the reader to understand the perspective of the researcher and how this perspective may have influenced the study. In addition,
throughout the data collection and analysis process, the researcher bracketed her personal feelings in order to maintain objectivity so as to limit the influence of bias on the study findings (Maxwell, 2005; Mears, 2009).

The researcher was careful to generate interview transcripts that included the entirety of each participant interview (Maxwell, 2005). Furthermore, in reporting the study findings, as many descriptive perspectives related to each theme as possible were included. This presentation of rich, thick descriptions is considered an important strategy to increase validity (Creswell, 2009). This can also improve reliability as it provides readers with information that allows them to understand the research context in order to determine the transferability of the findings to their specific circumstance (Creswell, 2007; King & Horrocks, 2010; Merriam, 2009).

The study findings considered discrepant responses that did not align with the themes. Thorough examination of discrepant responses increases validity of findings by allowing the researcher to consider whether the discrepant responses reveal an error in analysis of the data that would cause the conclusions to change (Maxwell, 2005). Where discrepant data were found, they were presented in Chapter IV in the discussion of each theme in order that readers would be aware of this evidence and could base their conclusions in its existence (Maxwell, 2005).

Possible participant bias due to “power differentials” was addressed in this study by specifically excluding from participation principals who were employed in the school district where the researcher holds the position of assistant superintendent (Mears, 2009).

Future Research

This study presents an early examination of the perceptions of principals related to the Massachusetts Educator Evaluation System. The Massachusetts Educator Evaluation Regulations were enacted in June 2011. Since that time, public school districts in Massachusetts
have been in the process of implementing evaluation systems that comply with the regulations. These systems have been phased in over a multi-year period according to Massachusetts Department of Elementary and Secondary Education requirements. Full implementation will not be complete until October 2017, when the final district DDM data is submitted to the DESE. It will be important to continue to study the Massachusetts Educator Evaluation System as all public school educators become more familiar with it to continually monitor the impact that it has on educator practice and performance, and student achievement.

This study was limited to perceptions of elementary principals who are the sole evaluators of teachers in their schools, and it examined three very specific research questions related to the effect of the Massachusetts Educator Evaluation System on teacher growth and development, principals’ ability and opportunity to support teacher’s growth and development, and teacher-principal relationships. It would be interesting to examine these same research questions from the perspective of teachers as well as from the perspective of secondary level principals. Furthermore, because the Massachusetts Educator Evaluation Regulations also apply to the evaluation of principals and district-level leaders such as superintendents, assistant superintendents, special education directors, and business managers, future studies could investigate the perceptions of those who evaluate educators in these positions to consider commonalities and differences that may exist as they relate to specialized educator roles.

The ultimate goal of educator evaluation is the improvement of student achievement. Studies examining the link between changes in educator practice that took place as a result of the Massachusetts Educator Evaluation System and improvements in student achievement are recommended as well.
The variability in training among the 11 principals in this study is of concern given that lack of training was cited as a weakness in teacher evaluation systems in Massachusetts prior to the adoption of the Massachusetts Educator Evaluation System. Further studies should investigate the consistency of training for evaluators, as well as the quality, based on learning outcomes. Additionally, it would be interesting to see if districts are continuing to train new principals who join them, given that two principals in this study were not offered district training when they transitioned to new school districts as experienced principals.

**Recommendations**

The findings from this study point to some benefits and some areas of concern about the Massachusetts Educator Evaluation System as perceived by principals. The findings related to concerns may be beneficial to district, school and state leaders as they examine the System’s effectiveness.

School committees and district leaders, may want to consider increasing the number of evaluators in each school so that there is adequate time available for them to effectively implement the Massachusetts Educator Evaluation System. Aside from the performance ratings, principals did not feel that any aspects of the System were negatively impacting teacher growth. Rather, they expressed an inability to implement all aspects effectively and fully in the appropriate time frame. Although adding administrators is expensive, doing so might have the benefit of long term savings since there is a financial cost to having less than effective teachers in schools. Lower performing teachers produce lower performing students. When students do not perform as expected, schools must intervene to help them catch up. Typically, intervention for struggling students requires additional teaching staff to provide targeted small group instruction. If, with additional evaluators and an effective evaluation system, teachers in a
school could all become highly skilled, there would be far fewer low-performing students and the cost for additional teachers would be minimized.

The results presented herein may help district-level committees that oversee educator evaluation identify and consider modifications to the requirements for written feedback to make it more manageable for principals. They might also consider adding language that requires periodic meetings between the evaluator and each teacher to discuss the teacher’s performance and their progress toward achievement of their goals. Finally, with regard to the performance ratings, district-level evaluation committees may consider renaming the performance ratings that are used internally to names that suggest stages of growth and development.

As the Massachusetts Department of Elementary and Secondary Education reviews the Massachusetts Educator Evaluation System, it may reconsider the names used for performance ratings state-wide. The DESE may also wish to study the effect of making public the ratings of teacher performance each year. Based on this study, the ratings do not cause teacher performance to improve, and release of this information causes more harm than good.

**Personal Comments**

I embarked on this study to learn from the practitioners what effects the Massachusetts Educator Evaluation System was having on teachers and principals, and to understand some of the benefits and challenges that principals as evaluators perceived. My interest in this topic began many years ago as a classroom teacher who had but one instance in her 14 year teaching career of constructive and supportive evaluative feedback. My interest grew as I became an evaluator and found the challenges of helping teachers grow and develop through evaluation overwhelming. As a central office administrator charged with responsibility for curriculum, instruction and assessment in two different school districts, I grew further concerned as I worked with school-level administrators to support their work as evaluators, seeing and hearing many
principals speak about struggling with evaluation for many reasons. When the Massachusetts Educator Evaluation regulations were enacted I was hopeful that the strong research base upon which the System was built would provide Massachusetts educators with an outstanding system to help educators continually improve their professional practice, ultimately resulting in improvements in student achievement. When the opportunity to conduct a study presented itself, I was curious about what principals thought about this new System and how, in the early stages of implementation, it helped teachers and principals. I was also curious about whether the new System had been successful in addressing some of the challenges that I had experienced as a building-level evaluator, as well as challenges that principals with whom I have worked over the years have expressed. I was hopeful that the study results would provide me with information that I could use to help the principals in my district use the new System most effectively to help support teacher growth and development. Prior to conducting the study, I undertook an in-depth look at the Massachusetts Educator Evaluation System with a cohort of nine educators from my school district as soon as the model contract language (developed by the DESE) was released. We spent several months reviewing the regulations, the Model Contract Language, and the implementation guides to examine whether we felt it best to adopt, adapt or develop our own system. In the end, we adopted the Model System in our district, as did all of the districts represented in this study. For our committee, the process of scrutinizing the supporting documents that the DESE had provided, line-by-line, gave us confidence that the system we would put in place was not only well-intentioned to support teacher growth and development, but that it would in fact successfully do so. As each phase was implemented, I was directly responsible for communication of the changes to all evaluators and teachers in my school district. I participated, along with all evaluators and a few volunteer teachers, in two courses
taught by an outside consultant who was well versed in the System to ensure that evaluators knew how to use it effectively, and that there was consistency from school to school related to the implementation of the System, but also, and more importantly in my mind, to ensure that principals’ judgment of teaching practice was calibrated across schools. When we began to use the System in our schools with our teachers, the consultant continued to work with the team of evaluators. Periodically, he and I would meet with small groups of evaluators and we would together conduct an evaluation of a teacher who had volunteered to support our learning by opening in his or her classroom up to us for observation practice. During those sessions each of us acted as if we were evaluating on our own and after the observation and judgments were made by each of us independently, we discussed what we had observed and what our judgments of the teacher’s practice were. The lack of variation in the judgments between evaluators during these coaching sessions was a relief to me. I felt good that there was consistency among our evaluators and as a result, felt that our System was fair to teachers in this regard. In addition to the evaluator calibration coaching sessions, the evaluators in our district continued to meet five to six times per year with me and with the consultant to discuss teachers who were considered by their evaluator to be low performing. In most cases, the teachers under discussion at these meetings were on Directed Growth or Improvement Plans designed to provide them with increased support and increased oversight intended to accelerate their improvement. In these sessions, the identity of the teacher remained anonymous, but evaluators were given advice about how to best provide the teacher with the specific information needed to understand the areas in which improvement was necessary, as well as suggestions related to the steps he or she could take to realize the improvement. We also discussed how to provide the teacher with the supports that he or she needed, and at the same time reminded evaluators that
they needed to continue, more frequently, to evaluate and document in writing the teacher’s performance to ensure that at any point in the future if there were to be legal challenges to a case of dismissal for performance, the school district would have the necessary documentation to support that the teacher was informed and supported to improve in weak areas and that the legal obligations for evaluation were carried out well and responsibly. These meetings were particularly difficult for evaluators in my district because they struggled with finding a balance between supporting teachers in these situations and evaluating them.

