TEACHER SUPERVISION AND EVALUATION: A CASE STUDY OF ADMINISTRATORS’ AND TEACHERS’ PERCEPTIONS OF MINI OBSERVATIONS

A dissertation presented by

Thomas F. Campbell

to
The School of Education

In partial fulfillment of the requirements for the degree of Doctor of Education

In the field of Education

College of Professional Studies
Northeastern University
Boston, Massachusetts
January 2013
Abstract

This case study will investigate teachers’ and administrators’ perceptions of the relationship between mini observations and teacher performance to understand what effect, if any, a system of mini observations has on teacher performance, and if mini observations influences a teacher’s pedagogical practice differently than a traditional evaluation model. Evaluating public school teachers has been studied for over one hundred years, yet little evidence exists as to the efficacy of mini observations on teacher performance. This study, including ten administrators and thirty seven teachers, will provide a small sample of data to help public schools evaluate and potentially implement this new method.

Four questions will guide this research: (1) What is the experience of teachers and administrators who have been exposed to mini observations? (2) How does the mini observation model impact teacher’s instructional practices as perceived by both teachers and administrators? (3) What are teachers’ perceptions about the reflective feedback and its ability to lead them to modify their practice? (4) Do teachers and administrators perceive any change in their relationship as a result of participating in mini observations? This qualitative study will gather data from six to ten participants from across one elementary school and one middle school through semi-structured interviews, observations of the teachers’ and administrators’ reflective conversations, and responses to a written survey. Findings, implications for practice, and limitations of the study will be discussed at length in the conclusion.

Keywords: Teacher supervision and evaluation, observation, mini observation, reflective conversation, teacher growth.
Acknowledgements

Writing a dissertation, by design, is an individual endeavor, yet there are many people who helped to make this project possible and need to be thanked. Most importantly, my family who are my life and strength. Without their support, I could not have made it through the process. To my amazing wife, Sue, who supported me every step of the way, thank you. Your patience and constant encouragement helped me to realize my dream. My son, Derek, who was a junior in high school when I began this adventure, never complained when Dad was again missing in action. I hope that my perseverance is an example of what happens when you set your sights on a goal and work really hard—anything is possible. My parents, Don and Ann Campbell, who have been my best teachers of all. They taught me about how important family is, and that caring about others and helping those who are less fortunate are the most important values in life. My dad, who worked two jobs to provide for his family, and my mom, who kept the entire family together, gave my four siblings and myself lessons in life that cannot be learned in school. Because of their guidance, I chose the path of becoming a special education teacher. Thank you for your many, many sacrifices to make our lives so rich and blessed. Finally, to my in-laws, Bill and Sylvia Stickney, who could not have been more supportive and caring. I’m lucky to have a second mother and father—it doesn’t happen very often.

Special thanks to my Superintendent, Dr. Dan Gutekanst, who quietly supported me every step of the way. Your friendship and kindness made all the difference in the world. To my colleagues in Needham, particularly those who graciously volunteered to be part of this study, I am deeply grateful for your time and support.

I’m grateful for the support of my advisor, Dr. Sara Ewell, who has served as a wonderful guide and mentor. She knew just when to push, and her encouragement helped me to reach the
finish line. Thanks also to Dr. Chris Unger and Dr. Ray Shurtleff who served on my committee. I appreciate their support and guidance.

Finally, I’d like to dedicate my doctorate to the memory of my very good friend, Dr. Joyce Reardon, who went to a better place on November 30, 2012. I sure wish you were here to see me accept my diploma, but somehow, I think you might be. The greatest instructions you gave me were to laugh hard and not be too serious. I love you, Joie.
# Table of Contents

Abstract ............................................................................................................................................2  
Acknowledgements ..........................................................................................................................3  
Table of Contents .............................................................................................................................5  
Chapter I: Introduction .....................................................................................................................8  
  Problem of Practice ..................................................................................................................8  
  Significance of the Problem .................................................................................................10  
  Practical and Intellectual Goals .......................................................................................14  
  Summary of Research Questions .......................................................................................15  
  Organization of the Document ...........................................................................................16  
  Theoretical Framework ......................................................................................................17  
Chapter II: Literature Review ........................................................................................................21  
  History of Teacher Evaluation ...........................................................................................22  
  1700–1900: Inspection .......................................................................................................22  
  1900–1945: Observations, Ratings and Professional Development ...................................23  
  1945–1980: Competency and Quality of Instruction .........................................................26  
  1980–1990: Accountability, Public Perception and Teacher Preparation .........................30  
  1990–Present: Focus on Standards and Value-Added Models ...........................................39  
  Current Models of Teacher Evaluation ..............................................................................43  
  Implications for Further Research .....................................................................................49  
Chapter III: Methodology ..............................................................................................................51  
  Research Questions .............................................................................................................51  
  Rationale for Qualitative Design .........................................................................................53
Chapter I: Introduction

Problem of Practice

Public schools in Needham, Massachusetts, like most public schools across the nation, rely on a system of teacher evaluation that was developed decades ago. This system consists of up to four formal classroom observations during the course of the school year, followed by extensive written summaries of each observation and a post-observation conference. The final summative evaluation is a detailed narrative document that evaluates the teacher based on nine performance standards: knowledge of the discipline, establishment of a learning environment, instructional strategies, assessment of student knowledge, promotion of equity and diversity, quality of relationships, managerial responsibilities, professional growth, and evidence of student learning.

Since the 1990s, teacher supervision and evaluation has shifted, in some school districts, to a standards-based approach with a focus on student learning. This shift has resulted in evaluating teachers based on statements of expectations that detail what teachers should know and be able to do. Many current supervision models, however, rely on a limited number of classroom observations and extensive written reports that typically require four to six hours to produce. Teachers and administrators say that the process has significant problems: the amount of time required for administrators to complete the process (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984); the limited number of classroom observations on which administrators base their perceptions and decisions about teacher performance (Marshall, 2009); the often rote-like quality of written feedback to teachers (Marshall, 2009); and a lack of evidence that the process improves professional practice or leads to improvement in student achievement (Wise et al., 1984).
In most districts, a principal implementing a traditional evaluation model formally observes approximately .1% of a teacher’s total teaching time. Many teachers are alone in their classrooms the remaining 99.9% of the time (Marshall, 2009, p. 22). Often teachers know when a principal is coming to observe and can present a lesson that may not represent their typical work; students will usually behave differently as well when an authority figure is in the room.

An alternative model of teacher supervision and evaluation that fits the standards-based approach focusing on student learning was developed by Kim Marshall (1996, 2005, 2006, 2009) and is based on increased classroom observations and reflective conversations with teachers. Through a series of unannounced “mini observations,” lasting no more than 10-15 minutes and conducted several times during the school year, the evaluator gathers a more in-depth understanding of how students are learning in a teacher’s classroom as well as the teacher’s strengths and skills that may require improvement. The final summative evaluation is a rubric-based model written in narrative form. What currently remains unknown is whether reflective conversations actually lead teachers to think differently about their methods of instruction.

There are no empirical studies to support Marshall’s claim other than his own anecdotal reports. The purpose of this study is to investigate teachers’ perceptions of the relationship between mini observations and teacher performance to determine what effect, if any, a system of mini observations has on teacher performance by influencing planning, pedagogy, student assessment, differentiation, and decisions to re-teach the lesson. The process developed by Marshall requires the principal to engage in four strategies: short, but frequent classroom observations followed by feedback conversations with the teacher; close involvement with curriculum planning; monitoring teams in their analysis of interim assessment data; and the use of a rubric-based summative evaluations at the end of the year. Marshall (2009) states, “there is
no pure exemplar and, to date, no definitive research on what happens when all four components are implemented together” (p. xviii). From 1993-2002, Marshall (2009) implemented mini observations while principal at the Mather School in Boston. Despite nine years of implementation, only anecdotal data about the efficacy of mini observations is available. By focusing on two schools with a combined total of 10 administrators and 37 teachers participating in mini observations, this study investigates the value of Marshall’s model and provides research results that will support or not support his claim by looking closely at the impact of mini observations of three teachers and three administrators.

I propose to address this problem of practice by conducting research that investigates teachers’ and administrators’ perceptions of mini observations and the impact that mini observations have on the teacher’s overall performance. Negotiated through the collective bargaining process, two schools in the selected district have implemented mini observations since September 2011. My research will focus on these two schools with ten administrators and thirty-seven teachers participating in mini observations.

**Significance of the Problem**

The purpose of teacher supervision and evaluation is varied but centers on several themes, including the diagnosing and improving of a teacher’s skills to improve student achievement; accountability for stakeholders; assisting administrators with decisions on continued employment; developing remediation plans for underperforming teachers; promotions and salary enhancement; providing professional development for teachers; helping students to select courses; decisions regarding school improvement; and collecting data for research on teaching (Doyle, 1983; Millman & Darling-Hammond, 1990; Harris, 1986; Johnson, 1988; Darling-Hammond, Wise, & Pease, 1983; Tucker & Stronge, 2005; Nolan & Hoover, 2010;
Danielson, 2007; Howard & Gullickson, 2010; Pullin, 2010; Moss, 2010). As stated by Howard and Gullickson (2010), teachers are critically important and, therefore, “it stands to reason that continuing investigation into the links among teaching practices, student performance, and the school reform efforts is a crucial need” (p. 340).

Teacher supervision and evaluation has spanned the ages from the early Christian era to the present day. The focus on teacher evaluation is cyclical, often emerging as a response to wartime conditions, political pressures, or international events, such as the launch of Sputnik in 1957 (Darling-Hammond, 1988; National Commission on Excellence in Education, 1983; Nolan & Hoover, 2010). Reform in teacher evaluation often follows national and global socioeconomic changes, and was documented as early as the 1800s (Doyle, 1983: Darling-Hammond, 1988). Historically, the most significant changes to teacher evaluation occurred at the turn of the century, in the 1960s and the 1980s (Darling-Hammond, 1988). In the late 1990s, a focus on standards-based education once again shifted the evaluation of teacher performance onto student achievement (Danielson & McGreal, 2000; DuFour, 2002; Marshall, 2005; Tucker & Stronge, 2005; Schmoker, 2006; Elmore, 2008).

Until the turn of the 20th century, formal evaluation of teachers was rare (Shinkfield & Stufflebeam, 1995). It was not until the 1970s that teacher evaluation became a political issue; it particularly gained significant attention after the release of A Nation At Risk (1983), which documented the shift of focus on education from equity and access to accountability.

Improving teacher supervision and evaluation methods is a national goal, and a priority in Massachusetts as well as most other states. President Obama’s education initiative, Race to the Top, requires significant changes to teacher evaluation including measures of student growth, the use of multiple rating categories, providing timely and constructive feedback, using the
evaluation system as a way to make tenure decisions, as well as a tool for removing ineffective teachers (Race to the Top Program Executive Summary, 2009). As a Race to the Top state, Massachusetts has proposed sweeping changes in this regard, such as instituting observations of teacher’s work that are shorter in duration but far more frequent throughout the school year (Massachusetts Department of Elementary and Secondary Education, 2012).

Research has shown that a sequence of highly effective teachers can result in a difference of 50 percentile points in student achievement after only three years, as compared to a student subjected to a series of inadequate teachers (Sanders & Rivers, 1996). Sykes and Winchell (2010) agree that a series of highly effective teachers from grade-to-grade can significantly improve student achievement. School principals and curriculum staff, as instructional leaders, are charged with ensuring the highest level of instructional practice to effect student achievement at high levels. Danielson (2009) states, “the single most important factor under the control of the school influencing the degree of student learning is the quality of teaching” (p.3). Howard and Gullickson (2010) state, “the quality of teacher performance in the classroom supersedes all other elements in its significant effect on student learning” (p. 338), and a student’s overall experience is linked to the teacher’s ability to “create an engaging learning environment” (p. 338).

Pullin (2010) examined data from the U.S. Department of Education concerning Massachusetts’ dismissal rates of teachers during the 2003-04 school year. Massachusetts then employed approximately 84,500 teachers in 371 school districts. At the conclusion of the school year, an average of 4.2 teachers were dismissed in each school district. Pullin (2010) points out that a vast majority of those teachers who were dismissed were within their first three years of employment. Pullin (2010) states the data is not surprising as the law pertaining to professional
status provides a clear advantage to the superintendent, who can dismiss teachers without cause within their first three years of employment (Massachusetts Education Reform Act, 1993).

One manner of assessing teachers’ performances and helping them improve their efficacy is to observe them frequently throughout the school year, and to provide meaningful and constructive feedback to promote greater self-reflection and help develop new strategies. For too many years, average and below-average teachers have been allowed to languish in classrooms throughout the United States, resulting in costly outcomes for children. Many principals do not have the tools to weed out ineffective teachers and two observations per year, as provided in most current teacher evaluation protocols, certainly does not provide enough evidence to make such important decisions.

The research on mini observation may help teachers move from being isolated, independent contractors to collaborative members of a professional learning community. Meeting frequently with principals and curriculum supervisors may help teachers strengthen their professional practice. Isolated professional development, such as attending workshops after school or outside the school district are disconnected and often times irrelevant. Job-embedded conversations about practice, meaningful discussions about student learning, and regular reflection about teaching and learning may lead to improved student achievement in a way that is deeper than infrequent, disconnected professional development workshops.

Finally, mini observations set the stage for meaningful engagement on the part of school principals. In many schools across the nation, principals are tied to endless administrative tasks such as building maintenance, upset parents, state testing requirements, local district initiatives, federal mandates, budget oversight and implementation and staffing issues—which only represent a few of the hundreds of priorities that principals must manage each day. Every
administrative task requiring a principal’s attention takes away from the single most important responsibility of instructional leadership: student learning. Mini observations require the principal and curriculum leaders to reprioritize and focus on classroom visits to observe teachers in action throughout the course of the school year. The protocol pushes instructional leaders to engage in conversations about how students learn best and the instructional practices that lead to higher student achievement. Lastly, this research will help evaluate if a system of mini observations improves the relationship between teachers and supervisors to a level where teachers actively engage in reflective conversations that are comfortable, collaborative, low stress and less likely to be viewed as less punitive.

Through this research, I hope to “bring people closer to an understanding of the experiences of others so that situations and programs and policies can be better understood and, if necessary, improved” (Mears, 2009).

**Practical and Intellectual Goals**

The practical goal of this research is to determine if the mini observation model is effective in Needham Public Schools and if the model should be implemented in the district’s six other schools, including the district’s largest school, Needham High School, with over 135 professional staff members. The purpose is to “explore the experiences of others so that lessons might be learned” (Mears, 2009) about the viability of using the model in other schools in the district. This case study will help to determine if mini observations lead teachers to reflect more deeply on their practice. In addition, the Department of Elementary and Secondary Education (DESE) is in the process of developing new evaluation protocols designed to meet the requirements of the federal Race To The Top grant recently awarded to Massachusetts. The DESE has developed a conceptual model that links teacher evaluation to student performance by
using standardized tests and other metrics of student achievement developed by each school district. While the mini observation model will not provide a systematic measurement linking student achievement with teacher performance, it will provide evidence of teachers’ and principals’ perceptions of the model with regard to student performance. This data could be instrumental to the Needham School district, as well as the DESE, to help develop one facet of the new state evaluation model.

The intellectual goals of this research are to gain a deeper understanding of using research to solve problems within the K-12 community. I hope to gain the necessary experience and skills to conduct empirical research to evaluate programs to determine their effectiveness within my school district. In addition to enhancing my skills as a researcher, I hope to become a more effective consumer of research literature when contemplating programs and services within my school district. In my opinion, educators spend too little time investigating the research literature before making decisions that are both important and costly to the school district. With limited public dollars, school districts must carefully consider the outcomes of decisions and invest wisely. Learning to use the vast body of research knowledge will be important as I further my career as an educator.

Research Questions

The following research questions are proposed for this study:

1. What is the experience of teachers and administrators who have been exposed to mini observations?

2. How does the mini observation model impact teacher’s instructional practices as perceived by both teachers and administrators?
3. What are teachers’ perceptions about the reflective feedback and its ability to lead them to modify their practice?

4. Do teachers and administrators perceive any change in their relationship as a result of participating in mini observations?

These questions are intended to examine the relationship between mini observations and teacher performance and to consider if a system of mini observations might have a positive impact on teacher performance. To date, there is little evidence to support Marshall’s claim that mini observations are a “powerful strategy for improving teaching and learning” (Marshall, 2009, p. 62). This qualitative research project will utilize a case study methodology with a series of interviews and questionnaires to address the research questions.

**Organization of this Document**

This proposal is organized into several key sections: the theoretical framework, literature review, research design, protection of human subjects, conclusion, bibliography, and appendices. Leadership and organizational theory comprise the theoretical framework used to inform the design of the research study and analysis of data. This theory systematically details how individuals and groups of people act and impact an organization. The literature review will be organized chronologically to provide the historical context and document the changes in supervision and evaluation, expectations, and practices resulting from political and social demands. A qualitative case study will comprise the following section, the research design. The study investigates teachers’ and administrators’ experience of participation in mini observations as a teacher evaluation tool and documents the themes that emerge in response to short but frequent classroom observations. This section will detail the research questions of the study, a description of the site and participants, data collection procedures and analysis, and issues of
validity and credibility. The next section will present how teachers and administrators, who volunteered to be part of the study, are protected and under the established criteria of the Institutional Review.

**Theoretical Framework**

The theoretical framework that will best inform investigation of the problem is leadership and organizational theory. This theory presents how individuals and groups of people act and impact the organization. The leadership and organizational theory providing the theoretical framework is based on the work of DuFour (2004); Elmore (2008); Bush (2009); Gallos (2008); Bolman and Deal (2008); Schmoker (2006); and Marzano, Waters, and McNulty (2005).

Schmoker (2006) discusses the many changes that individual schools implement without measuring effectiveness or improvement. This "logic of confidence," or constant implementation of new initiatives without evaluation, takes time away from "authentic leadership in instructional improvement" concludes Schmoker (p. 31). Schmoker (2006) identifies two things that matter most in schools: what is being taught and how well it is taught. Supporting Schmoker’s premise, DuFour (2002) presents the need for schools to view the principal as the learning leader and not just as an instructional leader, as principals must carefully evaluate student learning and not just teacher performance. In this new role, the principal can focus on a teacher’s instructional strategy as a way to improve student outcomes (p. 13). According to Marshall (2009), the mini observation model of supervision places a sharp focus on what students are learning to determine if teachers are effective.

Elmore (2008) states, "never have the demands for new skills and knowledge on the part of teachers and administrators been greater" (p. 89). He notes little difference in the organization and culture of schools over the last two centuries, and further describes teachers as independent
practitioners who do not have extensive opportunities to collaborate with other teachers (p. 92). Elmore (2008) believes the next chapter of performance-based accountability requires teachers to improve their instructional capacity to improve student achievement (p. 125). Marshall (2009) believes improved instructional capacity can be achieved through effective supervision and evaluation.

Bush (2003) discusses the features of collegial models, suggesting that "power and decision-making should be shared among some or all members of the organization" (p. 64). He states that collegial models are well suited for schools as teachers possess an "authority of expertise" (p. 65) and that "decisions are reached by consensus rather than division or conflict" (p. 67). Bush's (2003) collegial model fits well with Marshall's (2009) approach of reflective conversations after mini observations as the discussion is designed to share power and ideas about a specific aspect of the mini observation. Bush (2003) identifies three leadership models that are "particularly relevant for collegiality" (p. 76). These three models include transformational leadership, participative leadership, and interpersonal leadership. Transformational leadership focuses on school outcomes; participative leadership brings staff together; and interpersonal leadership "stresses the importance of collaboration and interpersonal relationships" (p. 79). The collegial model integrating transformational, participative, and interpersonal leadership provides the foundation for implementing a new observation model utilizing Marshall's approach to evaluating teachers.

Leithwood (1994) identifies four specific areas that leaders must focus on in schools: personal attention, helping others to solve problems in new ways, communicating high expectations, and model behavior. Marzano, Waters, and McNulty (2005) discuss the four dimensions of instructional leadership, concluding that an instructional leader is a "resource
Marzano et al. (2005) identifies 21 "responsibilities" of school leaders and how they correlate with student achievement. Monitoring and evaluating school practices has one of the highest correlations with academic achievement. He identifies specific behaviors and characteristics of school leaders to promote continual monitoring of the effectiveness of the school's curricula, instructional, and assessment practices, and to encourage continual awareness of the impact of the school's practices on student achievement.

Marzano et al. (2005) labels change in organizations as first- and second-order change. First-order change is slow and obvious while second-order change involves a "dramatic departure from the expected, both in defining a given problem and in finding a solution" (pp. 65-66). First-order change, as described by Marzano et al. (2005), places the responsibility of monitoring and evaluating as the most important factor in managing this type of change. In response to feedback from school principals about the limitations and complexity of the current evaluation model, the Needham Public Schools implemented first–order change by convening a committee to review models of teacher supervision and evaluation over two years. Consisting of seven members, including two principals, one assistant principal, three teachers, and the director of human resources, the committee represented administrators and teachers from elementary, middle, and high school levels. The first year included an in-depth study of the research on teacher evaluation. One principal, familiar with Marshall’s (2009) evaluation model through her experience teaching educational leadership at Simmons College, suggested the committee evaluate Marshall’s (2009) work, among others, in standards-based teacher evaluation (Danielson & McGreal, 2000; Tucker & Stronge, 2005). Because of the enthusiasm generated by Marshall’s (2009) work, in the second year the evaluation committee focused on
understanding the Marshall model through reading his articles and book. In addition, all members of the committee attended a full-day professional development workshop in 2010 where Marshall presented the mini observation model, and five of the seven members conducted a site visit to the Groton-Dunstable school district, which had implemented the model in all K-12 schools.

The final leadership and organizational theorists reviewed are Bolman and Deal (2008), and Gallos (2008). Bolman and Deal (2008) discuss leadership in the context of reframing or “deliberately viewing a situation from multiple perspectives” (p. 37), and offer the human resource frame as one perspective. “Human resources theorists typically advocate openness, mutuality, listening, coaching, participation and empowerment” (p. 41); these qualities are at the heart of mini observations. The supervision process shifts from the context of evaluations and documentation of performance to that of reflective conversation where the supervisor asks questions and listens, to help the teacher’s professional growth. Bolman and Deal (2008) articulate that “Human Resources Leaders Are Visible and Accessible” (p. 41) is one of three leadership principles. The mini observation model requires that principals observe teachers 12-15 times per year, and follow up each classroom visit with a personal conversation with the teacher.

Gallos (2008) notes the ability of a leader to make sense of everyday situations. She states, “Sense making involves three fundamental steps: noticing something, deciding what to make of it, and determining what to do about it” (p. 163). Marshall (2010) has done exactly what Gallos has suggested—notice that traditional models of evaluation are ineffective and developing an alternative method to measure effectiveness of student learning and enhance the professional growth of teachers.
Yin (2009) concludes:

An important step in all of these replication procedures is the development of a rich, theoretical framework. The framework needs to state the conditions under which a particular phenomenon is likely to be found…as well as the conditions when it is not likely to be found. (p. 54)

Chapter II: Literature Review

Teacher supervision and evaluation has evolved over the past 300 years with the most significant changes occurring in the last three decades. Despite the changes and increased focus on standards, supervision and evaluation continue to be mediocre at best and limited in helping teachers change their pedagogy for the benefit of their students (National Commission on Teaching and America’s Future, 1996; Howard and Gillickson, 2010; Darling-Hammond, 1998; Wise, 1884; Harris, 1986; Bridges, 1990; Tucker and Stronge, 2005). Teachers, principals, and other administrators are burdened by a process that relies on a limited number of classroom observations, demands a great deal of the administrator’s time to complete necessary paperwork, and lacks evidence demonstrating that the process actually improves professional practice (Darling-Hammond, Amrein-Beardsley, Haertel and Rothstein, 2012; Bouchamma, Godin, and Godin; 2008; Moss, 2010; Taylor and Tyler, 2011; Danielson and McGreal, 2000; Papay, 2012; Danielson, 2009; Darling-Hammond, 2010, Marshall, 2009).

This chapter presents a historical review of the relevant literature, divided by the state of teacher supervision and evaluation during the following five time periods:

1. 1700-1900: Inspection. In this era, administrators traveled to classrooms infrequently to observe teachers and ensure the curriculum was being followed.

2. 1900-1945: Observations, ratings, and professional development. These years saw the implementation of teacher rating scales in evaluation and the development of a clinical model to improve professional development. In addition, the importance of a
teacher’s personal qualities and characteristics received new focus as the understanding of teaching shifted from a scientific process to one that was more humanistic.

3. 1945-1980: Competency and quality of instruction. Focus moved toward teacher competency to increase effectiveness in the classroom, as well as the use of constructive appraisal and the diagnosis of strengths and weaknesses to help improve the quality of instruction.

4. 1980-1990: Accountability, public perception, and teacher preparation. For the first time in American education, the public became highly involved in teacher accountability amid the development of formative and summative assessment in teacher evaluation. Reports written for the public expressed a high demand for districts to develop instruments to evaluate teachers, set performance criteria, develop goals, and to change how teachers were prepared for the classroom. In this decade, the use of standards in teaching was articulated and recognized as an important aspect of advancing student achievement.

5. 1990-present: Focus on standards, value-added models, and professional conversations. This section reviews the literature on standards-based teacher evaluation and the shift in focus toward the relationship between student performance and teacher performance

**History of Teacher Evaluation**

**1700-1900: inspection.** Irregular inspection of teacher’s work through direct observation of a traveling supervisor was the focus of teacher supervision and evaluation in the early centuries in the United States and abroad (Shinkfield & Stufflebeam, 1995; Blumberg, 1985).
Shinkfield and Stufflebeam (1995) describe the writing of Charles Hoole, an English grammar school teacher who in 1659 published a pamphlet that highlighted characteristics of an effective teacher. Hoole wrote that it was the child’s responsibility to learn from the teacher, but the teacher could not be held responsible if the child failed to learn. In the late Victorian era, England provided a merit pay system to teachers called “payment by results” (p. 11). If the Queen’s inspectors believed that children were learning, based on their inspections, the teacher’s pay was supplemented. In 1902, the system was abolished due to public opinion that the education system was corrupted.

According to Blumberg (1985), teacher supervision models date back to the 1840s, where circuit superintendents traveled to schools to rate teachers and award teacher licenses. In the 1890s, supervision was conducted as an inspection to ensure that the curriculum was being followed and students were learning (Robinson, 2009).

Dewey (1900) advocated for change, particularly in the aftermath of the Industrial Revolution, where schools were required to examine the “science resulting in great inventions” (p. 9). Dewey (1900) stated that the curriculum needed to change to meet the needs of a changing, increasingly industrialized society. He advocated strongly that while schools should prepare students for vocational work, schools should also help children understand the world. He also stressed that traditional learning did not provide a social context, and that students should be working and learning together in a cooperative environment. Dewey (1900) stated, “Knowledge is no longer an immobile solid; it has been liquefied. It is actively moving in all the currents of society itself” (p. 25).

1900-1945: Observations, ratings, and professional development. In the early part of the 20th century, supervision and evaluation of teachers shifted from inspection to a focus on
research and professional development. For the first time, evaluation of teachers relied on newly developed teacher rating scales. The focus shifted from evaluating interpersonal characteristics of teachers to their professional development and interactions with students (Robinson, 1998; Millman & Darling-Hammond, 1990; Stufflebeam 1995).

According to Robinson (1998), supervision of teachers began after World War II and soon progressed to a clinical supervision model developed by Morris Cogan at Harvard University in the 1950s and 1960s. In the 1950s, supervision shifted from that of inspection to one of professional development where teachers were encouraged to enhance their professional growth. Since the 1980s, supervision has continued to promote personal reflection and changing one’s own behavior (Robinson, 1998).

Millman and Darling-Hammond (1990) examined the historical information on teacher effectiveness and the use of rating scales. Between 1896 and 1955, numerous studies documented the use of the student observations to identify the effective traits of teachers. Good and Mulryan (1990) state “Kratz’s (1896) study, ‘Characteristics of the Best Teachers as Recognized by Children’ was the pioneering work in this area” (p. 192).” In 1915, research began to look at the traits of effective teachers and continued through 1949. Several rating scales were developed, including the Elliot scorecard in 1915, the Boyce Card in 1915, the Rugg self-appraisal in 1920, and the Barr quality scales in 1931 (Millman & Darling-Hammond, 1990).

In the 1900s, non-written evaluations of teachers conducted by principals were based primarily on personal characteristics such as grooming, enthusiasm, and integrity. In the early part of the century, industry, typically ahead of education, tried to influence the educational process by “produc[ing] predictable and improved results” in schools (Shinkfield & Stufflebeam, 1995, p. 12). Darling-Hammond (1988) noted that at the turn of the century, progressives
pressed for schools of education to become more like schools of law, medicine, and applied science. Bobbit (1912), a prominent writer at the time, tried to bring industrial principles to the schoolhouse so that society’s needs would be met through the education of students. By 1925, it was reported that 75% of schools were using systems to rate teacher’s performance, an increase attributed to Bobbit’s (1912) influence and the influence of others in industry. In addition to personal characteristics, the evaluation rating system targeted areas of instructional technique, classroom management, and recordkeeping. According to Shinkfield and Stufflebeam (1995), the Hawthorne studies of the early 1930s prompted the shift in teacher evaluation from that of a “scientific approach” (p. 13) to a more humanistic approach. The change from production standards to interpersonal relationships paved the way for a focus on the teachers’ interactions with students as well as their own professional growth.

By 1920, the Committee on Child Development of the National Research Council developed methods of evaluation by observation. In 1945, the Ohio Teaching Record was developed; it was one of the first instruments where the supervisor and teacher worked cooperatively to focus on student learning (Good & Mulryan, 1990).

Robinson (2009) referred to the work of Stanford University professor Ellwood P. Cubberley, who stated in his forward to H. Nutt’s book, The Supervision of Instruction (1920), that supervision is only successful when the shift is made from inspection to helping the teacher to become more efficient in the classroom through constructive feedback. In 1925, students at Harvard produced the Confidential Guide to Courses, which contained an evaluation of instructors for students to use when selecting courses. In 1925, Washington University compiled student evaluations of their professors throughout the campus.
**1945-1980: Competency and quality of instruction.** The focus of teacher supervision and evaluation in the middle of the 20th century focused on teacher competency and the quality of instruction. For the first time, the literature revealed a shift from authority-based evaluation to peer evaluation and self-reflection. While most evaluation systems were based on a formative approach, they were viewed as largely ineffective, as administrators were not using the evaluation tool to help teachers grow or to move ineffective teachers out of the classroom (Shinkfield & Stufflebeam, 1995; Bridges, 1990; Darling-Hammond, 1990).

Bridges (1990) wrote the evaluation practices of the 1950s, 1960s, and 1970s were fundamentally weak and insufficient. Administrators were not well trained to conduct meaningful evaluations, and less stringent evaluation practices meant that less qualified teachers were granted tenure. Bridges (1990) described a system where mediocre teachers, once granted tenure, survived absent complaints from students and parents. Where complaints were lodged, teachers were moved to other buildings, assigned to teach elective courses, or worked as substitutes. When these “escape hatches” (p.149) were unavailable, only then would the principal confront teachers about their performance, intending that teachers would either improve, resign, or retire. Bridges noted that improvement was not often the outcome as identified teachers were fed years of positive feedback on their performance and principals had few tools or resources to promote improvement. When efforts to improve failed, negotiations with the union focused on removal and often involved incentives such as cash, extended medical benefits, or long-term consulting positions or employment as a substitute teacher. Bridges (1990) stated:

> A relatively small percentage of the teaching force—less than 1 a year—was forced to resign or retire early under this complaint-driven approach. As a result, nearly two million students a year were being shortchanged by marginal performance in the classroom, enough students to fill all the classrooms in the 14 smallest states. Moreover,
the students who were most likely to be cheated by the escape hatches in this approach were minorities and students from low SES [socioeconomic status] backgrounds. (Bridges, 1990, p. 150)

In 1958, Ryans studied the characteristics of 5000 elementary and secondary teachers to identify the quality and characteristics of teaching. He concluded that there were three patterns differentiating teachers, which he labeled as Pattern X, Y, and Z. Pattern X included being friendly, understanding, and sympathetic versus aloof, egocentric, and restricted. Pattern Y included those who were responsible, systematic, and businesslike versus unplanned, slipshod, and haphazard. Pattern Z included teachers being stimulating and creative versus rote and boring (Ryans, 1960).

From 1945-1975, Shinkfield and Stufflebeam (1995) described the abundance of literature produced on teacher evaluation, much of which was opinion and not research-based. Despite the lack of a scientific basis, the consensus was that a more constructive appraisal system to improve teacher performance was needed. The authors (1995) identified the themes that emerge in the literature, noting in particular that the work of David Ryans provided the foundation for the teacher competency era. Ryans (1960) stated, “if the teachers are misfits or are indifferent to their responsibilities, the whole program is likely to be ineffective and largely wasted” (p. 1). Ryans’ (1960) study looked carefully at identifying the “patterns of classroom behavior, attitudes, viewpoints, and intellectual and emotional qualities” (p. 9) to help quantify competent teachers. The study examined personal and social characteristics, and the patterns that identified incompetent teachers. From Ryans’ (1960) work investigating the patterns and characteristics of effective teachers, the Houston Needs Assessment was developed in 1973; two years earlier, in 1971, the Pennsylvania Competency Based Teacher Education program articulated the competencies that all teachers should possess.
In the 1960s through the 1970s, the evaluation protocol shifted to systematic accountability of teachers and the quality of classroom instruction. In the 1970s, the literature was mixed about evaluations. While teachers in principle agreed with the notion of evaluation, they did not believe the system was reliable or that the principal would always use the instrument for professional growth (Shinkfield & Stufflebeam, 1995).

Literature from the 1970s explored the shift from principals as primary evaluators to a system of self-assessment, peer assessment, and student assessment of teachers. Research found that when teachers took their evaluations positively, they were more likely to benefit from the process (Shinkfield & Stufflebeam, 1995).

Finally, Shinkfield and Stufflebeam (1995) found that evaluation systems in the 1970s and 1980s were formative in nature and did not provide the required, ongoing feedback that teachers needed to reflect on professional improvement. The authors concluded that up until the 1970s, teacher evaluation systems did not measure teacher effectiveness and that future efforts must address these issues as well as its relationship with student learning. Rosenshine and Furst (1971) explicitly stated there is little evidence to show the link between teacher behavior and improved student achievement, and, moreover, additional research was required to better understand the performance criteria measures to assess good teaching.

In the 1970s, a significant number of books were written on instructional evaluation (Doyle, 1983). Bloom, Hastings and Madaus (1971) differentiated evaluation into three different categories: diagnostic, formative, and summative. Educators who used diagnostic evaluation methods sought the underlying reasons for deficiencies; formative evaluation methods determined the level of the teacher’s mastery and identified the areas in which teachers were less
competent; and summative evaluation methods focused on global outcomes and general assessment of the product.

Stufflebeam (1971) reviewed the Context, Input, Process, Product (CIPP) model, developed through evaluating factors concerning content, input, process, and product. The content factors included outside events that affected classroom teaching such as the changes to curriculum, social forces, and teaching practices. Input factors included money, materials, equipment, textbooks, and expectations. Process factors were the interactions between students and teachers and included questions, directions, wait time, studying, and organization. Product factors included outcomes such as skills, habits, and understanding. This new approach to evaluating teachers represented a shift from an objective-based approach to looking at other factors influencing a teacher’s craft.