At the same time that I was working with other evaluators in my district, I also had the responsibility for evaluating a few educators using this same System. I found myself struggling with how to make a fair rating decision, how open and honest I should be when I met with someone to discuss their practice and when I gave them written feedback, and whether or not all the time I was devoted to this work was making a difference. When these same struggles surfaced with the other evaluators in my district during the calibration and coaching sessions, I was compelled to understand whether other evaluators across the state were feeling the same way and if they were or had been, how they reconciled these challenges.

I had hoped to find that there were solutions to the challenges that I had seen and experienced in my district so that I could share those solutions with principals and I could employ them as I evaluated as well. What I discovered is that the challenges that we face are not unique, and that there is much work to do at the school, district and state level to ensure that educator evaluation consistently causes improvements in professional practices for all educators so that the students we serve have access to the high quality education that they deserve.
References


Marzano, R. J., & Toch, M. D. (2013). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development


Nye, B., Konstantopoulas, S., & Hedges, L. V. (2004). How large are teacher effects? 


Appendix A: Initial Participant Recruitment Email

Dear Principal: (Each principal’s name was inserted into the greeting line in order to personalize the invitation)

My name is Sheila Muir and I am a doctoral candidate in the College of Professional Studies at Northeastern University. I am also the assistant superintendent of school for the Quabbin Regional School District. As part of my dissertation research, I am conducting a study about the Massachusetts Educator Evaluation System.

In order to gather data about this research, I am inviting you to participate in my study. You have been asked to participate in this project because you are an elementary school principal in Massachusetts.

Due to the nature of my research I will need to limit participation to principals who are the only administrators who evaluate teachers in their school buildings. In addition, participants must have used the Massachusetts Educator Evaluation System for at least two years so that they have experienced the full evaluation cycle.

If you meet both of the criteria above, I would love to have you participate in my study. Participation will involve a 45 to 60-minute interview. This interview can be conducted in–person at a location and time of your choice, or over the telephone at a time that is convenient for you. During the interview I will ask you several questions about your experiences using the Massachusetts Educator Evaluation System. Your participation in the study will be confidential; names and other personal information will not be used. Your participation is entirely voluntary.

Please respond via e-mail to muir.s@husky.neu.edu if you are interested in volunteering or have any questions. Thank you in advance for your time.

Sincerely,

Sheila Muir
Appendix B: Signed Informed Consent Form for In-Person Interviews

Northeastern University, College of Professional Studies

Principal Investigator: Dr. Chris Unger  
Student Researcher: Sheila Muir

Title of Project: Principals’ Perceptions of the Massachusetts Educator Evaluation System

Why am I being asked to take part in this research study?  
You have been asked to participate in this research study because you are a principal of an elementary school who has used the Massachusetts Educator Evaluation System to evaluate teachers in your school. In addition, you are the only administrator in your school who evaluates teachers, and you have been using the Massachusetts Educator Evaluation system to evaluate teachers for at least two years.

Why is this research study being done?  
The purpose of the research is to explore the experiences of principals who are the sole evaluators of teachers in their schools as they implement the Massachusetts’ Educator Evaluation System. Principal perceptions as to the intended and actual effects of the Educator Evaluation System on teacher growth and development, their ability as leaders to support teacher development and the relationships between principals and teachers, may assist school and district leaders who seek to improve evaluation practices as well as state leaders who may consider revisions to the regulation that govern the Massachusetts Educator Evaluation System or other systems like it that may exist.

What will I be asked to do?  
The researcher will be looking for you to participate in an individual interview session, 45 to 60-minutes in length. You will choose to be interviewed in person at a location and time of your choice, or over the telephone at a time that is convenient for you. The interview will be audio taped.

Where will this take place and how much time will it take?  
Individual interviews will take approximately 45 to 60 minutes each.

Will there be any risk or discomfort to me?  
There are no significant risks involved in being a participant in this study.

Will I benefit by being in this research?  
There are no direct benefits for you. However, a potential benefit includes the ability to reflect on how you have implemented the Massachusetts Educator Evaluation System with teachers in your school. With your insight and feedback, your participation could potentially provide information that would assist principals in implementing the Massachusetts Educator Evaluation System. In addition, the outcomes of this study may inform parties who may be interested in improving the Massachusetts Educator Evaluation System.
Who will see the information about me?
Your part in the study will be completely confidential. Pseudonyms will be used for all study participants. Only the researcher will be aware of the participants' identities. No reports or publications will use information that can identify you in any way. In rare instances, authorized people may request to see research information about you and other people in this study. This is done only to be sure that the research is done properly. The researcher would only permit people who are authorized by organizations such as Northeastern University to see this information. No identifying information will ever be shared with anyone else.

If I do not want to take part in the study, what choices do I have?
You are not required to take part in this study. If you do not want to participate, you do not have to sign this form.

What will happen if I suffer any harm from this research?
There are no significant risks involved in being a participant in this study.

Can I stop my participation in this study?
Participation in this study is voluntary. You may discontinue your participation in this research program at any time without penalty or costs of any nature, character, or kind.

Who can I contact if I have questions or problems?
Sheila A. Muir, Student Investigator
Doctoral Student
P.O. Box 952
Barre, MA 01005
Cell: (978) 877-9681
Email: muir.s@husky.neu.edu

Chris Unger, Principal Investigator
College of Professional Studies
360 Huntington Avenue (BV 20)
Northeastern University, Boston, MA
Office: (617) 373-2400
Email: C.unger@neu.edu

Who can I contact about my rights as a participant?
Nan C. Regina, Director
Human Subject Research Protection
490 Renaissance Park
Northeastern University
Boston, MA 02115
(617) 373-4588
n.regina@neu.edu
You may call anonymously if you wish.

Will I be paid for my participation?
There is no compensation for participation in this study.