Harris (1986) stated that performance areas with specific descriptions and details should be elements of a teacher evaluation system that provided the opportunity for objective observations used to help diagnosis a teacher’s strengths and weaknesses. Harris (1986) identified several models to help develop performance areas, including a theoretical model based on B.F. Skinner’s stimulus-response theory of learning, which “call[ed] for teaching performance categories that stress student response modes, sequencing responses, reinforcement techniques and testing for behavioral outcomes” (p. 79). Another theoretical model discussed by Harris (1986) is the work of Steel and Stone (1973), who developed a model based on the work of Guthrie, Piaget, Maslow, and Skinner. The model specified “five ‘competency clusters’ with learning process ‘elements’ identified with each cluster” (p. 79). These competencies were designed to help assess the teacher’s ability to manage learning, relate to students, and help
students to understand, facilitate the student’s learning to new contexts, and provide feedback to the student.

Darling-Hammond (1990) observed that the public believed the quality of schools could only improve when the quality of teachers improved. Citing a 1979 Gallup poll on how public schools could improve, the results favored improving teacher quality.

Evaluation systems developed in the 1980s were based on models developed in the 1960s and 1970s. Those systems had a number of deficiencies, including an overreliance on a small number of observations, low level instructional strategies, little differentiation for novice and experienced teachers, and the principals’ lack of expertise in many disciplines (Danielson & McGreal, 2000). In addition, the clinical model developed in the 1960s relied on a one-to-one relationship between the supervisor and teacher to develop a collaborative approach to supervision (Hillyer, 2005).

1980-1990: Accountability, public perception, and teacher preparation. This 10-year period in education marked a significant change in how public perception of teachers in the classroom influenced policy. Led by highly publicized documents about the teaching profession, the general public was more involved than ever before in calling for accountability in education.

In 1983, a landmark report, *A Nation At Risk*, was produced for the American public. It called for a number of changes to the nation’s educational system because there was “widespread public perception that something is seriously remiss in our educational system” (p. 1), and that the United States’ educational system was at risk (United States National Commission on Excellence in Education, 1983). One recommendation for teacher preparation was “Persons preparing to teach should be required to meet high educational standards, to demonstrate an
aptitude for teaching, and to demonstrate competence in an academic discipline” (United States National Commission on Excellence in Education, 1983).

Shinkfield and Stufflebeam (1995) concluded that the publication of *A Nation at Risk* ignited the awareness of the American public that teacher performance and accountability in education were paramount. As a result, 98% of school districts put a teacher evaluation model in place, although predominately a summative model. Reform in education, with accountability as the primary focus, was clearly at the top of the agenda for national and local leaders (United States, National Commission on Excellence in Education, 1983; The National Commission on Teaching and America’s Future, 1996; Darling-Hammond, 1988; A Nation Prepared: Teachers for the 21st Century in response to A Nation At Risk, 1983). Helping teachers to grow through a system of formative and summative assessment, as well as providing support through mentoring programs, were new ideas for this decade (Harris, 1986; Darling-Hammond, 1988).

The National Commission on Excellence in Education called for using teams of experienced teachers to evaluate new instructors and extending probation periods from two to three years prior to granting tenure. The Commission also made bold recommendations with regard to teacher preparation and evaluation:

> Salary, promotion, tenure, and retention decisions should be tied to an effective evaluation system that includes peer review so that superior teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated. (The National Commission on Excellence in Education, 1983, p. 30)

Darling-Hammond (1988) stated that the reform in the 1980s had a new focus on teacher preparation and recognizing concerning trends that fewer high school students were choosing teaching as a career, and those who did were less capable than others entering other professional positions. Teachers judged to be academically capable were leaving the professional in greater numbers than those similarly situated in other industries. Darling-Hammond (1998) described
the initial change of educational reform as one that included legislation to assess teacher competency. In 1986, the next chapter of reform focused on teacher competence “through more rigorous preparation, certification and selection...” (p. 5).

In response to *A Nation At Risk*, Wise, Darling-Hammond, McLaughlin, and Bernstein (1984) documented the interest of school districts in creating systems of merit pay, career ladders, and master-teacher incentives in 1984. In order for the initiatives to succeed, the authors called for the improvement of teacher evaluation, which “must be capable of yielding fairly objective, standardized, and externally defensible information about teacher performance” (p. v).

Wise et al. (1984) studied the evaluation systems of 32-school district and found great variance among them. The researchers concluded that teacher evaluation was not well understood in terms of its purpose and goals. They observed that teachers were evaluated through three dimensions of teaching: labor, craft, and profession. Labor involved planning for teaching, while craft referred to the use of specialized skills in teaching. Profession was the ability to use judgment about the need for specialized skills. Wise et al. (1984) found that the two most frequently cited problems among the 32 schools were a sense that principals were not effective evaluators and many teachers were resistant to change and apathetic. Positive outcomes resulting from teacher evaluations were enhanced communication between the administrator and teacher, improvement of teachers’ instructional skills, the development of common goals and shared expectations, and a feeling of less isolation in the classroom.

Wise et al. (1984) concluded that a number of variables need to be present in an effective evaluation system and made several recommendations. First, the object of evaluation should match the district’s mission, values, goals, and operating assumptions. Second, districts should commit time and resources to teacher evaluation, making it a top priority. Further, principals and
other administrators conducting teacher evaluations should be trained. Finally, the authors recommended that master teachers assist with conducting evaluations, as they could offer content expertise that principals might lack.

In 1986, the Carnegie Forum on Education and the Economy produced *A Nation Prepared: Teachers for the 21st Century*, in response to *A Nation At Risk* (1983). Several key components of their framework to improve student outcomes included the creation of a National Board for Professional Teacher Standards to set standards and evaluate teachers. The authors of the report wrote:

> Virtually every occupation regarded by the public as a true profession has codified the knowledge, the specific expertise, required by its practitioners, and has required that those who wish to practice that profession with the sanction of its members demonstrate that they have a command of the needed knowledge and the ability to apply it. That is, the leading members of the profession decide what professionals in that area need to know and be able to do. (Carnegie Forum on Education and the Economy, 1986, p. 55)

The Carnegie Forum recommended the National Board also take responsibility for issuing two certificates: a teacher’s certificate and an advanced teacher’s certificate, both based on high standards for teaching and influenced by research on teaching and learning. Awarding these certificates would be based upon a candidate’s success in a three-stage assessment process: content knowledge, mastery, and observations of the candidate’s teaching. The Carnegie Forum recommended that salary advancement be available to those teachers holding National Board Certification. Furthermore, the Forum recommended using the standards produced by the National Board in colleges and universities in conjunction with teacher preparation programs.

Additionally, the Forum recommended that teachers earn a bachelor’s degree as a prerequisite to teaching, and that graduate schools develop a new master’s degree in teaching. The report suggested that an undergraduate degree alone did not provide the full preparation necessary to teach, and that states should consider a combined bachelor’s and master’s degree in
teaching. The report further suggested that the master’s program require two years, with the first year focused on course work and the second year devoted to teaching as a “resident” under the supervision of a lead teacher. The report criticized the way elementary teachers were being prepared, as most of their coursework was in pedagogy, not in the subject matter. Conversely, high school teachers did not have substantial coursework in pedagogy and too many courses focused on the subject matter. The result was that K-12 teachers were not properly prepared for the classroom. The report recommended that all teachers become experts in content and could teach using sound research practices combined with a solid understanding of child growth and development. Schools were also encouraged to develop lead teachers to focus on curriculum and instruction, and to help improve or recommend for dismissal their underperforming colleagues.

Harris (1986) suggested an eight-step approach to teacher evaluation: establishing criteria (what will be evaluated); instrumentation (how data will be quantified); data gathering (procedures); analysis (how data will be used); interpretation (understanding the findings); valuing (interpretation); decision-making (agreement and disagreements about the data); and action (plan for the future). He also argued that evaluation systems designed to judge, punish, reward, or control are ineffective and should be replaced by a system that offers a collaborative process for supporting teachers and helping them to improve.

According to Harris (1986), the practice of evaluating teachers across the nation was uniform and provided judgmental ratings with little data gathering, feedback for the teacher, or action. Stipnieks (1981), through her dissertation research on models of teacher evaluation in the United States, observed that some school districts in Georgia developed performance specifications and classroom observations that included pre- and post-conferences. During this
decade, teacher self-reflection or self-evaluation was a trend developing in some school districts in the country (Harris & Hill, 1982).

Harris (1986) argued strongly for the need to employ both formative and summative evaluations for teacher improvement. Many teacher evaluation systems were for summative purposes, he contended, yet districts’ stated goals were to help teachers understand their strengths and weaknesses (p. 211). Harris and Hill (1982) identified nine objectives for both formative and summative approaches: promotions, dismissals, merit awards, accountability to authorities, public information, reassignments, improving administration, improving program operations, and improving individual performance (p. 211).

Darling-Hammond (1988) advocated for mentor programs for new teachers, to reduce the high attrition rates. She reviewed those states requiring new teacher induction programs and found that 25 states required beginning teachers either to pass a performance assessment or they were provided with a support component in their first year. Some states awarded aspiring teachers a provisional certification upon completion of their preparation program, and then required a formal assessment during their first year as a full-time teacher. Teachers were evaluated by the principal two or three times during the year in several areas, often ignoring curriculum decisions.

According to Peterson and Peterson (1984), the evaluation of beginning teachers was uniquely problematic in that feedback from the principal was often different than feedback offered by the college or university supervisor. Other problems were that the principal was placed in a conflicting role of both supporter and judge, and often the same person conducted the formative and summative evaluations. The authors suggested that multiple sources of data be
used to give the new teacher appropriate feedback, including student and parent surveys, peer review of materials, student achievement data, teacher tests, and observations.

Evaluation instruments suggested that teaching competence was the same no matter the level or subject being taught; thus, the evaluator was to look at teachers’ behavior as evidence of competence or incompetence (Darling-Hammond, 1988). The outcome, according to Darling-Hammond (1988), was that assessment strategies for beginning teachers needed to differ from those for experienced teachers, who required more than minimum competencies to be evaluated.

Manatt (1982) advocated for an evaluation system that involved the teacher in the process, used performance criteria, involved goal setting, incorporated self and peer assessment, reviewed the results with the teacher, formulated goals, and included pre- and post-observation conferences. This clinical supervision model, developed by Harvard School of Education faculty in the 1960s, relied heavily on frequent observations and professional discussions with teachers, which was not well implemented in most local school districts. In his review of the literature, Manatt (1982) believed that holding a conference with a teacher after the observed lesson had the most impact on changing teacher behavior.

Stiggins and Duke (1988) conducted a research analysis of four school districts in Oregon and Washington to review the nature of teacher evaluation and collect data on perceptions about the evaluation system. They concluded that evaluators needed more training on how to evaluate teachers; more time for professional conversations about instructional improvement was necessary; teachers needed to get more involved in the process; and more sources of data were needed to include student and peer perspectives and to develop performance criteria relevant to student outcomes.
Bridges (1990) advocated designing evaluation systems that could account for the possibility of dismissal. The system should help the teacher to improve, increase the probability that a dismissal will be upheld if performance does not improve, and ensure that minority students and students from low socioeconomic backgrounds have equal access to a quality experience in the classroom.

In 1996, The National Commission on Teaching and America’s Future published What Matters Most: Teaching for America’s Future, reporting: “what teachers know and do is the most important influence on what students learn…research confirms that teacher knowledge of subject matter, student learning, and teaching methods are all important elements of teacher effectiveness” (p. 6). The report stated that a teacher’s thorough knowledge of content and strategies to motivate students were the most important factors for high student achievement. The report also criticized the number of teachers in the United States unqualified to teach due to the lack of teacher licensure and preparation. In 1990-91, only 72% of newly hired teachers were fully licensed to teach (p. 15). Several steps were recommended for improving student outcomes, including the development of standards for student learning and teaching.

The Commission further recommended a three-pronged approach to improving teacher preparation. First, all schools of education with teacher preparation programs were required to be accredited through the National Council for Accreditation of Teacher Education. In 1996, only 42% of all schools of education had successfully met the accreditation standards. Second, each state was to adopt performance standards for a teacher license that included testing for content. Third, states should adopt National Board for Professional Teaching Standards to develop policies concerning practice. The Commission also embraced the idea of an extended teacher preparation program whereby teachers completed a four-year undergraduate degree in a
specific content area, and a master’s degree in teaching. The Commission (1996) believed this practice, which had been adopted in a number of other countries, would provide the necessary preparation to become a more effective teacher.

A bold recommendation of the National Commission on Teaching and America’s Future (1996) was to remove incompetent teachers from the classroom by working closely with the union to establish peer support and review programs. The Commission (1996) cited better outcomes for removing teachers with a peer review model than traditional evaluation systems, and, accordingly, proposed a formal system utilizing peer assistance and review. The principal, as the primary evaluator, was responsible for many tasks including evaluation and could not devote the time necessary to provide the intensive support to a struggling teacher. A system of peer review, however, where pedagogical weaknesses could be addressed over time, was deemed to be more effective in improving instruction or recommending removal of the teacher. Yet most collective bargaining agreements are structured to protect veteran teachers. Stiggins and Duke (1988) wrote that collective bargaining agreements have historically interfered with the development of effective evaluation procedures. They stated, “Collective bargaining seemingly has done little to promote links between teacher evaluation and individual development” (p. 12).

The National Commission on Teaching and America’s Future (1996) cited work in Toledo, Cincinnati, and Columbus, Ohio; Rochester, New York; and Seattle, Washington. In each district, 7-10 teachers and administrators comprised a peer review and assistance panel. Teachers selected to be on the panel provided intensive assistance to teachers who were performing below expectations. The Commission (1996) stated that in each of the three states, “more teachers have been given help and have made major improvements in their teaching, and more teachers have been dismissed than ever had occurred under the old systems of
administrative review” (p. 99). In Toledo and Cincinnati, approximately one-third of teachers referred for intervention left teaching at the end of the school year. In Columbus, approximately 2% of teachers (144) were assigned for peer intervention over an eight-year period, and 28 teachers resigned or retired. In Cincinnati, 39% of teacher dismissals came from administrators, as compared to 61% initiated by the peer review and assistance panel. In Rochester, about 6% voluntarily left teaching out of the 60 teachers assigned intervention. “When teachers take on the task of professional accountability, it not only improves instruction but it profoundly changes the roles of teachers’ unions” (p. 99).

In the ten-year period from 1980-1990, teacher supervision and evaluation models experienced more change than any other time in history. The public demanded accountability and wanted to see change; policymakers similarly stressed reform through accountability. Holding teachers to high expectations was clearly articulated in the literature produced in this decade, as was the demand for an overhaul in teacher preparation. This ten-year period paved the way for the next two decades and eventual shift in focus toward standards.

1990-Present: Focus on standards and value-added models. Over the past two decades, teacher supervision and evaluation has taken a new direction, focusing on standards and value-added models. The past decade has seen a laser-like concentration on helping teachers to internalize what students should know and be able to do, and what exemplary teaching looks like. The standards-based movement has produced complex and complicated teacher evaluation models that can be cumbersome and exhaustive; yet the last two years has seen a shift from complex standards and measurement to engaging in professional conversations about teaching and learning (Tucker & Stronge, 2005; Danielson & McGreal, 2000; Marshall, 1996, 2005, 2006, 2009; Papay, 2012; Johnson, 2012; Peterson, 2011). The development of standards in the
teaching profession has been the single most effective step toward enhancing the work and learning in the classroom (Moss, 2010).

Taylor and Tyler’s (2011) research of the Cincinnati Teacher Evaluation System, based on the standards developed by Danielson (1996), showed improvement in students’ math scores when teachers were evaluated using standards. Furthermore, student achievement in math was sustained at higher levels over time when taught by a teacher who was subject to a standards-based evaluation. Darling-Hammond, Amrein-Beardsley, Haertel and Rothstein (2012) found standards-based evaluation had a positive impact on student learning gains.

In 2000, Danielson and McGreal paved the way for a new lens on teacher evaluation as “we are now interested in more complex learning in problem solving, in the application of knowledge to unfamiliar situations” (p. 3). Danielson and McGreal continue:

This is not an indictment of earlier models; indeed they represented the best of what was known at the time. Like other professions, however, education is built around conception of practice based on current and emerging research findings; as those findings suggest new approaches, pedagogical practice must also move forward. (p. 3)

The authors developed a set of teaching standards to support teacher growth and development, with a focus on helping teachers to reflect on their practice and engage in professional conversations that identified strengths and areas for improvement. Teachers would also engage in self-assessment activities as well as create a professional development plan for further growth. Danielson and McGreal (2000) strongly advocated for differentiating the evaluation process among new teachers, those with experience, and teachers who require assistance. With novice teachers, Danielson and McGreal (2000) recommended that one of several classroom observations should be conducted for an entire school day.

In 2005, Tucker and Stronge continued the standards-based approach to teacher evaluation and included an additional component of linking teacher performance to student
performance. They examined four models employed by school districts in Colorado; Virginia; Tennessee; and Oregon. Each school district had effectively incorporated some form of student performance in the evaluation of teachers. The outcome suggested that a focus on pedagogy student helped to improve student learning.

Kennedy (2010) examined standards-based evaluation of teachers and found five central themes: (a) the teacher’s resolve to students and the learning process; (b) content knowledge; formative assessment; (c) thinking analytically; and (d) making adjustments based on new learning and (e) participating in the learning with colleagues. Kennedy (2010) found that standards-based evaluations focused on the teacher and the ability to meet performance standards.

Bouchamma, Godin, and Godin (2008) built a model based on Danielson’s standards-based evaluation work in 1996. Citing concern about burdening supervising administrators due to the complexity of standards-based evaluation, four levels of performance measures (unsatisfactory, basic, proficient, and distinguished) were used to document four domains of teaching: (a) planning and preparation; (b) classroom environment; (c) instructional performance; and (d) professional responsibilities. Teachers were placed in five-year cycles of evaluation, where the first cycle reported on all domains and the remaining cycles focused on areas related to the teachers’ goals and professional development.

Continuing the work of Danielson’s standards-based work in 1996, Saginor (2008) developed a Diagnostic Classroom Observations protocol to help evaluators carefully look at the many facets of teaching so as to better diagnose teachers’ strengths and weaknesses. The process emphasized the depth of student learning over the content of teachers’ lessons. Understanding that inquiry and investigation were at the heart of good teaching, Saginor’s (2008) protocol
embraced a hands-on learning environment for students. Knowing what and how students think about a problem and the solution was essential for the teacher to understand. Saginor (2008) further argued that teachers needed to understand how students organized their knowledge, so as to assess their students’ ability to retain what they learned. The protocol required that students engage in dialog with each other; this social exchange is what promoted learning.

Despite the many advances in teacher supervision and evaluation, Papay (2012) argues that teacher evaluation in the United States is ineffective, and suggests a combination of standards-based practices aligned to a value-added model, where student achievement data is used to measure a teacher’s effectiveness (Papay, 2012). However, value-added models have come under fire as student gains can be attributed to many different factors, not just the influence of the classroom teacher. Further, it is difficult to isolate the many variables that contribute to higher achievement (Darling-Hammond, Amrein-Beardsley, Haertel & Rothstein, 2012).

Papay (2012) argues that an effective evaluation system must include a component that addresses ongoing teacher development and improvement, and that “assessment tools must be able to measure accurately and be unbiased, valid and reliable” (p. 127). Interrater reliability is an important aspect of any evaluation system and evaluators require extensive training to ensure consistency (Papay, 2012; Wei & Pecheone, 2010).

While neither system alone provides all of the answers, Papay (2012) believes that any effective evaluation system must focus on helping teachers to be more effective in their instructional capacity, including professional conversations about how teaching can improve student learning. Papay (2012) continues:

The evaluation system can and should be seen as a professional development tool and should be evaluated on its ability to raise instructional proficiency and student learning. Assessing the prospects of an evaluation system as a tool for continuous instructional
improvement requires not only examining its reliability, validity and bias but also identifying the system’s prospects for driving instructional change, (p. 133)

The Race to the Top initiative has prompted school districts to compete for federal dollars by using student performance data as part of each teacher’s yearly evaluation (Johnson, 2012). Performance-based pay systems have been tied to evaluation in a number of school districts in the country. According to Johnson (2012), the “theory of change” (p. 114) focuses on two assumptions: The notion that teachers will work harder if a bonus is offered, and less-effective teachers will leave when they are excluded from any pay increase. Using performance-based pay offers mixed results, particularly when the program focuses on the individual rather than rewarding groups of teachers. Knowing how to use student data in teacher evaluation continues to be a weakness for many evaluators (Peterson, 2011).

**Current Models of Teacher Evaluation**

**Professional conversations.** The focus on standards-based teacher evaluation in the last decade, and the more recent shift from the principal as the sole assessor to a shared model where professional conversations are at the heart of the learning and growth, has taken shape in the literature over the last several years (Holland, 2004; Danielson, 2009; Nolan & Hover, 2010; Marshall, 2009).

Danielson (2009) proposes that conversations about teaching and learning can lead teachers to reflect on their pedagogy and work with students. Administrators who fail to engage in this powerful method “decline to take advantage of one of the most powerful approaches at their disposal to promote teacher learning” (p. ix). These informal, professional conversations help teachers to think deeply about instructional strategies and to examine methods to improve student learning. Danielson (2009) argues that traditional supervision models require a teacher to be passive while administrators observe the teaching, spend days writing their observations,
Danielson (2009) continues:

Of all of the approaches available to educators to promote teacher learning, the most powerful...is that of professional conversation. Reflective conversations about practice require teachers to understand and analyze events in the classroom. In these conversations, teachers must consider the instructional decisions they have made and examine student learning in light of those decisions. There can be no doubt that conversation contributes to thinking. (p. 5)

Danielson (2009) reminds us that students learn best when they are given timely, specific feedback based on standards. The standards clearly outline what needs to be done to improve. She argues the same holds true for teachers—the feedback must be timely and specific, based on standards (p. 10). Danielson promotes the role of administrators as “coaches,” when trust has been established and teachers view them as a resource (p. 26).

Danielson (2009) argues that professional conversations must be built upon a shared understanding of essential ideas common among the staff in the school. These ideas set the vision and foundation for conversations; thus, school leaders must work with the faculty to focus on the most important and central ideas. Based on research and best practices, the shared culture is shaped by five guiding questions: (a) what constitutes important learning; (b) what causes learning; (c) how are students motivated; (d) what is intelligence and (e) how do students’ views influence their actions (pp. 27-28).

Danielson (2009) further provides several different topics for professional conversations after the principal or evaluating administrator has conducted a brief visit to a teacher’s classroom.

**Shared expertise.** Nolan and Hover (2010) similarly argue for collaborative, professional conversations between the teacher and principal. Older models, where the principal drives the process by providing all feedback and suggestions, assume the principal has the
content and pedagogy expertise. A teacher’s self-evaluation, however, is an important part of the growth process, as is a discussion between teacher and principal of the teacher observation.

Nolan and Hover (2010) argue shared expertise is necessary for understanding the teaching and learning process. They further contend that a high-performing model of teacher evaluation includes standards that are completely understood by both teacher and administrator (p. 19). These standards for evaluations are differentiated among novice teachers, experienced teachers, and veteran teachers who are underperforming (p. 19).

Nolan and Hover (2010) outline several principles that are required for an effective evaluation system for teachers: (a) focusing on the broad responsibilities of teachers; (b) collecting data from multiple sources using multiple methods to base judgments; (c) providing comprehensive training for evaluators based on best teaching practices; (d) community participation in designing the evaluation model and (e) ensuring compliance with local contracts. The authors conclude that teacher evaluation should be differentiated based on the performance of each teacher. Thus, teachers who perform at a high level should be evaluated differently than those teachers who underperform.

Professional development. The most recent research on teacher supervision and evaluation has focused on teacher professional development as the method to improve performance in a standards-based environment (Howard & Gullickson, 2010; Darling-Hammond, 2010; Papay, 2012).

Howard and Gullickson (2010) identify professional development as one avenue to improve teacher effectiveness, but caution that the teacher’s path of professional development must relate to the school’s or district’s goals, and not be left solely to the individual teacher’s discretion. Howard and Gillickson (2010) also argue that for too long, teacher evaluation has
been comprised of a series of behavioral outputs observed by the principal. They believe the evaluation system must measure the depth of the teachers’ knowledge base and pedagogy skills, as well as their ability to work with a diverse student population. The authors further contend that teacher evaluation has failed in many school districts for a variety of reasons, including the promotion of teaching styles that do not reflect research-based best practices, reliance on poorly-trained administrators, the failure to provide substantial follow-up, and the implementation of professional development goals inappropriately aligned to the school’s and district’s goals.

The work of high-performing nations. Darling-Hammond (2010) contends that no school district in the United States supports teachers in developing better teaching skills, in contrast to other high-achieving nations such as Finland, Korea, and Singapore. Darling-Hammond (2010) states the answer to improving student achievement, particularly in high-needs schools that are typically impoverished, is not in pay-for-performance models, but rather in a well-defined structure that enables practitioners to share effective practices with each other.

Darling-Hammond (2010) outlines the many problems with teachers who enter the profession, such as the varying level of preparation and skills, disparate salaries across the nation, placement of least skilled teachers in the poorest schools, higher class sizes in poorer communities, poor mentoring and induction programs, and the cost of undergraduate and graduate teacher preparation programs. She also states, “[Eighty percent] of the United States teacher’s time is spent with students, while teachers in high-achieving nations spend 60% of their time with students and 40% of the time planning, sharing or learning to improve their craft” (p. 201).

In contrast, Darling-Hammond (2010) notes, other high-achieving nations invest significantly in teacher preparation, training, and support and include the following:
• Universal, high quality teacher education at the government’s expense. Teacher candidates in Finland, Sweden, Norway, and the Netherlands receive two to three years of graduate-level preparation at the government’s expense, with a stipend for living expenses (p. 199).

• Mentoring for beginning teachers with fewer class sections to teach.

• Ongoing professional development with 15 to 25 hours of planning and collaboration with colleagues, and two to four weeks a year of institutes and seminars.

• Leadership opportunities for expert teachers.

• Equitable and competitive salaries (p. 198).

• Provision of 100 hours per year of professional development (p. 201).

Darling-Hammond (2010) recommends a standards-based teacher evaluation system that requires evidence showing that a student’s performance improves as a teacher becomes more effective. A standards-based teacher evaluation model can also help with decisions about who remains in the profession, and which teachers should become mentors and coaches.

“Professional standards hold members of a profession accountable for developing shared experience and applying it appropriately” (p. 219).

**Mini observations.** Observing teachers while they practice their craft has been a staple of supervision and evaluation since the 1700s. The last decade has focused observing teachers related to performance standards. Traditionally, classroom observations encompassed a full period, watching the teacher conduct a lesson from start to finish. Recent literature has articulated the value of shorter, more frequent observations of teachers (Milanowski, 2011; Marshall, 1996, 2005, 2006, 2009; Holland, 2004).
It is important to distinguish the difference between mini observations and walk-through’s, also referred to as learning walks and instructional rounds, developed in the 1990s. A walk-through typically involves an individual or group of observers who tour classrooms or the entire school with focus on a specific aspect of instruction that is followed by a debriefing session with the building principal (City, Elmore, Fiarman, & Teitel, 2009; Odden, 2011). Downey, English, Frase, and Poston (2004) state walk-through’s are not intended to evaluate teachers but rather gather data on how teachers make decisions about instructional practices. In contrast, mini observations are time intensive, involving more frequent visits to the classroom and providing individual feedback to the teacher that focuses on evaluation and instructional improvement (Marshall, 2009).

Milanowski (2011) advocates for the use of walk-through as a method to evaluate teachers. These frequent observations of a teacher’s practice provide a better picture of the typical classroom experience. Milanowski (2011) advocates that school leaders, instructional coaches, and peers should engage in walk-through’s as a method for measuring instructional practice and competency. Peterson (2004) cautions that walk-through’s are limited, and thus best used as a means for gathering data.

To measure productivity, Milanowski (2011) advocates for the use of a value-added approach to teacher evaluation in combination with teacher assessment methods. The value-added approach measures, through student assessment data, how effective a teacher is with a particular group of students, as compared to the expected growth of an average group of students with similar characteristics.

with evaluation models in several public and charter schools, Marshall (2009) developed his own rubric that consisted of domains, criteria, and a rating scale. The principal would observe teacher briefly in the classroom, and in team meetings and other school events. The mini observations in the classroom lasted 5-15 minutes, which were followed up within 24 hours by face-to-face conversations. Teachers conducted a self-assessment under the same rubric used by the principal for summative evaluations. Teachers who failed to meet standards were informed well before the end of the year. The principal assumed responsibility for prescribing remedial steps.

In 2009, Marshall refined the teacher evaluation rubrics after conducting further research of five public and charter schools that had developed rubrics for supervision and evaluation. The new rubrics provided a clear path for teachers to understand areas needing improvement in order to reach proficiency and expert levels.

Despite the decades of research devoted to teacher evaluation, however, no one clear method has emerged linking teacher performance to higher levels of student learning. Teachers are assessed throughout their careers, beginning as early as college, and extending through student teaching, the licensure process, the job interview, and in the classroom. Yet none of these assessments have successfully predicted who will have the greatest impact about student’s ability to learn (Kennedy, 2010). The most recent literature suggests that frequent visits to classrooms with clear feedback for improvement may result in higher achievement (Marshall, 2009; Danielson, 2009; Milanowski, 2011; Papay, 2012; Johnson, 2012; Taylor & Tyler, 2011).

**Implications for Further Research.**

An abundant amount of literature exists on teacher supervision and evaluation, particularly since the 1980s. In the last two decades, the literature has focused primarily on standards-based teacher evaluation. This approach to supervision and evaluation has provided a
framework for teachers to understand better what a high performing teacher looks like and how they can become one. Through the use of rubrics, standards, and indicators, teachers have—for the first time in the history of teaching—a clear indication of how to improve their instructional practice, through benchmarks in all domains of teaching that clearly determine whether their teaching is unsatisfactory, improving, proficient, or exemplary.

Within the last decade, the standards-based movement has been supplemented with literature on the use of professional development, professional conversations, and mini observations to help teachers improve their practice and thereby improve student achievement. Through the use of mini observations, evaluators can see a broader perspective of a teacher’s responsibilities through a series of unannounced visits throughout the course of the school year. Each mini observation is then followed by a focused professional conversation about a specific aspect of pedagogy. This conversation is a two-way communication that requires teachers to think deeply about their practice and use feedback from the evaluator to improve their instructional strategies.

The most recent literature has begun to address the use of student performance data in the evaluation of teachers. In September 2013, Needham Public Schools, like every public school district in the Commonwealth of Massachusetts, will be required to implement new regulations adopted by the Board of the Massachusetts Department of Education, which require incorporating student data into the evaluation of teachers and administrators.

The evolution of teacher evaluation began in the 1700s, when circuit superintendents visited infrequently to ensure the curriculum was being taught. In the 1900s, evaluations began to focus on a teacher’s personal characteristics. The mid-century saw increased emphasis on teacher competency and the quality of classroom instruction. The 1980s brought innovation to
teacher evaluation through public accountability. The 1990s shifted the focus to standards and
today, State Departments of Education use student achievement data to determine a teacher’s
effectiveness.

Changes in teacher evaluation were typically in response to major social upheaval such as
war, or it was spurred by publications demonstrating that American students were academically
behind their peers in other countries. The 1970s brought the most significant change in teacher
evaluation the United States had seen to date, with further alarm bells being sounded in the early
1980s in response to the publication of *A Nation at Risk* (1983).

During the last two decades in the United States, education reformers have focused on
evaluation models that pay particular attention to a teacher’s content knowledge, standards for
effective teaching, looking at student performance as a way to measure a teacher’s effectiveness,
and examining the amount of time evaluators were to spend in classrooms.

The agenda is clear for the future of teacher evaluation: The presidential initiative to
implement differentiated pay structures for teachers based on student achievement will spur the
development of new models of evaluation, no longer focused solely on teacher performance.

**Chapter III: Methodology**

**Research Questions**

Developing research questions is a critical part of an effective research study (Yin, 2009).
Mears (2009) recommends that research questions be “reduced to the smallest, definable
elements [that]…target a discernible, specific issue and population” (p. 78). Four research
questions guide this study:

1. What is the experience of teachers and administrators who have been exposed to mini
   observations?
2. How does the mini observation model impact teachers’ instructional practices as perceived by teachers and administrators?

3. What are teachers’ perceptions about reflective feedback and its ability to lead them to modify their practice?

4. Do teachers and administrators perceive any change in their relationship as a result of mini observations?

Leadership and organizational theory provides a framework for considering the efficacy of mini observations and its impact on teacher pedagogy and practice. Schmoker (2006) states there are two variables that matter most in schools: what is being taught, and how well it’s being taught. The system of mini observations provides a supervisor with the opportunity to determine the essence of curriculum and pedagogy for every teacher in the school. DuFour (2002) clearly articulates that school principals will not be effective unless they focus on learning. Elmore (2008) urges school leaders to focus on the quality of instruction as a means for improving student learning, and for teachers to engage in an ongoing assessment of how to improve their practice. Finally, Bolman and Deal (2008) suggest that leaders must be visible and accessible to effect change in the organization.

These four research questions are intended to explore the perceived relationship between mini observations and teacher performance, and to determine if a system of mini observations may have a positive impact on teacher performance. To date, there is little evidence to support Marshall’s claim that mini observations are a “powerful strategy for improving teaching and learning” (Marshall, 2009, p. 62).

Recently, several Massachusetts school districts have implemented Marshall’s system of mini observations as a model for teacher supervision and evaluation (Groton-Dunstable Public
Schools, Westwood Public School, Sudbury Public Schools, and North Andover Public Schools). However, little research has been conducted to determine if the model is more effective in improving a teacher’s instructional practice leading to higher student achievement.

The research questions are best answered within the framework of leadership and organizational theory. This theory details how individuals and groups of people act within and impact the organization. Evaluators and teachers are both subjects of this investigation as each has an important, yet different, role in the process. The evaluator is responsible for implementing the mini observation process with fidelity, and the teacher is responsible for reflecting on the supervisor’s observations and impressions shared during the reflective conversation.

**Rationale for a Qualitative Design**

This study is a qualitative case study of one suburban school district in Massachusetts. The design of this study will link the research questions to conclusions (Yin, 2009). A qualitative approach allows the researcher to study a phenomenon in a natural setting and draw conclusions based on the meaning the individual brings to the experience (Denzin and Lincoln, 2005). Merriam (2009) states that qualitative research is using words to detail data. This case study will investigate the teacher’s and administrator’s perceptions of a system of mini observations as a form of supervision and evaluation. The study will include “the voices of the participants, the reflexivity of the researcher, and a complex description and interpretation of the problem” (Creswell, 2007, p. 37).

Merriam (2009) states that education, like other fields such as health and social work, is ideal for qualitative research because it centers on people and everyday problems. Improving conditions is often accomplished by asking questions that can be researched. Merriam (2009)
continues, “research focused on discovery, insight, and understanding from the perspectives of those being studied offers the greatest promise of making a difference in people’s lives” (p. 1), and, furthermore, qualitative researchers want to know how people understand their experiences and how that meaning changes their lives. Merriam (2009) states, “The overall purposes of qualitative research are to achieve an understanding of how people make sense out of their lives…” (p. 14). This is the primary responsibility of the researcher who serves as the data collector and interpreter.