Will it cost me anything to participate?
There is no cost to participate in this study.
I have read, understood, and had the opportunity to ask questions regarding this consent form. I fully understand the nature and character of my involvement in this research program as a participant and the potential risks. I agree to participate in this study on a voluntary basis.

Research Participant Signature____________________________________________________

Research Participant Name___________________________________

Date______________
Appendix C: Unsigned Informed Consent Form for Telephone Interviews

Northeastern University, College of Professional Studies

Principal Investigator: Dr. Chris Unger
Student Investigator: Sheila Muir

Title of Project: Principals’ Perceptions of the Massachusetts Educator Evaluation System

Why am I being asked to take part in this research study?
You have been asked to participate in this research study because you are a principal of an elementary school who has used the Massachusetts Educator Evaluation System to evaluate teachers in your school. In addition, you are the only administrator in your school who evaluates teachers, and you have been using the Massachusetts Educator Evaluation system to evaluate teachers for at least two years.

Why is this research study being done?
The purpose of the research is to explore the experiences of principals who are the sole evaluators of teachers in their schools as they implement the Massachusetts’ Educator Evaluation System. Principal perceptions as to the intended and actual effects of the Educator Evaluation System on teacher growth and development, their ability as leaders to support teacher development and the relationships between principals and teachers, may assist school and district leaders who seek to improve evaluation practices as well as state leaders who may consider revisions to the regulation that govern the Massachusetts Educator Evaluation system or other systems like it that may exist.

What will I be asked to do?
The researcher will be looking for you to participate in an individual interview session, 45 to 60-minutes in length. You will be interviewed over the telephone at a time that is convenient for you. The interview will be will be audio taped.

Where will this take place and how much time will it take?
Individual interviews will take approximately 45 to 60 minutes each.

Will there be any risk or discomfort to me?
There are no significant risks involved in being a participant in this study.

Will I benefit by being in this research?
There are no direct benefits for you. However, a potential benefit includes the ability to reflect on how you have implemented the Massachusetts Educator Evaluation System with teachers in your school. With your insight and feedback, your participation could potentially provide information that would assist principals in implementing the Massachusetts Educator Evaluation System. In addition, the outcomes of this study may inform parties who may be interested in improving the Massachusetts Educator Evaluation System.
Who will see the information about me?
Your part in the study will be completely confidential. Pseudonyms will be used for all study participants. Only the researcher will be aware of the participants' identities. No reports or publications will use information that can identify you in any way.

In rare instances, authorized people may request to see research information about you and other people in this study. This is done only to be sure that the research is done properly. The researcher would only permit people who are authorized by organizations such as Northeastern University to see this information. No identifying information will ever be shared with anyone else.

If I do not want to take part in the study, what choices do I have?
You are not required to take part in this study. If you do not want to participate, you do not have to sign this form.

What will happen if I suffer any harm from this research?
There are no significant risks involved in being a participant in this study.

Can I stop my participation in this study?
Participation in this study is voluntary. You may discontinue your participation in this research program at any time without penalty or costs of any nature, character, or kind.

Who can I contact if I have questions or problems?
Sheila A. Muir, Student Investigator
Doctoral Student
P.O. Box 952
Barre, MA 01005
Cell: (978) 877-9681
Email: muir.s@husky.neu.edu

Chris Unger, Principal Investigator
College of Professional Studies
360 Huntington Avenue (BV 20)
Northeastern University, Boston, MA
Office: (617) 373-2400
Email: C.unger@neu.edu

Who can I contact about my rights as a participant?
Nan C. Regina, Director
Human Subject Research Protection
490 Renaissance Park
Northeastern University
Boston, MA 02115
(617) 373-4588
n.regina@neu.edu
You may call anonymously if you wish.

Will I be paid for my participation?
There is no compensation for participation in this study.

Will it cost me anything to participate?
There is no cost to participate in this study.
You may keep this consent form for yourself.

Thank you,
Sheila A. Muir
Appendix D: In-Person Interview Guide

Interviewee (Title and Name): Elementary Level Principals
Interviewer: Sheila A. Muir
Date: ________________________
Location of Interview: In-Person Interviews (site chosen by participant)

INTRODUCTORY PROTOCOL

Hello. I am Sheila Muir. I am the assistant superintendent in the Quabbin Regional School District, a regional district located in central Massachusetts. I appreciate your willingness to talk with me today and to participate in my study. You were invited to participate in this study because you are an elementary level principal who is the only evaluator of teachers in your school and because you have at least two years’ experience using the Massachusetts Educator Evaluation System.

To meet our human subjects’ requirements at the university, you have received the consent form that I sent to you. Essentially, that document stated that: (1) all information will be held confidential, (2) your participation is voluntary and you may stop at any time if you feel uncomfortable, (3) we do not intend to inflict any harm and (4) that I will be audio taping this interview. Do you have any questions about the interview process or the consent form at this time? If you consent to participating in this interview, please sign the consent form and I will begin the interview process.

I will begin today by explaining the purpose of my research. I will then ask you to respond to some questions about your experiences implementing the Massachusetts Educator Evaluation System with teachers in your school. I will be audiotaping this session because I want to be sure to capture your words accurately. I will also be taking written notes during the interview. I can assure you that everything you say here will remain confidential and anonymous. I will be assigning a pseudonym to your recorded responses prior to sending them to the transcriptionist and that pseudonym will be used when quoting from the transcripts. Tapes will be destroyed after they are transcribed. I will be the only one privy to your identity.

It is my goal to make you comfortable to share your experiences openly today. Please remember that I am interested in learning about your experiences and ideas. There are no incorrect answers.

INTENT OF THE STUDY

I am conducting a study on principal’s perceptions of the implementation of the Massachusetts Educator Evaluation System. My questions will focus on your experiences in evaluating teachers with the new system.

ESTABLISH RAPPORT
I have planned this interview to last no longer than 60 minutes. During this time, I have several questions that I would like to cover. If time begins to run short, I may need to interrupt you so that I can capture your responses to the full range of questions. I would like to begin by asking you a few questions about your experience as principal. I am asking these questions to allow you to become comfortable with the interview process and also so that I have information that may inform the conclusions from this research. After the initial questions, I will move into questions that pertain to your experiences using the Massachusetts Educator Evaluation System with teachers in your school.

INTERVIEW

Questions to Gain Background Information

1. How long have you been a principal at the elementary level?

2. Have you even been a principal at the secondary level? If so, when and for how long?

3. How long have you held your current position?

4. Have you evaluated teachers with an evaluation system other than the Massachusetts Educator Evaluation System, adopted in 2011? If so, what other evaluation system have you used? When and for how long did you use the other evaluation system?

5. Could you describe the how you use the Massachusetts Educator Evaluation System in your school?

6. When did you begin to use this new system?

7. How did you learn about it? Did your district provide training? Could you describe the training for me?

Questions About Educator Evaluation System Impact on Teacher Growth and Development

8. In theory, what do you think the goals of the Massachusetts Educator Evaluation system are?

9. One of the stated goals of the Massachusetts Educator Evaluation System is the improvement of teacher practice. What has been your experience with this aspect of the Evaluation System?

10. Could you give me some examples of your experiences related to teacher growth and development within the Massachusetts Educator Evaluation System? Please do not use any teacher names when you provide examples. You could make up a name for the teacher or you could just say “teacher x”.
11. What are the components of the Evaluation System that you think are most beneficial to your teachers in supporting their growth and development? Could you provide a specific example for me without identifying the teacher involved?