Qualitative research is preferable when the researcher needs to hear the stories and experiences of others to understand the complexity of an issue at a very detailed level (Creswell 2007). Mears (2009) points out that storytelling has been part of history for many generations, but only recognized as a bona fide methodology since the mid-20th century. The researcher interprets what “they see, hear and understand” (Creswell, 2009). Creswell (2007) states, “Unquestionably, the backbone of qualitative research is extensive collection of data, typically from multiple sources” (p. 43). Maxwell (2005) identifies four main components to qualitative research: the establishment of a relationship with participants in the study, the site and participant selection process, data collection, and data analysis.

Case Study

Creswell (2007) describes case study research as “the study of an issue explored through one or more cases within a bounded system…through detailed, in-depth data collection involving multiple sources of information…and reports a case description and case-based themes” (p. 73). The case study approach is best when the research has defined cases and seeks to study the phenomenon in-depth (Creswell, 2007). Yin (2009) agrees and defines a case study as “an empirical inquiry that investigates a contemporary phenomenon in-depth within its real-life...
context” and “relies on multiple sources of evidence, with data needing to converge in a 
triangulating fashion” (p. 18). Yin (2009) also advocates that case study research is appropriate 
when the researcher is trying to “describe an intervention” (p. 20).

The efficacy of a case study is linked to the quality of the research questions and 
triangulation of the data (Stake, 2005). Stake (2005) describes three types of case studies: 
intrinsic, instrumental, and multiple. The research conducted on mini observations is primarily 
instrumental as the case is examined “mainly to provide insight into an issue or redraw a 
generalization” (Stake, 2005, p. 445).

Merriam (2009) describes a case study as “one particular program” (p. 41) where a 
limited number of participants are part of a “bounded system” (p.41) and a particular concern or 
phenomenon. Stake (2005) describes the ongoing concentration of a case study as a system with 
a specific focus and attempts to “understand its complexities” (p. 444). This research on mini 
observations is clearly part of a bounded system (two schools within a school district), where 
participants were part of a new evaluation program not previously used in the district (mini 
observations). Like Creswell (2007), Merriam (2009) defines a case study as “an in-depth 
description and analysis of a bounded system” (p. 43) but offers a further definition including the 
notion that a case study has “particularistic” meaning:

Case studies focus on a particular situation, event, program or phenomenon. The case 
itself is important for what it reveals about the phenomenon and for what it might 
represent. The specificity of focus makes it an especially good design for practical 
problems – for questions, situations, or puzzling occurrences arising from everyday 
practice. (p.43)
The use of several sites is described by Merriam (2009) as a “multisite case study” (p. 49) 
and involves collecting and analyzing data from individuals experiencing the same phenomenon. 
The advantage of studying several cases is in the variation of the data. Patton (2002) suggests 
that a single case study is “likely made up of smaller cases” (p. 297) where themes are developed
across the cases. Merriam (2009) cites the strengths of the case study approach as “result[ing] in a rich and holistic account of a phenomenon. It offers insights and illuminates meanings that expand its readers’ experiences” (p. 51), and “has proven particularly useful for studying educational innovations, evaluating programs, and informing policy.”

Yin (2009) recommends using the case study method “when (a) “how” and “why” questions are being posed, (b) the investigator has little control over events, and (c) the focus is on a contemporary phenomenon within a real-life context” (p. 2).

Stake (1998) provides a clear understanding of the limitations of a case study with regard to generalization, writing:

The real business of case study is particularization, not generalization. We take a particular case and come to know it well, not primarily as to how it is different from others but what it is, what is does. There is emphasis on uniqueness, and that implies knowledge of others that the case is different from, but the first emphasis is on understanding the case itself. (p. 8)

This study is a single-case design with two schools in a district where the study could be replicated in other schools within the district or in other school districts within or outside of the state (Yin, 2009). Because people are at the heart of this study and are the “unit of analysis” (Patton, 2002), the case study approach is appropriate for the research questions identified.

**Site Participants**

The study was conducted in a suburban school district in Massachusetts. According to the Needham Public Schools Performance Report (2012), the Town of Needham consists of 28,886 residents. A total of 10,341 households are reported in the Town with 43% of households having children under the age of 18. In 2010, the median income for a family was $144,042; and 72% of the adult population held a bachelor’s degree or higher. The school district serves students from preschool through the twelfth grade, with five elementary schools,
one sixth grade school, one seventh and eighth grade school, and one high school covering
grades nine through twelve. The total school population in 2011 was 5,409 students. More than
1,000 students who live in the district attend private independent schools. Six percent of the
student population receives a free or reduced lunch, and 7% is either an English language learner
and/or English is not their first language. Three percent of the school population lives below the
poverty line.

By all accounts, the Town of Needham is considered an upper middle class community
and the Needham Public Schools district is considered high-performing. In 2011, the following
scores were reported on the Massachusetts Comprehensive Assessment System: 96% of tenth
graders were proficient or advanced in English Language Arts, 93% in Mathematics, and 91% in
Science and Engineering. The average combined SAT scores on Math, Critical Reading, and
Writing was 1878, and 93% of Needham High School graduates pursued post-secondary
education (Needham Public Schools Performance Report, 2012).

The Needham Public Schools, through contractual negotiations, agreed to form a study
committee of administrators and teachers with the goal of improving the district’s teacher
evaluation system. The committee met for three academic years including 2007-08, 2009-10,
and 2010-11 to explore models, review the research, and make a recommendation to the union
and superintendent. In January 2011, the committee recommended a pilot of the mini
observation model in two schools within the district to begin in September 2011. Two schools
were selected to participate in the pilot—one K-5 elementary school of 450 students and one
middle school with 430 sixth-graders. The Haggerty Elementary School and Jonas Stone School
were chosen because of their relative small size, and because evaluations are conducted by a
small number of administrators, thereby increasing the likelihood that the model can be
implemented consistently. In addition, the principals of schools were enthusiastic about the model and asked to be included in the pilot. A combined total of 37 teachers and 10 administrators have continued to participate in the pilot teacher evaluation system.

Merriam (2009) states the participants in a qualitative study are often “nonrandom, purposeful and small” (p. 16). After identifying potential participants, the researcher wrote a letter to all individuals in the pilot inviting them to participate in the research study (see Appendix A). In addition, the informed consent to participate form was distributed with the letter, which outlined the purpose of the study, procedures, and potential risks (see Appendix B). One 35-minute information session was held in each school, and the study was reviewed using a PowerPoint presentation. Ten teachers and administrators attended the information session. Following the presentation and question and answer period, participants were reminded to return the signed informed consent form if they wanted to participate in the research study. A total of 13 signed consent forms were returned to the researcher.

Of the 37 teachers in the pilot, seven agreed to an in-depth interview lasting approximately 45 minutes to one hour. Six of the 10 administrators also agreed to an in-depth interview. Three teachers and three administrators were selected for the in-depth interviews representing the typical cases as opposed to extreme or convenience selection (Mears, 2009). Teachers selected to participate included a fifth-grade teacher who had been teaching in the district for 13 years, a special education teacher who was in her first year of teaching in Needham, and a sixth-grade literacy specialist in her second year of teaching in the district. Administrators included the elementary literacy curriculum coordinator with six years in the district, a principal with three years’ experience, and a math and science curriculum coordinator
with nine years of experience as a teacher in the district, and three years’ experience as an administrator.

All 47 teachers and administrators in the study were given the opportunity to participate in an online survey using Google Docs. Twenty-five teachers and eight administrators completed 21 questions (see Appendix C and D).

Mears (2009) points out that a random sample is not the objective in qualitative research, but rather the goal is to “interview a small number of people who know a great deal about the topic” (p. 89). Seidman (2006) suggests two criteria to determine the number of participants: sufficiency and saturation. The researcher must consider if the number of participants “reflect the range of participants and sites that make up the population” (p. 55) and determine when continuing the data collection will not yield new information.

Data Collection

Yin (2009) identifies six different sources of data for case studies, “documentation, archival records, interviews, direct observation, participant-observation and physical artifacts” (p. 99). For this research study, documents such as email correspondence, meeting agendas, written online surveys, focus group records, written notes, and audio recordings from interviews were reviewed. The primary source of data consisted of teacher and administrator interviews, surveys, and focus groups, which provided an in-depth understanding of the cases and an analysis of themes across the cases. Yin (2009) describes participant interviews as “one of the most important sources of case study information” (p. 106). Asking questions that address “why” is likely to generate important data while at the same time helping the participant feel more comfortable (Yin, 2009).
Patton (2002) defines six different types of interview questions, which are experience and behavior, opinions and values, feeling, knowledge, sensory, and background questions. The research questions for this study, as well as the interviews, focus groups, and surveys, were designed to solicit opinions, values, and feelings.

Merriam (2009) suggests that “good interview questions are those that are open-ended and yield descriptive data, even stories about the phenomenon…Tell me about time when…Give me an example of…Tell me more about that…What was it like for you when…” (p. 99). The questions asked in the individual interviews and focus groups in the current study followed Merriam’s (2009) recommendation by asking questions such as “Tell me about your perception of the 10-minute observation” and “Describe your experience with the mini observation model.” Merriam (2009) suggests the interviewer avoid multiple-part questions, asking leading questions that make assumptions, or asking questions that only solicit a yes or no response. Stake (2005) advises that the researcher pay particular attention to the critical issues and states, “One might say a personal contract is drawn between researcher and phenomenon” (p. 449).

Individual interviews were recorded using the application Dictimus on an Apple iPad and by taking notes. The recordings were then transcribed by a professional transcription service. The researcher reviewed each recorded interview and transcript simultaneously for accuracy. For each focus group, the researcher recorded written notes. The anonymous survey was conducted online using Google Docs to record all responses, and a report was printed for both the teacher and administrator surveys.

**Individual interviews.** Merriam (2009) states, “Any and all methods of gathering data, from testing to interviewing, can be used in a case study” (p. 42); data collection through interviewing is common in education and other social science fields. Creswell (2005) states,
“Your research questions formulate what you want to understand; your interview questions are what you ask people in order to gain that understanding” (p. 92). Interviews are commonly conducted in person, either one-on-one or in a group, with the intent to gather data that is otherwise difficult to observe, such as an individual’s perceptions or feelings (Merriam, 2009). Merriam (2009) further states, “Interviewing is also the best technique when conducting intensive case studies of a few selected individuals” (p. 88). Interviewing is a way to gather research so that others not involved as participants in the study learn from those who did participate, thereby “creating a community of deeper understanding” (Mears, 2009, p. 49). Interviewing allows participants to provide meaning to their experience, and encourages an “understanding the lived experience of other people and the meaning they make of that experience” (Seidman, 2006, pp. 8-9). Yin (2009) believes a good interviewer must be prepared to ask relevant questions and have the ability to interpret the answers while being an active listener. Interviews also allow the researcher to understand another’s perspective that cannot otherwise be observed (Patton, 2002).

Seidman (2006) provides a framework for conducting interviews, classifying three distinct levels of listening that the researcher must employ during the interview. The first level is listening carefully to the substance and detail. The second level is listening to the participant’s “inner voice” (Seidman, 2006, p. 78) and digging deeper when the participant is using words that are not specific. Finally, the third level of listening is hearing the participant while keeping track of the logistics of the interview, such as time and the participant’s energy level, as well as any “non-verbal cues he or she may be offering” (Seidman, 2006, p. 79).

The interview structure can include three levels of interviewing: highly structured, semistructured, or unstructured (Merriam, 2009). The highly structured interview is ordered and questions are established in advance with no variation. The semistructured interview mixes
questions that are predetermined and structured with those that are less structured. The interview questions are also flexible with no set order. The unstructured interview is highly flexible, open-ended, and “more like a conversation” (Merriam, 2009, p. 89). Stake (2005) notes the importance of the researcher’s ability to be reflective when interpreting the interviews, as the researcher is gathering significant data to make meaning of the case.

Asking open-ended questions (Mears, 2009; Seidman, 2006; Yin, 2009) allows the researcher to hear an in-depth account of the participant’s experience through “the personal essay to access meaningful insights and thoughtful reflection” (Mears, 2009, p. 110). It is important for the researcher to ask additional questions when something is not understood, avoid leading the participants with questions, and take a neutral stance by neither agreeing nor disagreeing with the participant’s responses (Seidman, 2006).

Seidman (2006) provides an excellent summary of the purpose of interviewing for qualitative research:

"The purpose of in-depth interviewing is not to get answers to questions, nor to test hypothesis, and not to “evaluate” as the term is normally used. At the root of in-depth interviewing is an interest in understanding the lived experiences of other people and the meaning they make of that experience. (p. 9)"

Individual interviews (Yin, 2009) were employed to gather data and learn about teacher and administrator perceptions concerning the efficacy of mini observations. Individual interviews were conducted with three teachers and three administrators, whose years in the district ranged from 1 to 13 years. The interviews were semistructured and consisted of 19 open-ended questions that allowed interviewees to elaborate on their responses (see Appendix E). While the questions were predetermined and asked sequentially, the structure of the interview allowed for flexibility and the opportunity for the researcher to ask clarifying and follow-up questions. The interviews were conducted at each participant’s school, either in the classroom or
office. One administrator who shared an office with two other employees was interviewed in the researcher’s office, which was located one floor above the administrator’s office.

Mears (2009) advises that analysis of each interview examine how “individual patterns fit together, overlay, coincide or contradict” (p. 136), and that the researcher interpret that data through the theoretical framework and research questions identified for the study.

As stated by Merriam (2009), interviewing brings the researcher into the world of the participants. For the purpose of analyzing this study, the research questions served as the anchor while reviewing all documents. Further, notes from each interview were made that identified “themes, patterns, or discrete elements” (Mears, 2009, p. 137).

Focus group interviews. According to Merriam (2009), the introduction of focus groups in social science research dates back to the 1950s. Focus groups bring people together to share their knowledge on a shared topic (Merriam, 2009). They generally consist of six to ten participants “who know the most about the topic” and, with the help of a facilitator, come together to share their experiences (Merriam, 2009, p. 94).

For this research, three focus groups were conducted on October 25, 2011, December 7, 2011, and April 5, 2012. The first two focus groups consisted of the 10 administrators participating in the study. The questions for these two focus groups were unstructured. The final focus group consisted of seven administrators and eight teachers who met together, with semistructured questions (see appendix F). All three focus groups were voluntary and conducted either during the school day (for the administrator focus groups) or after school (for the combined administrator and teacher focus group).

Survey. On June 6, 2011, a web-based survey was sent to all participants (see Appendix C). Sixty-nine percent of teachers and 80% of administrators responded. The survey consisted
of 19 structured questions designed to solicit opinions, feelings, and perceptions about the mini observation model (Patton, 2002). Some of the questions were open-ended, such as “Describe how the mini observations helped you to grow and improve your teaching.” Other questions were structured and asked the participant to provide a rating, such as “To what degree did the mini observation model help you to grow and improve student learning?” The participants had five choices ranging from “Not at all” to “Significantly.” As pointed out by Merriam (2009), the survey did not allow for follow-up questions or probing for deeper meaning.

Data Analysis

Seidman (2006) advises researchers to carefully organize the data as interviews are completed, but to avoid in-depth analyses until all interviews are completed. He continues:

The researcher must come to the transcripts with an open attitude, seeking what emerges as important and of interest from the text…The interviewer must come to the transcript prepared to let the interview breathe and speak for itself. (p. 117)

Recording the interviews is highly recommended to ensure the preservation of the data and the ability to conduct in-depth analyses afterward (Seidman, 2006, Yin, 2009, Creswell, 2007). Once the recordings are completed, “The ideal solution is for the researcher to hire a transcriber” (Seidman, 2006, p. 115).

Case study analysis involves an in-depth description of the participants and the site, and parsing of the data to delineate themes (Creswell, 2009). For this research study, data has been coded and categorized to help develop comparisons (Maxwell, 2005). Stake (1998) advises the researcher that the “search for meaning often is a search for patterns, for consistency, for consistency within certain conditions” (p. 78). The data must be reviewed to identify and explain connections, confirmations, surprises, consistencies, and inconsistencies (Seidman,
2006). Finally, the researcher evaluates the new learning and “asks what meaning [has been] made of their work” (Seidman, 2006, p. 129).

Following Creswell’s (2009) steps for processing data, this study began with organizing the data by transcribing all interviews, collating survey and focus group data, reading through all the data and noting general themes, and coding the data into categories. Seidman (2006) advises that data be reduced and analyzed for meaning either by “creating profiles…of a participant’s experience” (p. 119), or by collating the interviews based on categories or themes. Data analysis is categorizing text and making meaning through interpretation (Creswell, 2009) and the researcher is “exercising judgment about what is significant in the transcript (Seidman, 2006, p. 118).

Seidman (2006) cautions the researcher to ensure retaining the participant’s voice when coding into categories and themes—not the “themes [the researcher] already has in mind” (p. 128). Creswell (2007) also stresses the importance of voice and urges the researcher to pay particular attention to ensuring that a participant’s “voice is not silenced, disengaged or marginalized” (p. 212). While data must be interpreted, Seidman (2006) states that researchers are compelled to examine what they have learned from those who have been interviewed and how new learning connects to the literature and prior knowledge. Interpreting the data “asks researchers what meaning they have made of their work” (Seidman, 2006, p. 129). Stake (1998) reports that researchers who use the case study method often “make assertions on a relatively small database” (p. 12).

**Coding.** Each transcript in this research study was carefully read and labeled according to common themes (coding). Saldaña (2009) defines a code as a word or phrase that summarizes the portion of recorded data; the goal of the researcher is to understand the repetition of the codes
to make meaning of the patterns and understand the links among the data. Saldaña (2009) states the act of coding is not “a precise science” but rather an “interpretive act” (p. 4). He further provides an excellent explanation of the difference between decoding (understanding meaning) and encoding (labeling the meaning with a code). For the researcher, coding is exploratory and requires one to discover the patterns and features of the data, to then “arrange things in a systematic order, to make something part of a system or classification, to categorize” (Saldaña, 2009, p. 8).

Saldaña (2009) urges the researcher to constantly think about the ways that datum are linked together:

One of the most critical outcomes of qualitative data analysis is to interpret how the individual components of the study weave together. The actual integration of key words...—a technique I call ‘codeweaving’—is a practical way of insuring that you’re thinking how the puzzle pieces fit together. (p. 36)

Saldaña (2009) advises the researcher to keep in mind seven attributes for coding: (a) a high degree of organization; (b) perseverance; (c) the ability to deal with ambiguity; (d) remaining flexible; (e) creativity; (f) maintaining high ethical standards; and (g) using language “to find just the right words for your codes, categories, themes, concepts and theories (pp. 29-30).

Saldaña (2009) describes qualitative coding in two phases. “First cycle coding” (p. 48) methods include seven different possible processes used by the researcher when first reviewing the data. For the purposes of this study, the process of “elemental methods” (Saldaña, 2009, p. 66) was used to gain a deeper understanding of participants’ experiences. Specifically, descriptive coding (single words and short phrases used to describe the data) and in vivo coding (quoting words and phrases) were used to understand and interpret the data (Saldaña, 2009). In addition to elemental methods, “affective methods” (Saldaña, 2009, p. 86) were used in coding,
specifically “values coding” (p. 89), where the “participant’s values, attitudes, and beliefs, representing his or her perspective or worldview” (p. 89) are analyzed and interpreted, particularly with data derived from the online survey.

“Second cycle coding” (Saldaña, 2009, p. 149) is the process of categorizing codes and organizing the themes that emerge from the first cycle coding. While not always necessary, second cycle coding can be beneficial in merging and reducing codes from the first cycle coding. For this study, the researcher employed “focused coding” (Saldaña, 2009, p. 155), where “data similarly…coded are clustered together and reviewed to create tentative category names with an emphasis on process through the use of gerunds” (p. 156).

While the coding methods vary from study to study, and even migrate within a specific study, Saldaña (2009) recommends the researcher ensure they are “making new discoveries, insights, and connections about your participants, their processes, or the phenomenon under investigation” (p. 51).

A qualitative software program, MAXQDA, was used to analyze codes into themes based on the data collected during individual interviews, focus groups, and written surveys. The data was reviewed “to determine whether any meaningful patterns are emerging” (Yin, 2009, p. 128). Seidman (2006) again cautions the researcher not to “force the excerpts into categories” (pp. 127-128) and to allow the data to be categorized based on participants’ experiences rather than researcher bias. Interviews are conducted with participants “to find out what their experience is and the meaning they make of it, and then to make connections among the experiences of people who share the same structure” (Seidman, 2006, p 128).

Yin (2009) advises the use of multiple sources of data (such as documentation, archival records, interviews, direct observation, participant-observation, and physical artifacts) as a means
of triangulating the data, and, furthermore, to use the evidence to develop “converging lines of inquiry” (p. 117), thus allowing the researcher a greater opportunity for accuracy in interpreting the data and presenting outcomes and conclusions. The data for this study was triangulated by using data from interviews, focus groups, and surveys.

**Validity and Credibility**

Merriam (2009) points out that case studies are limited by the integrity of the investigator and the commitment to use all the data that was collected as part of the study. In the past, case study research has been criticized for the lack of rigor, investigator bias, and the difficulty in generalizing a case to the larger population (Merriam, 2009). She suggests, “The inclusion of multiple cases is, in fact, a common strategy for enhancing the external validity and generalizability of your findings” (p. 50).

The ethical conduct of a researcher conducting a qualitative research project will reduce any threats to validity and reliability (Merriam, 2009). 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). Maxwell (2005) argues that the collection of solid evidence will help to mitigate possible concerns with validity.

Mears (2009) describes validity as the degree to which the investigation is creditable with respect to the consistency of design and process, as well as disclosure of inconsistencies within the process. Because interviewing involves human subjects, their recollections can be inconsistent; thus it is responsibility of the investigator to link the stories together to “reveal the greater truth” (Mears, 2009, p. 25). Mears (2009) further notes that reliability is best vetted by allowing the participants to review the transcripts for accuracy. For this research study, each participant was emailed a copy of the recorded transcript and encouraged to provide feedback on
accuracy. In all six cases, no participants provided any feedback on accuracies or inaccuracies in the transcripts. Creswell (2007, 2009) provides a framework for understanding validation in a qualitative research study, which is to include the “accuracy of findings” (p. 206), the amount of time the researcher spends collecting data and engaging participants, examining the process as a way to confirm validity, and “employ[ing] accepted strategies to document the ‘accuracy’ of their study” (p. 207). He emphasizes several strategies for validation, such as including the strength of the relationship developed between the researcher and participant; using multiple sources of data; identifying a “peer reviewer…who keeps the researcher honest” (p. 208); openly disclosing the researcher bias; allowing the participants to review the data and provide comment and clarification; providing “rich, thick descriptions” (p. 209) so that the reader can determine if the research is applicable to other settings; and using a third party expert to “audit” (p. 209) the data and conclusions.

Mears (2009) cautions the researcher about objectivity and bias in conducting interviews. There is the potential for the researcher to lead the participant in certain directions with both questions and follow-up inquiries. Having a background in the subject matter can either increase or diminish the likelihood of bias; accordingly, caution must be taken when conducting the research in a site familiar to the researcher, such as the workplace (Mears, 2009). “Research from within the setting becomes more challenging, for it requires overcoming your personal lens in order to understand from other’s point of view” (Mears, 2009, p. 83).

Threats to validity exist in any research study. Yin (2009) suggests that case study research is best protected when a “chain of evidence” is maintained and the researcher investigates “rival explanations” (p. 3). Merriam (2009) contends, “Issues of validity and reliability can be best addressed through a careful construction of the research and the way in
which data are collected, analyzed and interpreted” (p. 210). For the purpose of this research study, a careful chain of evidence has been maintained in the case study database (Yin, 2009).

“Power differentials” (Mears, 2009, p. 100) did exist in this study as the researcher also held a position of authority in the school district, as Director of Human Resources. In addition, the participants were aware that this study was not only for research purposes, but also for the district’s desire to develop and pilot a new model for teacher evaluation.

Case study research, like other designs, must pay attention to the quality of the design. Yin (2009) recommends four tests to measure the quality:

**Construct validity.** “Identifying correct operational measures for the concepts being studied” (Yin, 2009, p. 40).

**Internal validity.** The researcher recognizes that other factors could have played a role in the outcome other than the relationship factors identified by the researcher and the vetting of inferences. Merriam (2009) explains that the match between reality and findings is the internal validity and “hinges on the meaning of reality” (p. 213). Internal generalization examines the degree to which the findings can be applied to others within the setting, such as other teachers and administrators in the Needham district (Maxwell, 2005).

In the present study, the researcher, an employee of the district, oversees the supervision and evaluation process and has observed that bias does exist. While the researcher was cognizant of his role as both a researcher and employee with potential influence of the employee participants, attention was paid “not to eliminate this influence, but to understand it and to use it productively” (Maxwell, 2005, pp. 108-109).

**External validity.** Explores how the results from one study generalize to others in the wider population such as teachers and administrators in the State and country.
Reliability. In any research study, a different investigator would find similar results if the same procedures were followed.

For the purposes of this study, the triangulation of data reduced the “potential problems of construct validity…because multiple sources of evidence provide multiple measures of the same phenomenon” (Yin, 2009, pp. 116-117). In addition, a “case study database” (Yin, 2009, p. 119) was developed to secure and record interview transcriptions, notes, survey responses, and focus group records. All data resides in files on the researcher’s computer and backed up daily by the use of Time Machine (Needham Public Schools server), CrashPlan, a cloud-based backup storage service and weekly back up using an external hard drive. Because the researcher was able to access the data in the case study database, the “reliability of the entire case study…markedly increases” (Yin, 2009, p. 119).

Protection of Human Subjects

Seidman (2006) and Mears (2009) articulate clearly that the researcher must be “guided by principles of fairness and equitability; and is committed to do no harm to the individuals in the study, to faithfully present findings, and to fulfill both the letter and the intent of legal requirements for research conduct” (Mears, 2009, p. 32). The National Commission for the Protection of Human Subjects of Biomedical and Behavior Research (1979) specifies the ethical principles for the protection of human subjects, which are (a) respect for persons; (b) beneficence; and (c) justice. Respect for persons includes the right to volunteer to be part of the study, to be fully informed, and to protect those who have diminished capacity. Beneficence requires the participant to be protected from harm and to ensure that all participants are cared for at all times. Justice refers to ensuring that all participants are selected as well as treated equally, ethically, and fairly.
Seidman (2006) points out that “in-depth interviewing does not pose the life and death risk of biomedical research, but it is not risk free” (p. 60). Because the interviewer is spending time with the participants who are sharing personal experiences, the researcher must be aware that some discomfort may be caused by the interview itself (Seidman, 2006). “Informed consent is the first step towards minimizing the risks participants face when they agree to be interviewed (Seidman, 2006, p. 61).

Seidman (2006) outlines eight significant parts of informed consent: (a) the invitation to participate; (b) the risks of participation; (c) participant rights; (d) possible benefits to participants; (e) confidentiality of records; (f) how the research will be disseminated; (g) parental consent for children under the age of 18; and (h) contact information for both the researcher and the university’s Institutional Review Board.

For this research study, each participant signed the informed consent form and each was given a pseudonym to protect confidentiality. Participants were informed that all data would be held in strict confidence (Fraenkel and Wallen, 2009), and that all audio recordings would be destroyed upon conclusion of the research.

**Conclusion**

Merriam (2009) describes seven competencies required for qualitative researchers: (a) developing questions aimed at discovery; (b) the ability to tolerate change as the nature of qualitative data may lead to different thinking; (c) skilled observations; (d) asking open-ended questions that allow for deeper inquiry; (e) “thinking inductively” (p. 17); and (f) the ability to write effectively.

This study was conducted to gain an in-depth understanding of how a system of mini observations impact both teachers’ and administrators’ practices in the classroom in an effort to
improve student learning. The theoretical framework of leadership and organizational theory guided this study. This theory systematically details how individuals and groups of people act within and impact the organizations with which they are involved, and was based on the work of DuFour (2004); Elmore (2008); Bush (2009); Gallos (2008); Bolman and Deal (2008); Schmoker (2006); and Marzano, Waters, and McNulty (2005).

Teacher supervision and evaluation has never been more scrutinized as it has been in the last decade. The primary focus of improvement must center on what is being taught in the classroom and how well it is being taught (Schmoker, 2006; DuFour, 2002). With a high demand for improving student achievement from policy-makers at both state and federal levels, the relationship between teaching and student learning is now being studied. The Massachusetts Department of Elementary and Secondary Education (2012) has developed a model evaluation system consisting of a five-step cycle that involves self-assessment, goal setting, developing an improvement plan, formative assessment, and a summative evaluation. Mini observations are part of the model, and a connection between student test scores and teacher performance will be introduced in 2014 for many school districts.

This research represents an attempt to gain a better understanding about how teachers and administrators perceive mini observations related to improved instructional practice in the classroom. The researcher conducted a qualitative case study to find answers to four research questions: (1) What is the experience of teachers and administrators who have been exposed to mini observations? (2) How does the mini observation model impact teacher’s instructional practices as perceived by both teachers and administrators? (3) What are teachers’ perceptions about the reflective feedback and its ability to lead them to modify their practice? (4) Do teachers and administrators perceive any change in their relationship as a result of participating
in mini observations? Participants in the mini observation pilot were carefully selected to represent a range of teachers and administrators with varying degrees of experience.

Data was collected using multiple sources, including face-to-face interviews with six participants, a survey with 33 teachers and administrators responding to 21 questions, and three focus groups conducted over the course of the school year. For this research study, data was coded and categorized to help develop comparisons (Maxwell, 2005) and interpret meaning.

The ability to generalize the findings of this study will be limited, based on the small number of participants (Mears, 2009). Yin (2009) states, “case studies, like experiments, are generalizable to theoretical propositions and not populations or universes” (p.15).

**Chapter IV: Report of the Research Findings**

This chapter describes and discusses the key findings from the research conducted over a 9-month period in two suburban schools in Massachusetts. The first section provides an overview of the study’s context. The following sections present the themes that have emerged from the individual interviews, surveys, and focus groups as related to the four research questions that have guided the study. The final section presents a summary of the most salient research findings.

**Study Context**

Through the decades, teacher supervision and evaluation have been ubiquitous topics in the professional literature, with the most significant changes in the field occurring in the last two decades. In 2009, the Federal Department of Education launched the Race to the Top initiative and redefined teacher evaluation in the nation’s public schools. Massachusetts, as a participant in this federal initiative, promulgated regulations requiring sweeping changes to the process of teacher evaluation. At the directive of the initiative, each school district negotiated the evaluation
process as a mandatory subject of collective bargaining. As a result, the Race to the Top regulations have provided a clear direction for all 351 districts in the Commonwealth by offering a model evaluation system that articulates a five-step cycle for every educator. One of the five steps includes “judgments based on observations and artifacts of professional practice, including unannounced observations of practice of any duration” [emphasis added] (Massachusetts Department of Elementary and Secondary Education, 2012, CMR 35.07).

In the winter of 2007, the Needham Public Schools entered into negotiations with the teachers’ union to discuss a new collective bargaining agreement. A key issue for the school committee and administration was the development of a new model to evaluate teachers. Many people in the district held the general belief that the traditional model of evaluation had become antiquated, inconsistent, ineffective, and imbalanced as it related to educational efforts and outcomes. One administrator during the interview described the traditional evaluation model in the following way: “. . . for me as [the] principal, it did not serve the purpose that it was supposed to, lifting teacher learning to lift student learning, to create relationships amongst the community members. I think, if anything, it fed mythology and fantasy . . .” Another teacher remarked, regarding her experience with the traditional evaluation model, “Observing just 2-3 times a year, for whole class periods, does not often give an accurate picture of my teaching, rather a quick snapshot of what happened on those days, with little time to reflect as a teacher. It felt more like a performance some days than an evaluation.”

When asked about the degree to which the traditional evaluation model helped Needham teachers to grow and improve student learning, 88% of the teachers surveyed felt either indifferent or that it did not help in a significant way (see Figure 1).
Figure 1. To what degree did the traditional model help you to grow and improve student learning?

In addition, when teachers were asked to rate their overall satisfaction with the traditional evaluation model, 92% of teachers surveyed indicated they were either indifferent or were not satisfied with the model (see Figure 2).

Figure 2. Please rate your satisfaction with the traditional evaluation model.
Needham school administrators were asked similar questions in a survey about the traditional observation model. As can be seen in Figure 3, 88% of the administrators surveyed were either indifferent to or believed that the traditional approach to evaluation did not help teachers to grow and 100% of the administrators surveyed were either indifferent or were not satisfied with the traditional evaluation model (see Figure 4).

Figure 3. In your opinion, to what degree did the traditional model help teachers to grow and improve student learning?
On January 13, 2010, a focus group was conducted with 40 volunteer teachers from all schools within the Needham district. The teachers believed the strengths of the traditional model included the following: different evaluation options for teachers who had obtained professional status; opportunity to discuss concerns about working conditions with administrators; opportunity to discuss goals prior to the observation; and a three-year probationary period for teachers who were new to the district. The teachers believed the weaknesses of the traditional model included: a lack of authenticity; assessment of too many categories; little monitoring of the administrator’s recommendations; the promotion of contrived lesson plans; only a small window of a teacher’s entire professional responsibilities was emphasized; and the model was judged to be high-stakes and stressful. In response, the teachers recommended improving the traditional evaluation model by including more frequent observations, an option for peer evaluation and mentoring, and opportunities for teachers to document their growth.

The district leadership team, consisting of 40 central office administrators, principals, assistant principals, department chairs and directors, and curriculum coordinators met on January
10, 2008 to discuss the strengths and weaknesses of the teacher evaluation model as well as recommendations for improvement. While the district leadership team agreed the traditional model was built on broad-based standards, included multiple classroom observations (up to four), and required evidence of student growth, the managers found the model to harbor many more weaknesses than strengths. For example, they cited the model lacked the use of multiple sources of data. In addition, it was a passive process for teachers and required the evaluator to do most of the work, which included significant time for the completion of written reports. Finally, the managers agreed that the model did not foster collaborative, collegial relationships with the teachers. The district leadership team advocated for a new model including more classroom observations, fewer written reports, more collaborative conversations with teachers, and a model based entirely on standards.

Table 1 provides a summary of the additional themes gathered from in-depth interviews, survey data, and focus group notes at that time.

Table 1

<table>
<thead>
<tr>
<th>Theme</th>
<th>Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lacking authenticity</td>
<td>What does it really measure? [The model] is not representative of what the teacher does every day.</td>
</tr>
<tr>
<td>High stakes</td>
<td>With the older evaluation, they thought I was there to catch them, to trick them, to criticize them, to find fault (administrator).</td>
</tr>
<tr>
<td></td>
<td>They were going to come in and watch for 45 minutes and I am just going to be sweating the whole time (teacher).</td>
</tr>
<tr>
<td>Ineffective model</td>
<td>I think part of the problem with the traditional evaluation is it was so long and so drawn out and we just didn’t have the effectiveness.</td>
</tr>
<tr>
<td>Lack of immediate feedback</td>
<td>If you’re not able to meet directly afterwards or very soon after the [observation] it becomes a lot more challenging to have those conversations because it’s not immediate in people’s head.</td>
</tr>
</tbody>
</table>
Writing challenges

I found the scripting [of the observation] sometimes to be really, really challenging especially in the content areas that I supervise. I can’t engage in the same ways if I were doing in the mini observation as I would ask the kids more about what they are doing and what the goal is and so I think that that is one piece.