12. What are the components of the Evaluation System that you think are least beneficial to your teachers in supporting their growth and development? Could you provide a specific example for me without identifying the teacher involved?

13. If you were able to change certain components of the Massachusetts Educator Evaluation System in order to have it better help teachers to grow professionally what would you change? Why? How do you think that this change would help?

Questions About Principal’s Capacity to Influence Teacher Growth and Development

14. How has implementation of the Massachusetts Educator Evaluation System affected your ability and capacity as an instructional leader to influence teacher growth and development? Could you provide a specific example related to this?

15. What are some ways that the Massachusetts Educator Evaluation System has helped you as an instructional leader to support teacher growth? Could you provide an example of that?

16. What are some ways that the Massachusetts Educator Evaluation System has hindered your ability and capacity as an instructional leader to support teacher growth? Could you provide an example of that?

17. If you were able to change certain components of the Massachusetts Educator Evaluation System in order to have it better help you as an instructional leader to support the growth of your teachers what would you change? Why? How do you think that this change would help?

Questions About Principal Teacher Relationships

18. How has implementation of the Massachusetts Educator Evaluation System affected your relationship with the teachers who you evaluate? Could you give me some specific examples that illustrate changes to principal-teacher relationships resulting from the Massachusetts Educator Evaluation System?

19. What are some ways that you think your relationship with your teachers has improved as a result of the Massachusetts Educator Evaluation System? Why do you think that is so? Could you give me an example of that?

20. What are some ways that you think your relationship with your teachers has been negatively impacted as a result of the Massachusetts Educator Evaluation System? Why do you think that is so? Could you give me an example of that?
21. If you were able to change certain components of the Massachusetts Educator Evaluation System in order to have it better help principals and teachers to collaborate professionally what would you change? Why? How do you think that your change would help?

DEBRIEF
This concludes all of the questions that I have for you.
Thank you very much for your participation today. I will be in touch with you in the future to offer you a chance to review the transcript of your interview. In addition, I will be happy to send you a copy of the final results upon completion of the study.
Appendix E: Telephone Interview Guide

Interviewee (Title and Name): Elementary Level Principals
Interviewer: Sheila A. Muir
Date: ______________________
Location of Interview: Telephone Interviews

INTRODUCTORY PROTOCOL

Hello. I am Sheila Muir. I am the assistant superintendent in the Quabbin Regional School District, a regional district located in central Massachusetts. I appreciate your willingness to talk with me today and to participate in my study. You were invited to participate in this study because you are an elementary level principal who is the only evaluator of teachers in your school and because you have at least two years’ experience using the Massachusetts Educator Evaluation System.

To meet our human subjects’ requirements at the university, you have received the consent form that I sent to you. Essentially, that document stated that: (1) all information will be held confidential, (2) your participation is voluntary and you may stop at any time if you feel uncomfortable, (3) we do not intend to inflict any harm and (4) that I will be audio taping this interview. Do you have any questions about the interview process or the consent form at this time?

At this time, do you consent to participating in this research study? (If consent is given interview will proceed.)

I will begin today by explaining the purpose of my research. I will then ask you to respond to some questions about your experiences implementing the Massachusetts Educator Evaluation System with teachers in your school. I will be audiotaping this session because I want to be sure to capture your words accurately. I will also be taking written notes during the interview. I can assure you that everything you say here will remain confidential and anonymous. I will be assigning a pseudonym to your recorded responses prior to sending them to the transcriptionist and that pseudonym will be used when quoting from the transcripts. Tapes will be destroyed after they are transcribed. I will be the only one privy to your identity.

It is my goal to make you comfortable to share your experiences openly today. Please remember that I am interested in learning about your experiences and ideas. There are no incorrect answers.

INTENT OF THE STUDY

I am conducting a study on principal’s perceptions of the implementation of the Massachusetts Educator Evaluation System. My questions will focus on your experiences in evaluating teachers with the new system.
**ESTABLISH RAPPORT**

I have planned this interview to last no longer than 60 minutes. During this time, I have several questions that I would like to cover. If time begins to run short, I may need to interrupt you so that I can capture your responses to the full range of questions. I would like to begin by asking you a few questions about your experience as principal. I am asking these questions to allow you to become comfortable with the interview process and also so that I have information that may inform the conclusions from this research. After the initial questions, I will move into questions that pertain to your experiences using the Massachusetts Educator Evaluation System with teachers in your school.

**INTERVIEW**

**Questions to Gain Background Information**

1. How long have you been a principal at the elementary level?

2. Have you even been a principal at the secondary level? If so, when and for how long?

3. How long have you held your current position?

4. Have you evaluated teachers with an evaluation system other than the Massachusetts Educator Evaluation System, adopted in 2011? If so, what other evaluation system have you used? When and for how long did you use the other evaluation system?

5. Could you describe the how you use the Massachusetts Educator Evaluation System in your school?

6. When did you begin to use this new system?

7. How did you learn about it? Did your district provide training? Could you describe the training for me?

**Questions About Educator Evaluation System Impact on Teacher Growth and Development**

8. In theory, what do you think the goals of the Massachusetts Educator Evaluation system are?

9. One of the stated goals of the Massachusetts Educator Evaluation System is the improvement of teacher practice. What has been your experience with this aspect of the Evaluation System?

10. Could you give me some examples of your experiences related to teacher growth and development within the Massachusetts Educator Evaluation System? Please do not use any teacher names when you provide examples. You could make up a name for the teacher or you could just say “teacher x”.

11. What are the components of the Evaluation System that you think are most beneficial to your teachers in supporting their growth and development? Could you provide a specific example for me without identifying the teacher involved?

12. What are the components of the Evaluation System that you think are least beneficial to your teachers in supporting their growth and development? Could you provide a specific example for me without identifying the teacher involved?

13. If you were able to change certain components of the Massachusetts Educator Evaluation System in order to have it better help teachers to grow professionally what would you change? Why? How do you think that this change would help?

Questions About Principal’s Capacity to Influence Teacher Growth and Development

14. How has implementation of the Massachusetts Educator Evaluation System affected your ability and capacity as an instructional leader to influence teacher growth and development? Could you provide a specific example related to this?

15. What are some ways that the Massachusetts Educator Evaluation System has helped you as an instructional leader to support teacher growth? Could you provide an example of that?

16. What are some ways that the Massachusetts Educator Evaluation System has hindered your ability and capacity as an instructional leader to support teacher growth? Could you provide an example of that?

17. If you were able to change certain components of the Massachusetts Educator Evaluation System in order to have it better help you as an instructional leader to support the growth of your teachers what would you change? Why? How do you think that this change would help?

Questions About Principal Teacher Relationships

18. How has implementation of the Massachusetts Educator Evaluation System affected your relationship with the teachers who you evaluate? Could you give me some specific examples that illustrate changes to principal-teacher relationships resulting from the Massachusetts Educator Evaluation System?

19. What are some ways that you think your relationship with your teachers has improved as a result of the Massachusetts Educator Evaluation System? Why do you think that is so? Could you give me an example of that?

20. What are some ways that you think your relationship with your teachers has been negatively impacted as a result of the Massachusetts Educator Evaluation System? Why do you think that is so? Could you give me an example of that?
21. If you were able to change certain components of the Massachusetts Educator Evaluation System in order to have it better help principals and teachers to collaborate professionally what would you change? Why? How do you think that your change would help?

DEBRIEF
This concludes all of the questions that I have for you.
Thank you very much for your participation today. I will be in touch with you in the future to offer you a chance to review the transcript of your interview. In addition, I will be happy to send you a copy of the final results upon completion of the study.
Appendix F: Massachusetts Model System for Educator Evaluation—Teacher Rubric

Rubrics – defined in the regulations as “scoring tool[s] that describe characteristics of practice or artifacts at different levels of performance” (603 CMR 35.02) – are a critical component of the Massachusetts educator evaluation framework and are required for every educator. Rubrics are designed to help educators and evaluators (1) develop a consistent, shared understanding of what proficient performance looks like in practice, (2) develop a common terminology and structure to organize evidence, and (3) make informed professional judgments about formative and summative performance ratings on each Standard and overall. This appendix contains the ESE Model Teacher Rubric.

Structure of the Teacher Rubric

- **Standards**: Standards are the broad categories of knowledge, skills, and performance of effective practice detailed in the regulations. There are four Standards for teachers: *Curriculum, Planning, and Assessment; Teaching All Students; Family and Community Engagement; and Professional Culture*.

- **Indicators**: Indicators, also detailed in the regulations, describe specific knowledge, skills, and performance for each Standard. For example, there are three Indicators in Standard I of the teacher rubric: *Curriculum and Planning; Assessment; and Analysis*.
- **Elements:** The elements are more specific descriptions of actions and behaviors related to each Indicator. The elements further break down the Indicators into more specific aspects of educator practice and provide an opportunity for evaluators to offer detailed feedback that serves as a roadmap for improvement.

- **Descriptors:** Performance descriptors are observable and measurable statements of educator actions and behaviors aligned to each element and serve as the basis for identifying the level of teaching or administrative performance in one of four categories: *Unsatisfactory, Needs Improvement, Proficient, or Exemplary.*

**Use of the Teacher Rubric**

This rubric describes teaching practice. It is intended to be used throughout the 5 step evaluation cycle for all teachers, including teachers of whole classrooms, small groups, individual students, or any combination of the above. The rubric is designed to be applicable to general education teachers from pre-K through Advanced Placement, as well as teachers with specialized classes or knowledge, including teachers of English Language Learners, and special education teachers; districts may also choose to use this rubric for educators in other roles such as specialists.

The responsibilities of teachers to whom this rubric will be applied may vary. ESE encourages educators and evaluators to use the rubric strategically by discussing and agreeing upon certain Indicators and Elements that should be high priorities according to that educator’s role and responsibilities as well as his/her professional practice and student learning needs. There are a variety of ways to emphasize these components throughout the evaluation cycle. For example, high priority Indicators and/or elements can be analyzed in greater depth during self-assessment, targeted during goal setting, a focus for more comprehensive evidence collection, or all of the above. However, the expectation is that by the end of the evaluation cycle, educators and evaluators have gathered and shared a reasonable amount of evidence on every Indicator to support a rating for each Standard.
<table>
<thead>
<tr>
<th>Standard I: Curriculum, Planning, and Assessment</th>
<th>Standard II: Teaching All Students</th>
<th>Standard III: Family and Community Engagement</th>
<th>Standard IV: Professional Culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Curriculum and Planning Indicator</td>
<td>A. Instruction Indicator</td>
<td>A. Engagement Indicator</td>
<td>A. Reflection Indicator</td>
</tr>
<tr>
<td>2. Child and Adolescent Development</td>
<td>2. Student Engagement</td>
<td></td>
<td>2. Goal Setting</td>
</tr>
<tr>
<td>3. Rigorous Standards-Based Unit Design</td>
<td>3. Meeting Diverse Needs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Well-Structured Lessons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Assessment Indicator</td>
<td>B. Learning Environment Indicator</td>
<td>B. Collaboration Indicator</td>
<td>B. Professional Growth Indicator</td>
</tr>
<tr>
<td>2. Adjustments to Practice</td>
<td>2. Collaborative Learning Environment</td>
<td>2. Curriculum Support</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Student Motivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Analysis Indicator</td>
<td>C. Cultural Proficiency Indicator</td>
<td>C. Communication Indicator</td>
<td>C. Collaboration Indicator</td>
</tr>
<tr>
<td>3. Sharing Conclusions With Students</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Expectations Indicator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Clear Expectations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. High Expectations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Access to Knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Shared Responsibility Indicator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Professional Responsibilities Indicator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Judgment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Reliability and Responsibility</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How to reference parts of the rubric:

**Indicator terminology:** under the “Teaching All Students” Standard (II), the “Instruction Indicator” (A) can be referred to as *Indicator II-A*

**Element terminology:** under the Instruction Indicator (A), the Student Engagement Element (2) can be referred to as *Element II-A-2*

**Standard I: Curriculum, Planning, and Assessment.** The teacher promotes the learning and growth of all students by providing high-quality and coherent instruction, designing and administering authentic and meaningful student assessments, analyzing student performance and growth data, using this data to improve instruction, providing students with constructive feedback on an ongoing basis, and continuously refining learning objectives.

**Indicator I-A.** Curriculum and Planning: Knows the subject matter well, has a good grasp of child development and how students learn, and designs effective and rigorous standards-based units of instruction consisting of well-structured lessons with measurable outcomes.

<table>
<thead>
<tr>
<th>I-A. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-A-1. Subject Matter Knowledge</td>
<td>Demonstrates limited knowledge of the subject matter and/or its pedagogy; relies heavily on textbooks or resources for development of the factual content. Rarely engages students in learning experiences focused on complex knowledge or skills in the subject.</td>
<td>Demonstrates factual knowledge of subject matter and the pedagogy it requires by sometimes engaging students in learning experiences around complex knowledge and skills in the subject.</td>
<td>Demonstrates sound knowledge and understanding of the subject matter and the pedagogy it requires by consistently engaging students in learning experiences that enable them to acquire complex knowledge and skills in the subject.</td>
<td>Demonstrates expertise in subject matter and the pedagogy it requires by engaging all students in learning experiences that enable them to synthesize complex knowledge and skills in the subject. Is able to model this element.</td>
</tr>
<tr>
<td>I-A-2. Child and Adolescent Development</td>
<td>Demonstrates little or no knowledge of developmental levels of students this age or differences in how students learn. Typically develops one learning experience for all students that does not enable most students to meet the intended outcomes.</td>
<td>Demonstrates knowledge of developmental levels of students this age but does not identify developmental levels and ways of learning among the students in the class and/or develops learning experiences that enable some, but not all, students to move toward meeting intended outcomes.</td>
<td>Demonstrates knowledge of the developmental levels of students in the classroom and the different ways these students learn by providing differentiated learning experiences that enable all students to progress toward meeting intended outcomes.</td>
<td>Demonstrates expert knowledge of the developmental levels of the teacher’s own students and students in this grade or subject more generally and uses this knowledge to differentiate and expand learning experiences that enable all students to make significant progress toward meeting stated outcomes. Is able to model this element.</td>
</tr>
<tr>
<td>I-A. Elements</td>
<td>Unsatisfactory</td>
<td>Needs Improvement</td>
<td>Proficient</td>
<td>Exemplary</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>I-A-3. Rigorous Standards-Based Unit Design</td>
<td>Plans individual lessons rather than units of instruction, or designs units of instruction that are not aligned with state standards/local curricula, lack measurable outcomes, and/or include tasks that mostly rely on lower level thinking skills.</td>
<td>Designs units of instruction that address some knowledge and skills defined in state standards/local curricula, but some student outcomes are poorly defined and/or tasks rarely require higher-order thinking skills.</td>
<td>Designs units of instruction with measurable outcomes and challenging tasks requiring higher-order thinking skills that enable students to learn the knowledge and skills defined in state standards/local curricula.</td>
<td>Designs integrated units of instruction with measurable, accessible outcomes and challenging tasks requiring higher-order thinking skills that enable students to learn and apply the knowledge and skills defined in state standards/local curricula. Is able to model this element.</td>
</tr>
<tr>
<td>I-A-4. Well-Structured Lessons</td>
<td>Develops lessons with inappropriate student engagement strategies, pacing, sequence, activities, materials, resources, and/or grouping for the intended outcome or for the students in the class.</td>
<td>Develops lessons with only some elements of appropriate student engagement strategies, pacing, sequence, activities, materials, resources, and grouping.</td>
<td>Develops well-structured lessons with challenging, measurable objectives and appropriate student engagement strategies, pacing, sequence, activities, materials, resources, and grouping.</td>
<td>Develops well-structured and highly engaging lessons with challenging, measurable objectives and appropriate student engagement strategies, pacing, sequence, activities, materials, resources, technologies, and grouping to attend to every student’s needs. Is able to model this element.</td>
</tr>
</tbody>
</table>
Indicator I-B. Assessment: Uses a variety of informal and formal methods of assessments to measure student learning, growth, and understanding to develop differentiated and enhanced learning experiences and improve future instruction.