I would find myself at night writing and writing and writing.

Time intensive

[I spent a] significant amount of time in observations and writing [them] up. I was spending hours and hours and hours, and then the summative evaluation—hours and hours and hours, and I just did not get the impact in terms of relationship building, student learning. It just was a tremendous demand on my time as principal that did not really impact the culture, the community, or the learning.

High stress

I had first year teachers on the traditional [evaluation model]. I think for them it was very nerve-racking. Okay, the administrator is coming in to [the classroom] today to evaluate. So the first couple of times of me coming to their classroom they were like flush red. Because the initial reaction of me being in their presence again was oh my gosh, she is here (administrator).

I think the main thing is that when you have a scheduled observation you planned this lesson and it is what it is because you planned it like the way that an actor planned a play or speech. And there is this high anxiety about deviating from that.

Value in written reports

The write up to me is a real reflection of the conversation that you had (administrator).

I think one thing I liked that I got in the traditional observation was the written feedback. I know that that’s like an anathema for the administrators. They don’t like doing it, it’s a pain in the neck, you are kind of grasping for straws as to what to say. I get all that, it’s very time-consuming (teacher).

In 2007, a study committee of administrators and teachers began meeting with the goal of improving the teacher evaluation system. The committee met for three academic years including 2007–08, 2009–10, and 2010–11 to explore models, review the research, and make a recommendation to the teachers’ union and superintendent. In January 2011, the committee
recommended that a pilot program, including the mini observation model, be instituted in two schools within the district beginning in September 2011. In response, the superintendent and executive board of the Needham Education Association accepted the recommendation.

In preparation for the program’s implementation, Kim Marshall conducted training for teachers and administrators on February 8th and March 29th of 2011. The training focused on how to conduct short, five–ten minute observations. The training instructed evaluators on how to give effective feedback to teachers in brief, engaging, and informally reflective ways. Evaluators also learned organizational issues such as tracking and record keeping as well as how to use of standards-based rubrics as a method to conduct formative and summative assessments of teachers. All teachers and administrators assigned to the Haggerty and Jonas Stone schools, respectively, participated in the four-hour training program.

Participants. A total of 37 teachers and 10 administrators participated in the mini observation pilot program. The teachers participating in the pilot included first, second and third-year teachers as well as those who had achieved professional status. Kindergarten through grade 6 classroom teachers were represented, as well as specialist teachers including special education, reading, music, library, speech, and physical education. The administrators included two principals, three curriculum coordinators, three K–12 directors, and two special education directors.

Of the 47 participants in the pilot, 13 teachers and administrators volunteered for in-depth interviews as part of the researcher’s study on mini observations. Three teachers and three administrators were randomly selected for an in-depth interview with the researcher as they represented the typical cases (Mears, 2009). Table 2 shows the participants’ demographic information.
Table 2

**Participant Demographic Information**

<table>
<thead>
<tr>
<th>Participant Name</th>
<th>Role</th>
<th>Grade</th>
<th>Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamie</td>
<td>Administrator</td>
<td>6</td>
<td>Interview, survey, focus group</td>
</tr>
<tr>
<td>Bonnie</td>
<td>Administrator</td>
<td>K–5</td>
<td>Interview, survey, focus group</td>
</tr>
<tr>
<td>Tara</td>
<td>Administrator</td>
<td>6–8</td>
<td>Interview, survey, focus group</td>
</tr>
<tr>
<td>Dan</td>
<td>Administrator</td>
<td>K–12</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Diane</td>
<td>Administrator</td>
<td>K–12</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Kara</td>
<td>Administrator</td>
<td>6–8</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Kim</td>
<td>Administrator</td>
<td>K–12</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Meg</td>
<td>Administrator</td>
<td>K–5</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Mark</td>
<td>Administrator</td>
<td>K–5</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Wes</td>
<td>Administrator</td>
<td>6–8</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Millie</td>
<td>Specialist Teacher</td>
<td>6</td>
<td>Interview, survey, focus group</td>
</tr>
<tr>
<td>Rana</td>
<td>Teacher</td>
<td>5</td>
<td>Interview, survey, focus group</td>
</tr>
<tr>
<td>Edna</td>
<td>Specialist Teacher</td>
<td>K–5</td>
<td>Interview, survey</td>
</tr>
<tr>
<td>Don</td>
<td>Teacher</td>
<td>6</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Kathy</td>
<td>Specialist Teacher</td>
<td>K–5</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Tabitha</td>
<td>Teacher</td>
<td>6</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Cathy</td>
<td>Teacher</td>
<td>2</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Lela</td>
<td>Specialist Teacher</td>
<td>K–5</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Jack</td>
<td>Teacher</td>
<td>6</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Abigail</td>
<td>Specialist Teacher</td>
<td>6</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Edda</td>
<td>Specialist Teacher</td>
<td>6</td>
<td>Survey, focus group</td>
</tr>
<tr>
<td>Abbey</td>
<td>Teacher</td>
<td>4</td>
<td>Survey</td>
</tr>
<tr>
<td>Abra</td>
<td>Specialist Teacher</td>
<td>6</td>
<td>Survey</td>
</tr>
<tr>
<td>Bailey</td>
<td>Teacher</td>
<td>6</td>
<td>Survey</td>
</tr>
<tr>
<td>Cady</td>
<td>Specialist Teacher</td>
<td>K–5</td>
<td>Survey</td>
</tr>
<tr>
<td>Cindy</td>
<td>Specialist Teacher</td>
<td>K–5</td>
<td>Survey</td>
</tr>
<tr>
<td>Diana</td>
<td>Specialist Teacher</td>
<td>6</td>
<td>Survey</td>
</tr>
<tr>
<td>Debbie</td>
<td>Teacher</td>
<td>K</td>
<td>Survey</td>
</tr>
<tr>
<td>Edena</td>
<td>Specialist Teacher</td>
<td>K–5</td>
<td>Survey</td>
</tr>
<tr>
<td>George</td>
<td>Teacher</td>
<td>K</td>
<td>Survey</td>
</tr>
<tr>
<td>Hillery</td>
<td>Teacher</td>
<td>3</td>
<td>Survey</td>
</tr>
<tr>
<td>Jill</td>
<td>Teacher</td>
<td>6</td>
<td>Survey</td>
</tr>
<tr>
<td>Jackie</td>
<td>Specialist Teacher</td>
<td>K–5</td>
<td>Survey</td>
</tr>
<tr>
<td>Jeff</td>
<td>Teacher</td>
<td>6</td>
<td>Survey</td>
</tr>
<tr>
<td>Lucy</td>
<td>Teacher</td>
<td>5</td>
<td>Survey</td>
</tr>
<tr>
<td>Lisa</td>
<td>Teacher</td>
<td>5</td>
<td>Survey</td>
</tr>
<tr>
<td>Mary</td>
<td>Teacher</td>
<td>1</td>
<td>Survey</td>
</tr>
<tr>
<td>Meghan</td>
<td>Specialist Teacher</td>
<td>6</td>
<td>Survey</td>
</tr>
<tr>
<td>Melissa</td>
<td>Teacher</td>
<td>6</td>
<td>Survey</td>
</tr>
<tr>
<td>Nancy</td>
<td>Teacher</td>
<td>6</td>
<td>Survey</td>
</tr>
</tbody>
</table>
Note. Teachers and administrators who completed the survey did so anonymously. The survey was given to all 47 participants, and 33 participants completed the survey.

Program design. The Needham School Committee and the Needham Education Association agreed to pilot a new teacher evaluation model for the 2011–2012 school year at both the Haggerty and Jonas Stone Schools, respectively. The mini observation model consisted of 12–15 short observations lasting 10–15 minutes for each teacher over the course of the school year. The observations were varied throughout the day and unannounced to the teacher. The administrator and teacher were required to meet within 24–48 hours of the observation and to engage in reflective conversations about the mini observation.

At the end of the year, a summative evaluation was conducted using rubrics developed by Marshall (2009). The rubrics rated teachers on six standards subdivided into 60 indicators. For each indicator, the administrator rated the teacher as either highly effective, effective, improvement necessary, or does not meet the standard. Before the final summative conference was scheduled, both the teacher and administrator independently scored the rubric on all 60 indicators. At the summative meeting, both sets of rubrics were reviewed, and the teacher and administrator engaged in a conversation to discuss similarities and differences. The administrator then scored the final rubric evaluation, obtained the signature of the teacher, and sent the original copy to the Human Resources Department at the central office.
Each year, approximately half of the teachers in the Needham school district are on cycle for evaluation while the other half are not scheduled for evaluation. Teachers in their first three years in Needham schools are considered pre-professional status and are evaluated in each of the three years. Pre-professional status teachers are employees at-will and can be terminated without cause. Teachers who earn professional teachers status are evaluated every other year, and have the choice of either participating in classroom observations or electing to engage in a project-based activity. Teachers with professional status in Massachusetts can only be terminated for “inefficiency, incompetency, conduct unbecoming a teacher, insubordination or failure on the part of the teacher to satisfy teacher performance standards” (Mass. Gen. Law, ch. 71, §42, 1973).

Teachers at the Haggerty and Jonas Stone schools, respectively, who had obtained professional status or were in their third year in the district also had the option of participating in the pilot. Teachers in their first or second year were mandated to be part of the mini observation pilot. In total, 37 teachers and 10 administrators were part of the mini observation pilot. All 47 teachers and administrators were asked to be part of an in-depth interview as part of this research project. Thirteen teachers and administrators consented to be part of the study.

Research Question #1. What is the experience of teachers and administrators who have participated in mini observations?

Based on a review of the in depth interviews, results of the surveys and focus groups, both teachers and administrators believed that mini observations provided significantly more benefits than the traditional model of teacher supervision and evaluation. Administrators were surprised to learn how their focus shifted from always watching the teacher to interacting with the students leading to more informed observations. Teachers discovered that their level of
stress had decreased under the pilot’s mini observation model. They also felt more confident because administrators were able to see more of the teachers’ total job responsibilities with the new model.

The following themes emerged from the information gathered during the in-depth interviews, responses to the survey, and from notes taken during the focus group meetings:

1. Organizational benefits of the model
2. A greater focus on students
3. The stakes for teachers are significantly lower
4. The overall benefits of mini observations
5. Challenges with implementation

**Organizational benefits of the model.** The administrators spoke frequently about the ease with which they could conduct mini observations when they were not required to schedule a time to observe the teacher. Administrators reported that despite a very busy schedule, they often found small pockets of time to conduct a mini observation. Most administrators reported that the mini observation model gave them greater flexibility in doing their job. During the interview, one administrator remarked about the traditional evaluation model, “It’s a lot harder when you’re scheduling a preconference, scheduling an observation, scheduling the post conference, and then doing the write up, and then making sure also that you’re following up, and getting things back [signed observation reports] from people in a timely manner.”

Administrators described the benefits of not having to write reports after each observation or summative evaluations in the mini observation model. They commented that writing observation reports and summative evaluations was “painful” and “not worth doing.” One
administrator even said it was a “sigh of relief” not having to write reports. Not having to write reports also allowed administrators greater flexibility and more time to conduct observations.

Administrators agreed that having varied observation times gave them a better picture of the teaching and learning taking place in their schools. The traditional evaluation mode forced administrators to schedule 45 minutes for observations of classroom instruction. The mini observation model, because of the frequency of visits, allowed administrators to observe many different parts of a teacher’s day including the testing of students, during transitions from one class to the next, at the beginning of a lesson, at the end of a lesson, at the start of the school day, and the end of a school day. The varied observation schedule and the flexibility of needing only 10–15 minutes to conduct an observation were seen as a significant benefits to all administrators.

In spite of these benefits, all of the administrators did speak about the challenges of conducting 12–15 mini observations for each teacher they were supervising and the difficulty of having the reflective conversations within 48 hours of the observation. During the interview, one administrator noted about the reflective conversations:

I don’t think you can have them on the run. I don’t think you can have them over the phone. I think the benefit of this really comes from the developing of the relationship, the professional relationship that can only take place when you’re sitting eye to eye, knee to knee. That got complicated and that required not only flexibility on my part, but on the teacher’s part.

A greater focus on students. All administrators interviewed reported a significant increase with their focus on and interactions with students. The traditional observation model centered on the teacher and scribing the observation. The mini observation model, however, allowed them to interact with the students, to look over their work, and to ask questions such as, “What are your goals?” “Are you excited by this activity?” “What are you learning form this activity?” Administrators reported having made deeper connections with students because of the
The amount of time they were spending in the classrooms. Teachers agreed that it was highly beneficial having the administrators in their classrooms, especially to the benefit of the students. One teacher commented, “Having other adults in the classroom watch our lessons shows the students that others care about what we are doing.”

**The stakes for teachers are significantly lower.** Other than building stronger relationships and authentic conversations, teachers and administrators mentioned the reduction of anxiety with the mini observation model more frequently than any other theme. Teachers reported the brief and frequent observations had lowered the stakes considerably as they were no longer spending inordinate amounts of time planning for the longer observations and could resume focus on the practice of teaching each day. Several teachers commented that while the anxiety was lower, they must be “on their game” all of the time as an administrator could observe at any moment. During the interview, one teacher commented, “Because it wasn’t a planned visit, I didn’t stress out the night before. He [the principal] just he showed up and that was it, and so it really was a true reflection of me as a teacher in the classroom.” Another teacher stated, “I think the huge thing that’s better about the mini observation is that you don’t change your teaching, you just teach what you teach and they come, they see what they see.”

The anxiety-related results of the survey showed a significant decrease in anxiety with the mini observation model. Figure 5 shows that 80% of the teachers surveyed experienced less anxiety with the mini observation model. One teacher stated, “I feel much less nervous now when my principal comes into the room.” A different teacher commented, “I think it makes me less anxious when the supervisor walks in and I am better able to deliver instruction therefore increasing student learning.”
One administrator noted, due to frequent classroom visits, “I think a lot of that anxiety is disappearing because teachers are not thinking ‘I wonder what they are seeing? What are they looking for? What they are going to get me on?’ Because it is now part of the system.”

Administrators also reported the perception of intimidation had decreased, especially with teachers in their first year of teaching in the district. One administrator observed, “They’re much less intimidated with this model. It’s much more inclusive. It’s much more informal.” Another administrator stated, “This year, [with the mini observation model] I didn’t notice really any anxiety. This [model] is to support you in your learning. . . . I think people could really hear that. They got it. With the older evaluation, they thought I was there to catch them, to trick them, to criticize them.”

Figure 5. Compared to the traditional observation model, was your level of anxiety more, less, or about the same with the mini observation model?

Administrators were asked a similar question that related to how they perceived the teacher’s anxiety as tied to the evaluations. Eighty eight percent of the administrators believed that teachers experienced less anxiety with the mini observation model (See Figure 6).
Figure 6. Compared to the traditional observation model, do you think that the teacher’s level of anxiety was more, less, or about the same with the mini observation model?

The overall benefits of mini observations. Teacher and administrators agreed that mini observations and reflective conversations yield a deeper and more meaningful effect on a teacher’s performance. Many of the participants commented on the value of the 10-minute observation despite their initial apprehensions. In addition, administrators and teachers expressed surprise when they learned that administrators actually developed a deeper understanding of and appreciation for a teacher’s total responsibility, which extended far beyond the classroom. The frequency of classroom visits also mattered to the teachers. Teachers believed their observing administrators gained both a micro and macro perspective of curriculum and instruction because they observed a high degree of detail in each classroom. As a result of this dynamic perspective, administrators were able to make connections between each classroom and grade level. Administrators believed the value of frequent observations helped them to see the curriculum in a broader context within their respective school.
Ten minutes is longer than anticipated. During the training with Marshall, both teachers and administrators were skeptical about the efficacy of a 10–15 minute observation. However, based on interviews, survey, and focus group data, administrators reported that a 10–15 minute observation actually provided the right amount of time to understand the teaching and learning happening in the classroom. Because of the model’s flexibility, administrators could shorten or extend their observation as needed. Administrators further reported they were more focused during a mini observation. During the interview, one teacher reported:

I realized how much you actually can see in 10 minutes. You can see a lot, just the way questions are asked, the way you call on students, even looking at, ‘did I call boy-girl-boy-girl.’ I mean all in 10 minutes there were so many different things I could see about: the way I phrased things, who I called on, how much I accomplished in that 10 minutes, was time wasted, and was I using the time effectively. It was amazing, so I definitely think that you can see more in 10 minutes than most people would think you could. I think you almost have to watch it to understand it.

Despite shorter observations, the frequency of the classroom visits allowed administrators to paint a clearer picture of a teacher’s overall progress as related to instructional growth. One administrator commented, “I really got to see teachers using or not using the recommendations, growing or not growing in a way that was very different from before.” Administrators also reported feeling more focused while observing for a 10-minute period. One administrator said, “I grew to be quite sophisticated in looking at [instruction] in 10 minutes.” A different administrator stated, “I think I could acclimate really quickly, get the big picture really quickly and I got better at that.”

Administrators reported the mini observation model brought more flexibility to the supervision and evaluation process. Several administrators shared that, during the in-depth interviews and surveys, they were able to extend their observations if a few more minutes would provide further insights. One administrator commented, “The 10 minutes was long enough. I will also add, however, if we were into something really important and I needed to stay, I gave
myself permission. I felt very comfortable staying in for an extra five minutes. I felt this model had flexibility . . .” Another administrator reflected, “I think the 10 minutes is really nice. You get a really good perceptive on what’s going on and if there is a question you can stay for a little bit longer . . .”

One administrator provided a summary that best captures the value of the 10-minute observation. During the interview, she stated:

I pay attention to a lot more now. I am walking in and I am looking at the walls, I am looking at the kids, I am looking at the teacher, I am looking at the books, and I am walking in to get a ‘feel’ versus documenting every little behavior [emphasis added] like, teacher went to the left, or the teacher said ‘good morning’ to the kids.