<table>
<thead>
<tr>
<th>I-B. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-B-1. Variety of Assessment Methods</td>
<td>Administers only the assessments required by the school and/or measures only point-in-time student achievement.</td>
<td>May administer some informal and/or formal assessments to measure student learning but rarely measures student progress toward achieving state/local standards.</td>
<td>Designs and administers a variety of informal and formal methods and assessments, including common interim assessments, to measure each student’s learning, growth, and progress toward achieving state/local standards.</td>
<td>Uses an integrated, comprehensive system of informal and formal assessments, including common interim assessments, to measure student learning, growth, and progress toward achieving state/local standards. Is able to model this element.</td>
</tr>
<tr>
<td>I-B-2. Adjustment to Practice</td>
<td>Makes few adjustments to practice based on formal and informal assessments.</td>
<td>May organize and analyze some assessment results but only occasionally adjusts practice or modifies future instruction based on the findings.</td>
<td>Organizes and analyzes results from a variety of assessments to determine progress toward intended outcomes and uses these findings to adjust practice and identify and/or implement appropriate differentiated interventions and enhancements for students.</td>
<td>Organizes and analyzes results from a comprehensive system of assessments to determine progress toward intended outcomes and frequently uses these findings to adjust practice and identify and/or implement appropriate differentiated interventions and enhancements for individuals and groups of students and appropriate modifications of lessons and units. Is able to model this element.</td>
</tr>
</tbody>
</table>
### Indicator I-C. Analysis: Analyzes data from assessments, draws conclusions, and shares them appropriately.

<table>
<thead>
<tr>
<th>I-C. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-C-1. Analysis and Conclusions</td>
<td>Does not draw conclusions from student data beyond completing minimal requirements such as grading for report cards.</td>
<td>Draws conclusions from a limited analysis of student data to inform student grading and promotion decisions.</td>
<td>Individually and with colleagues, draws appropriate conclusions from a thorough analysis of a wide range of assessment data to improve student learning.</td>
<td>Individually and with colleagues, draws appropriate, actionable conclusions from a thorough analysis of a wide range of assessment data that improve short- and long-term instructional decisions. Is able to model this element.</td>
</tr>
<tr>
<td>I-C-2. Sharing Conclusions With Colleagues</td>
<td>Rarely shares with colleagues conclusions about student progress and/or rarely seeks feedback.</td>
<td>Only occasionally shares with colleagues conclusions about student progress and/or only occasionally seeks feedback from them about practices that will support improved student learning.</td>
<td>Regularly shares with appropriate colleagues (e.g., general education, special education, and English learner staff) conclusions about student progress and seeks feedback from them about instructional or assessment practices that will support improved student learning.</td>
<td>Establishes and implements a schedule and plan for regularly sharing with all appropriate colleagues conclusions and insights about student progress. Seeks and applies feedback from them about practices that will support improved student learning. Is able to model this element.</td>
</tr>
<tr>
<td>I-C-3. Sharing Conclusions With Students</td>
<td>Provides little or no feedback on student performance except through grades or report of task completion, or provides inappropriate feedback that does not support students to improve their performance.</td>
<td>Provides some feedback about performance beyond grades but rarely shares strategies for students to improve their performance toward objectives.</td>
<td>Based on assessment results, provides descriptive feedback and engages students and families in constructive conversation that focuses on how students can improve their performance.</td>
<td>Establishes early, constructive feedback loops with students and families that create a dialogue about performance, progress, and improvement. Is able to model this element.</td>
</tr>
</tbody>
</table>
**Standard II: Teaching All Students.** The teacher promotes the learning and growth of all students through instructional practices that establish high expectations, create a safe and effective classroom environment, and demonstrate cultural proficiency.

**Indicator II-A. Instruction:** Uses instructional practices that reflect high expectations regarding content and quality of effort and work; engage all students; and are personalized to accommodate diverse learning styles, needs, interests, and levels of readiness.

<table>
<thead>
<tr>
<th>II-A. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-A-1. Quality of Effort and Work</td>
<td>Establishes no or low expectations around quality of work and effort and/or offers few supports for students to produce quality work or effort.</td>
<td>May states high expectations for quality and effort, but provides few exemplars and rubrics, limited guided practice, and/or few other supports to help students know what is expected of them; may establish inappropriately low expectations for quality and effort.</td>
<td>Consistently defines high expectations for the quality of student work and the perseverance and effort required to produce it; often provides exemplars, rubrics, and guided practice.</td>
<td>Consistently defines high expectations for quality work and effort and effectively supports students to set high expectations for each other to persevere and produce high-quality work. Is able to model this element.</td>
</tr>
<tr>
<td>II-A-2. Student Engagement</td>
<td>Uses instructional practices that leave most students uninvolved and/or passive participants.</td>
<td>Uses instructional practices that motivate and engage some students but leave others uninvolved and/or passive participants.</td>
<td>Consistently uses instructional practices that are likely to motivate and engage most students during the lesson.</td>
<td>Consistently uses instructional practices that typically motivate and engage most students both during the lesson and during independent work and home work. Is able to model this element.</td>
</tr>
<tr>
<td>II-A-3. Meeting Diverse Needs</td>
<td>Uses limited and/or inappropriate practices to accommodate differences.</td>
<td>May use some appropriate practices to accommodate differences, but fails to address an adequate range of differences.</td>
<td>Uses appropriate practices, including tiered instruction and scaffolds, to accommodate differences in learning styles, needs, interests, and levels of readiness, including those of students with disabilities and English learners.</td>
<td>Uses a varied repertoire of practices to create structured opportunities for each student to meet or exceed state standards/local curriculum and behavioral expectations. Is able to model this element.</td>
</tr>
</tbody>
</table>
### Indicator II-B. Learning Environment: Creates and maintains a safe and collaborative learning environment that motivates students to take academic risks, challenge themselves, and claim ownership of their learning.

<table>
<thead>
<tr>
<th>II-B. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-B-1. Safe Learning Environment</td>
<td>Maintains a physical environment that is unsafe or does not support student learning. Uses inappropriate or ineffective rituals, routines, and/or responses to reinforce positive behavior or respond to behaviors that interfere with students’ learning.</td>
<td>May create and maintain a safe physical environment but inconsistently maintains rituals, routines, and responses needed to prevent and/or stop behaviors that interfere with all students’ learning.</td>
<td>Uses rituals, routines, and appropriate responses that create and maintain a safe physical and intellectual environment where students take academic risks and most behaviors that interfere with learning are prevented.</td>
<td>Uses rituals, routines, and proactive responses that create and maintain a safe physical and intellectual environment where students take academic risks and play an active role—individually and collectively—in preventing behaviors that interfere with learning. Is able to model this element.</td>
</tr>
<tr>
<td>II-B-2. Collaborative Learning Environment</td>
<td>Makes little effort to teach interpersonal, group, and communication skills or facilitate student work in groups, or such attempts are ineffective.</td>
<td>Teaches some interpersonal, group, and communication skills and provides some opportunities for students to work in groups.</td>
<td>Develops students’ interpersonal, group, and communication skills and provides opportunities for students to learn in groups with diverse peers.</td>
<td>Teaches and reinforces interpersonal, group, and communication skills so that students seek out their peers as resources. Is able to model this practice.</td>
</tr>
<tr>
<td>II-B-3. Student Motivation</td>
<td>Directs all learning experiences, providing few, if any, opportunities for students to take academic risks or challenge themselves to learn.</td>
<td>Creates some learning experiences that guide students to identify needs, ask for support, and challenge themselves to take academic risks.</td>
<td>Consistently creates learning experiences that guide students to identify their strengths, interests, and needs; ask for support when appropriate; take academic risks; and challenge themselves to learn.</td>
<td>Consistently supports students to identify strengths, interests, and needs; ask for support; take risks; challenge themselves; set learning goals; and monitor their own progress. Models these skills for colleagues.</td>
</tr>
</tbody>
</table>
## Indicator II-C. Cultural Proficiency: Actively creates and maintains an environment in which students’ diverse backgrounds, identities, strengths, and challenges are respected.

<table>
<thead>
<tr>
<th>II-C. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-C-1. Respects Differences</td>
<td>Establishes an environment in which students demonstrate limited respect for individual differences.</td>
<td>Establishes an environment in which students generally demonstrate respect for individual differences</td>
<td>Consistently uses strategies and practices that are likely to enable students to demonstrate respect for and affirm their own and others’ differences related to background, identity, language, strengths, and challenges.</td>
<td>Establishes an environment in which students respect and affirm their own and others’ differences and are supported to share and explore differences and similarities related to background, identity, language, strengths, and challenges. Is able to model this practice.</td>
</tr>
<tr>
<td>II-C-2. Maintains Respectful Environment</td>
<td>Minimizes or ignores conflicts and/or responds in inappropriate ways.</td>
<td>Anticipates and responds appropriately to some conflicts or misunderstandings but ignores and/or minimizes others.</td>
<td>Anticipates and responds appropriately to conflicts or misunderstandings arising from differences in backgrounds, languages, and identities.</td>
<td>Anticipates and responds appropriately to conflicts or misunderstandings arising from differences in backgrounds, languages, and identities in ways that lead students to be able to do the same independently. Is able to model this practice.</td>
</tr>
</tbody>
</table>
## Indicator II-D. Expectations: Plans and implements lessons that set clear and high expectations and also make knowledge accessible for all students.

<table>
<thead>
<tr>
<th>II-D. Elements</th>
<th>Un satisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-D-1. Clear Expectations</td>
<td>Does not make specific academic and behavior expectations clear to students.</td>
<td>May announce and post classroom academic and behavior rules and consequences, but inconsistently or ineffectively enforces them.</td>
<td>Clearly communicates and consistently enforces specific standards for student work, effort, and behavior.</td>
<td>Clearly communicates and consistently enforces specific standards for student work, effort, and behavior so that most students are able to describe them and take ownership of meeting them. Is able to model this element.</td>
</tr>
<tr>
<td>II-D-2. High Expectations</td>
<td>Gives up on some students or communicates that some cannot master challenging material.</td>
<td>May tell students that the subject or assignment is challenging and that they need to work hard but does little to counteract student misconceptions about innate ability.</td>
<td>Effectively models and reinforces ways that students can master challenging material through effective effort, rather than having to depend on innate ability.</td>
<td>Effectively models and reinforces ways that students can consistently master challenging material through effective effort. Successfully challenges students’ misconceptions about innate ability. Is able to model this element.</td>
</tr>
<tr>
<td>II-D-3. Access to Knowledge</td>
<td>Rarely adapts instruction, materials, and assessments to make challenging material accessible to all students.</td>
<td>Occasionally adapts instruction, materials, and assessments to make challenging material accessible to all students.</td>
<td>Consistently adapts instruction, materials, and assessments to make challenging material accessible to all students, including English learners and students with disabilities.</td>
<td>Individually and with colleagues, consistently adapts instruction, materials, and assessments to make challenging material accessible to all students, including English learners and students with disabilities. Is able to model this element.</td>
</tr>
</tbody>
</table>
**Standard III: Family and Community Engagement.** The teacher promotes the learning and growth of all students through effective partnerships with families, caregivers, community members, and organizations.

<table>
<thead>
<tr>
<th>Indicator III-A.</th>
<th>Engagement: Welcomes and encourages every family to become active participants in the classroom and school community.</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-A. Elements</td>
<td></td>
</tr>
<tr>
<td>III-A-1. Parent/Family Engagement</td>
<td><strong>Unsatisfactory</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Needs Improvement</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Proficient</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Exemplary</strong></td>
</tr>
</tbody>
</table>

**Indicator III-B. Collaboration: Collaborates with families to create and implement strategies for supporting student learning and development both at home and at school.**

<table>
<thead>
<tr>
<th>II-B. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-B-1. Learning Expectations</td>
<td>Does not inform parents about learning or behavior expectations.</td>
<td>Sends home only a list of classroom rules and the learning outline or syllabus for the year.</td>
<td>Consistently provides parents with clear, user-friendly expectations for student learning and behavior.</td>
<td>Successfully conveys to most parents student learning and behavior expectations. Is able to model this element.</td>
</tr>
<tr>
<td>III-B-2. Curriculum Support</td>
<td>Rarely, if ever, communicates with parents on ways to support children at home or at school.</td>
<td>Sends home occasional suggestions on how parents can support children at home or at school.</td>
<td>Regularly updates parents on curriculum throughout the year and suggests strategies for supporting learning at school and home, including appropriate adaptation for students with disabilities or limited English proficiency.</td>
<td>Successfully prompts most families to use one or more of the strategies suggested for supporting learning at school and home and seeks out evidence of their impact. Is able to model this element.</td>
</tr>
</tbody>
</table>
### Indicator III-C. Communication: Engages in regular, two-way, and culturally proficient communication with families about student learning and performance.

<table>
<thead>
<tr>
<th>III-C. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-C-1. Two-Way Communication</td>
<td>Rarely communicates with families except through report cards; rarely solicits or responds promptly and carefully to communications from families.</td>
<td>Relies primarily on newsletters and other one-way media and usually responds promptly to communications from families.</td>
<td>Regularly uses two-way communication with families about student performance and learning and responds promptly and carefully to communications from families.</td>
<td>Regularly uses a two-way system that supports frequent, proactive, and personalized communication with families about student performance and learning. Is able to model this element.</td>
</tr>
<tr>
<td>III-C-2. Culturally Proficient Communication</td>
<td>Makes few attempts to respond to different family cultural norms and/or responds inappropriately or disrespectfully.</td>
<td>May communicate respectfully and make efforts to take into account different families’ home language, culture, and values, but does so inconsistently or does not demonstrate understanding and sensitivity to the differences.</td>
<td>Always communicates respectfully with families and demonstrates understanding of and sensitivity to different families’ home language, culture, and values.</td>
<td>Always communicates respectfully with families and demonstrates understanding and appreciation of different families’ home language, culture, and values. Is able to model this element.</td>
</tr>
</tbody>
</table>
**Standard IV: Professional Culture.** The teacher promotes the learning and growth of all students through ethical, culturally proficient, skilled, and collaborative practice.

Indicator IV-A. Reflection: Demonstrates the capacity to reflect on and improve the educator’s own practice, using informal means as well as meetings with teams and work groups to gather information, analyze data, examine issues, set meaningful goals, and develop new approaches in order to improve teaching and learning.

<table>
<thead>
<tr>
<th>IV-A. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-A-1. Reflective Practice</td>
<td>Demonstrates limited reflection on practice and/or use of insights gained to improve practice.</td>
<td>May reflect on the effectiveness of lessons/ units and interactions with students but not with colleagues and/or rarely uses insights to improve practice.</td>
<td>Regularly reflects on the effectiveness of lessons, units, and interactions with students, both individually and with colleagues, and uses insights gained to improve practice and student learning.</td>
<td>Regularly reflects on the effectiveness of lessons, units, and interactions with students, both individually and with colleagues; and uses and shares with colleagues, insights gained to improve practice and student learning. Is able to model this element.</td>
</tr>
<tr>
<td>IV-A-2. Goal Setting</td>
<td>Generally, participates passively in the goal-setting process and/or proposes goals that are vague or easy to reach.</td>
<td>Proposes goals that are sometimes vague or easy to achieve and/or bases goals on a limited self-assessment and analysis of student learning data.</td>
<td>Proposes challenging, measurable professional practice, team, and student learning goals that are based on thorough self-assessment and analysis of student learning data.</td>
<td>Individually and with colleagues builds capacity to propose and monitor challenging, measurable goals based on thorough self-assessment and analysis of student learning data. Is able to model this element.</td>
</tr>
</tbody>
</table>
**Indicator IV-B. Professional Growth: Actively pursues professional development and learning opportunities to improve quality of practice or build the expertise and experience to assume different instructional and leadership roles.**

<table>
<thead>
<tr>
<th>IV-B. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-B-1. Professional Learning and Growth</td>
<td>Participates in few, if any, professional development and learning opportunities to improve practice and/or applies little new learning to practice.</td>
<td>Participates only in required professional development activities and/or inconsistently or inappropriately applies new learning to improve practice.</td>
<td>Consistently seeks out and applies, when appropriate, ideas for improving practice from supervisors, colleagues, professional development activities, and other resources to gain expertise and/or assume different instruction and leadership responsibilities.</td>
<td>Consistently seeks out professional development and learning opportunities that improve practice and build expertise of self and other educators in instruction and leadership. Is able to model this element.</td>
</tr>
</tbody>
</table>

**Indicator IV-C. Collaboration: Collaborates effectively with colleagues on a wide range of tasks.**

<table>
<thead>
<tr>
<th>IV-C. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-C-1. Professional Collaboration</td>
<td>Rarely and/or ineffectively collaborates with colleagues; conversations often lack focus on improving student learning.</td>
<td>Does not consistently collaborate with colleagues in ways that support productive team effort.</td>
<td>Consistently and effectively collaborates with colleagues in such work as developing standards-based units, examining student work, analyzing student performance, and planning appropriate intervention.</td>
<td>Supports colleagues to collaborate in areas such as developing standards-based units, examining student work, analyzing student performance, and planning appropriate intervention. Is able to model this element.</td>
</tr>
</tbody>
</table>
**Indicator IV-D. Decision-Making:** Becomes involved in schoolwide decision making, and takes an active role in school improvement planning.

<table>
<thead>
<tr>
<th>IV-D. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-D-1. Decision-Making</td>
<td>Participates in planning and decision making at the school, department, and/or grade level only when asked and rarely contributes relevant ideas or expertise.</td>
<td>May participate in planning and decision making at the school, department, and/or grade level but rarely contributes relevant ideas or expertise.</td>
<td>Consistently contributes relevant ideas and expertise to planning and decision making at the school, department, and/or grade level.</td>
<td>In planning and decision-making at the school, department, and/or grade level, consistently contributes ideas and expertise that are critical to school improvement efforts. Is able to model this element.</td>
</tr>
</tbody>
</table>

**Indicator IV-E. Shared Responsibility:** Shares responsibility for the performance of all students within the school.

<table>
<thead>
<tr>
<th>IV-E. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-E-1. Shared Responsibility</td>
<td>Rarely reinforces schoolwide behavior and learning expectations for all students and/or makes a limited contribution to their learning by rarely sharing responsibility for meeting their needs.</td>
<td>Within and beyond the classroom, inconsistently reinforces schoolwide behavior and learning expectations for all students, and/or makes a limited contribution to their learning by inconsistently sharing responsibility for meeting their needs.</td>
<td>Within and beyond the classroom, consistently reinforces schoolwide behavior and learning expectations for all students, and contributes to their learning by sharing responsibility for meeting their needs.</td>
<td>Individually and with colleagues develops strategies and actions that contribute to the learning and productive behavior of all students at the school. Is able to model this element.</td>
</tr>
</tbody>
</table>
### Indicator IV-F. Professional Responsibilities: Is ethical and reliable, and meets routine responsibilities consistently.

<table>
<thead>
<tr>
<th>IV-F. Elements</th>
<th>Unsatisfactory</th>
<th>Needs Improvement</th>
<th>Proficient</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-F-1. Judgment</td>
<td>Demonstrates poor judgment and/or discloses confidential student information inappropriately.</td>
<td>Sometimes demonstrates questionable judgment and/or inadvertently shares confidential information.</td>
<td>Demonstrates sound judgment reflecting integrity, honesty, fairness, and trustworthiness and protects student confidentiality appropriately.</td>
<td>Demonstrates sound judgment and acts appropriately to protect student confidentiality, rights and safety. Is able to model this element.</td>
</tr>
<tr>
<td>IV-F-2. Reliability &amp; Responsibility</td>
<td>Frequently misses or is late to assignments, makes errors in records, and/or misses paperwork deadlines; frequently late or absent.</td>
<td>Occasionally misses or is late to assignments, completes work late, and/or makes errors in records.</td>
<td>Consistently fulfills professional responsibilities; is consistently punctual and reliable with paperwork, duties, and assignments; and is rarely late or absent from school.</td>
<td>Consistently fulfills all professional responsibilities to high standards. Is able to model this element.</td>
</tr>
</tbody>
</table>

**Note:** At the Exemplary level, an educator’s level of expertise is such that he or she is able to model this element through training, teaching, coaching, assisting, and/or demonstrating. In this rubric, this level of expertise is denoted by “Is able to model.”