Administrators have a better understanding of the classroom and see a wider scope of a teacher’s responsibilities. Administrators reported gaining a broader lens when viewing each teacher’s full responsibilities under the mini observation model. With the traditional model of evaluation, administrators observed teachers for 45 minutes in the classroom setting. With frequent, short observations, they reported evaluating teachers in team meetings, interacting with colleagues, meeting with parents, and collaborating with guidance counselors which was not part of the traditional evaluation model.

Teachers also reported the value of the observing administrator seeing a wider view of their teaching, including their other responsibilities outside of the classroom, because of the mini observation model. One teacher remarked, “You [the administrator] get to see many transitions, many parts of my day. So it was transitions, morning meetings, one-on-one conferencing, students working independently, using technology, whole class lessons, small group work.” Another teacher remarked, “The mini observations offer a better sample, and a more representative sample of my craft than the single observation model.”
Both teachers and administrators believed the mini observation model provides an opportunity for the administrator to understand the culture of each teacher’s classroom. Eighty-eight percent of teachers and administrators believe the mini observation model helps administrators to better understand the classroom culture (See Figures 7 and 8).

*Figure 7.* Using the mini observation model, my evaluator has a good understanding of my classroom culture.
Figure 8. The mini observation model gave me a good understanding of each teacher’s classroom culture.

**Frequency of visits.** Both teachers and administrators agreed upon the value of frequent classroom visits. Teachers believed it was important for administrators to gain a deeper understanding of what goes on in the classroom. More frequent classroom visits facilitated this understanding and enabled administrators to become better attuned to each teacher’s skill set. On the whole, administrators gained a broader perspective of how the school was functioning overall.

Because of the frequency of observations, one administrator was able to see the results of the professional development, provided during department meetings, in action in the classroom. Administrators also made richer connections between the curriculum in one classroom and what was being taught in other classrooms. One teacher stated, “I actually think it is more helpful for the administrators to get into a lot of classrooms to see patterns that I think can help improve the school as a whole.”

Forty percent of teachers believed that 9–11 classroom mini observations will help them to grow professionally. Fifty percent of administrators believed 9–11 observations are sufficient for evaluation, and the other 50% believed 12–15 observations are necessary over the course of a school year to have an impact on teaching.
Figure 9. How many mini observations do you feel are necessary to have an impact on teaching?

Figure 10. How many mini observations do you feel are necessary to have an impact on teaching and learning?

Challenges with implementation. It is clear from the in-depth interviews, surveys, and focus groups that one prominent theme emerged as a challenge regarding the mini observation
model. Both teachers and administrators commented frequently about the issue of time. In some cases, administrators struggled to complete the required 10–15 mini observations. One teacher complained bitterly, "If you are supposed to be observed, it is my understanding that your administrator should observe you. It feels very badly to be ignored and reflects poorly on our district’s commitment to professionalism.”

Both teachers and administrators agreed it was difficult to find the time to engage in a reflective conversation within 48 hours of the mini observation given the schedule demands of both groups. One administrator shared her experience of conducting a reflective conversation in the hallway between two periods while the students were transitioning. The next day, the teacher met the administrator in her office and expressed that she preferred a reflective conversation that exceeded the 48–hours requirement to a conversation that felt rushed and like it was just “going through the motions.”

Data from in depth interviews, surveys, and focus groups clearly indicated the only challenge in the mini observation model was the issue of time. Both teachers and administrators reported the difficulty of finding the time to engage in the reflective conversation given the other demands of their jobs.

**Research Question #2. How does the mini observation model impact teachers' instructional practices as perceived by teachers and administrators?**

Teachers and administrators both report that the mini observations had a positive effect on classroom teaching practices. Because engaging in reflective questions was inherent in the mini observation process, both teachers and administrators agreed that instruction was immediately honed and refined, and often revisited in other follow-up conversations. Mini observations in classrooms allowed administrators to conduct frequent evaluations of student
learning, instruction, and assessment leading to focused conversations with teachers on a regular basis.

The two primary themes that emerged from the data regarding the answering of this question were:

1. Reflection on instruction and how it impacts students
2. Deeper thinking about practice

**Reflection on instruction and how it impacts students.** Teachers and administrators were not able to provide specificity about the impact on student learning. The perception was that the mini observation model has the potential to improve student outcomes. One administrator stated that “it allowed for more frequent check-ins on student learning, instruction, and assessment.” A different administrator said, “This has to result in positive teaching strategies in the classroom for kids.” Through a survey, 76% of teachers reported that the mini-observation model helped them to grow and improve student learning, while 100% of administrators believed the model was effective (See Figure 11).
Figure 11. To what degree did the mini observation helped teachers to grow and improve student learning?

There is uncertainty about whether there is data to support the claim that the mini observation model improves student achievement. Further research is warranted as one teacher wondered during the interview:

Well, I guess one of the things, I don’t know if it is so much of a challenge but one of my questions is this: What’s the impact on my student learning? Has this made a change or had an effect on the student learning? How could that be determined and how could it be assessed? I’m not sure I could even answer that at this point. I think there is a lot of things I’ve really liked about it [the mini observation model], but has it changed, has it had an impact on student learning? I’m not sure.

While uncertainty may exist as to the extent that mini observations affect student learning, anecdotal reports certainly suggest that teachers believe the practice did lead to improvement in the classroom due to the reflective nature of the process. One teacher stated, “I like the questions that arise [during the reflective conversations] and it sets me thinking differently on how to approach some kids.” Another teacher also praised the efficacy of the model when she stated, “Simply because of the reflective conversation afterward, it leads to additional teacher reflection, which in turn benefits student learning.” Another teacher observed, “Anytime a teacher is being truly reflective, student learning is benefited.” Finally, one teacher commented in a survey, “I was able to reflect and make changes to my teaching continuously throughout the school year. I looked forward to the observations knowing that they were brief and that I’d have time to discuss what was observed very soon after the lesson.”

Administrators agreed with the teachers’ perceptions about improved student achievement because of the quantity and quality of the mini observations and the subsequent conversations. One administrator noted, “Multiple professional conversations require teachers to reflect consistently on what is working in their classrooms and what is not.” Another
administrator shared, “[the model] provided a continuous, authentic structure for teacher reflection on teaching as it directly impacts student growth.”

While further research is necessary to specifically examine the correlation between mini observations and student achievement, results of this study suggested both teachers and administrators perceived that frequent classroom visits, followed by reflective conversations, led teachers to change practice with positive effects on students.

**Deeper thinking about practice.** Deeper thinking about practice was a consistent thematic response from individual interviews, surveys, and focus groups. Both teachers and administrators believed overwhelmingly that the reflective conversations after the mini observations caused teachers to “dig deep” and to think carefully about content and pedagogy. As one teacher commented, “this model really does kind of force the teacher into that mode of, ‘Well, what do you think about yourself?’”

Teachers reported they spend more time thinking about their students, what they are teaching, and asking, “Why is it important?” Teachers also reported being more focused on and reflective about their practice. This intensive reflection led to students having a clearer understanding about what was being taught and how it would help their learning. One teacher stated, “I think that that I’ve been more focused. And I think that’s just some of the little things [my principal] has made me think about or asked questions about. It kind of gets me thinking about my practice in that sense.” Administrators reported that because of reflective conversations, teachers are able to make changes and adjustments and constantly think about what is working and not working in their classrooms.

One teacher stated, “The mini observations have made me more reflective about my classroom—made me think about the ‘why’s’ of my classroom structure and routine.” Another
teacher commented, “Conversations with the evaluator are another critical piece for setting goals to enhance and enrich student growth and learning.” One administrator stated, “It allowed me to tap into my experience and skill as an evaluator, to ask questions, and to facilitate a reflective conversation. It allowed for the process to be shifted from judgment to inquiry, which in turn led to self-reflection by the teacher.”

Teachers reported the need to share the events that preceded the mini observations during the reflective conversation. It required them to “set the stage” and caused them to think about “sections” of their lesson plans in ways they had not thought about before. As one teacher remarked, “In order to unpack the 10 minutes the observer witnessed, it was necessary to think deeply about each of the parts of the lesson that either followed it or came before it.”

In sum, teachers’ and administrators’ responses to the mini observations illustrate that the act of self-reflection about practice is essential to improved instruction and pedagogy. As one teacher noted in the survey:

This process makes folks dig deeper and reflect on their role as a classroom teacher. . . . It does enable teachers to see their areas of strengths and weaknesses. Conversations with the evaluator are another critical piece for setting goals to enhance and enrich student growth and learning.

Research Question #3. What are teachers’ perceptions about the reflective feedback and its ability to lead them to modify their practice?

Both teachers and administrators agreed the reflective conversations after the mini observations provided an effective vehicle for teachers to improve their instructional practice. Teachers rated being observed and engaging in reflective conversations as the two top interventions that had the most impact on improving their teaching, while administrators rated having reflective conversations as the number one way to improve teaching (See Figure 12). Seventy six percent of the teachers believed that the feedback they received from an
administrator during a reflective conversation, which averaged 11 minutes, was sufficient to improve their practice (See Figure 13). When asked about the length of the reflective conversations, 80% of the teachers said it was “just right” (See Figure 14).

**Figure 12.** What has the most impact on improving teaching?

**Figure 13.** Was the feedback you received during the reflective conversation sufficient?
Figure 14. Do you think the length of the reflective conversation was too short, too long, or just right?

The primary themes that emerged from the data gleaned from the answers to research question #3 included the following:

1. Authentic conversations
2. Immediate feedback
3. Focused and rich conversations
4. Finding time to have conversations

**Authentic conversations.** Engaging in authentic conversations with the mini observation model emerged as an important theme for both teachers and administrators. In the reflective conversations, honest feedback was provided to the teacher in person (not in writing) and created the opportunity for teachers and administrators to engage in a two-way conversation. The conversations were authentic because they were “real,” “natural,” and “focused.” Several
teachers and administrators noted that the reflective conversation was not always focused solely on what the teacher needed to improve, but identified positive aspects of the observation. One teacher referred to the conversations as an opportunity to “problem-solve” in a very productive way, while another teacher said, “our reflective conversations were more honest and balanced.” One administrator stated, “I think you can have an authentic conversation that moves mountains in a way in which you truly cannot have in writing.” Another administrator conveyed that “The mini [observation] forced both evaluator and evaluatee to engage in very authentic conversations about instruction or other responsibilities appropriate to the evaluatee's position.”

**Immediate feedback.** Both teachers and administrators agreed that the immediate feedback played an important role in the mini observation model. Teachers frequently cited the benefit of hearing the feedback within a day or two, as opposed to several weeks later (common under the traditional observation model). Teachers viewed the feedback as an opportunity to change their practice much more quickly than they would have with the traditional model. One teacher expressed, “The immediate feedback enabled me to implement any changes I needed to make right away. The feedback was based on what was seen in that time and I could begin working on it immediately.” Another teacher, who appreciated receiving the feedback in a timely manner, stated, “I think that the consistent and immediate feedback helps me to change my practice more quickly than if we were sitting down a week or two later and a couple of things were pointed out.” A third teacher remarked that, “The immediate conversations after [the mini observations] were helpful because the lessons were fresh in my mind. I was also able to apply the advice given to me right away.”

Several administrators believed that the immediate feedback was a way for them to identify additional supports that a teacher could use to improve in a specific area. As one
administrator stated, “I think we are able to import supports for teachers sooner because they are revealing those needs for supports to us in a safe way.” In one case, an administrator described a teacher who was struggling to engage students in a meaningful way. Based on frequent observations and reflective conversations, the administrator enlisted the assistance of the curriculum specialist to provide support and model lessons that were of high interest and student centered. The teacher was able make the necessary adjustments to her planning and delivery of the curriculum.

**Focused and rich conversations.** Focused and rich conversations were identified as a consistent theme in this research study. Both teachers and administrators commented frequently about how rich and deep the reflective conversations had become, and how the conversations had “pushed their thinking.” A number of teachers and administrators noted that the conversations were always focused on teaching and learning. One administrator stated, “We have never had follow-up feedback or conversations that pushed our thinking in any other school [district] before we came here to Needham.” A different administrator commented, “I do think that it requires teachers to do a lot more of the thinking around what could I do to improve my practice rather than getting the suggestions from the administrator.” Several teachers remarked that the feedback they received from an administrator prompted them to have a conversation with a colleague about a similar topic. As one administrator said, “I think it’s a great way for cross pollination within the building.” A second administrator added, “Conversations are contagious and the conversation about teaching strategies is spreading across the staff. Staff not involved in the pilot are actually looking forward to doing it next year.”

Administrators complimented the ease of having focused reflective conversations and the discoveries that arose because of those conversations. One administrator commented, “He [the
observed teacher] started to transform the use of the walls in his room. In the old model, I would go in and saying ‘you need things on your walls’ but this change came from a reflective conversation and the feedback.”

Teachers also commented on the cultural shift from post-observation conferences in the old model to reflective conversations in the mini observation model. One teacher shared during the interview:

It was a good opportunity for [the principal] to say ‘is there anything else you want to tell me?’ and ‘Is there anything else going on that I need to know about?’ It was good just to have the time to talk to him. He would even make connections to other teachers to draw more information about what I was doing.

Another teacher said the administrator would ask questions like, “what prompted you to do this” and “What led to this” or “Where are you going from here.” “I feel like we were talking more about the craft of teaching than the actual the way a lesson went or what I exactly said.”

One administrator commented, “I think it is key having staff take more ownership for their professional growth and development and to be as a part of the conversation instead of, ‘here’s what I think about you and here’s my final up write up about you at work.’”

**Finding time to have conversations.** Every teacher and administrator interviewed, and some of the teachers and administrators who responded to the survey, indicated it was challenging to find the time to have the reflective conversations. This challenge was especially true for those administrators who traveled to different buildings to supervise teachers. As one teacher stated, “A challenge is figuring out a time to meet and converse, although I do think it is important to have feedback that is timely.”

**Research Question #4. Do teachers and administrators perceive any change in their relationship as a result of participating in mini observations?**
One significant focus of this research study was to evaluate the potential change in the relationship between the supervisor and subordinate, and to determine if the mini observation model of supervision and evaluation had any influence on improving that relationship. The vast majority of public schools in the Commonwealth of Massachusetts, like many schools in the nation, rely on the hierarchy of the principal acting as the instructional leader. To promote change and instructional improvement, a strong relationship built on mutual respect and trust is essential to any organization.

The three themes that emerged from the data pertaining to this research question included the following:

1. The development of stronger relationships
2. Increased comfort in taking risks
3. Increased trust between teachers and administrators

**The development of stronger relationships.** Developing stronger and deeper relationships between teachers and administrators was the most frequently identified theme in this research study. Both teachers and administrators overwhelming agreed that the mini observation model fostered a deeper professional relationship (See Figure 15).
Figure 15. The mini observation has helped to enhance the relationship between teachers and administrators.

The mini observation model provided many opportunities for structured interactions around student growth that facilitated the development of authentic, collaborative professional relationships. A teacher reported, “I feel like professionally we—our relationship—had taken a different turn in the sense of—we talk more professionally about the craft of teaching and that’s been really nice.” An administrator also commented, “With the mini observation model, the interpersonal relationship can grow more deeply face to face.”

Increased comfort in taking risks. Administrators reported that teachers were more comfortable taking risks in their teaching due to the trust that was built through frequent observations and the congeniality of the reflective conversations. Teachers asked administrators to visit classrooms when they were trying new things “without concern that it may not go well.” One administrator shared that, “The frequency of observations allows a teacher to experiment more. [They are] not fearful of the ‘one and done’ event.” Another administrator said,
“[Teachers] are very comfortable sharing the risks, sharing the failures.” One teacher reported during the interview:

I was able to reflect and make changes to my teaching continuously throughout the school year. I looked forward to the observations knowing that they were brief and that I'd have time to discuss what was observed very soon after the lesson. Often, I could ask for help with a group or a student by having my supervisor observe me.

One administrator observed that teachers were more willing to take risks with the mini observation model than the traditional model, “because they are very comfortable sharing the risks, sharing the failures. ‘I tried this, it didn’t work. I’m ready to pull my hair out’, whatever. [With the traditional model] It was about the pretty picture always.”

**Increased trust between teachers and administrators.** Administrators and teachers both reported that trust increased with the mini observation model. Given the increased frequency of observations and reflective conversations, professional relationships were strengthened between the evaluator and teacher, in turn leading to increased mutual trust.

According to one teacher, the evaluation process is not a “gotcha” moment but an opportunity to “see the good and the bad.” One administrator noted, “I believe there is more trust and expectations for constant conversation about teaching and learning.” Another administrator commented during the interview:

The idea is that there is trust and the coach [evaluator], at times, is going to push you to see if you can go farther, to guide you, and reflect back to you. You need to really have safe [conversations] and not walk away from hard conversations. The conversations, over a period of time, build trust. If I’m coming and questioning or asking, it is coming from a place they know is not from judgment, it’s let’s think outside of the box.

**Summary of Findings**

Forty seven teachers and administrators participated in the mini observation pilot conducted at the Haggerty Elementary School and Jonas Stone Sixth Grade Center. Of the 47 participants in the pilot, 13 teachers and administrators volunteered for in-depth interviews as
part of the researcher’s study on mini observations. Extensive data was collected through individual interviews of three teachers and three administrators, a 22-question survey given to all 47 participants, and three focus groups conducted over the course of a 9-month period. The collection of data from multiple sources provided the opportunity to understand how the participants perceived the mini observation model. This data will also assist in the decision-making process of whether or not to continue the model in the current schools and to expand its implementation to the remaining six schools in the district, including four elementary schools, the middle school, and Needham High School.

In summary, the data gathered from individual interviews, surveys, and focus groups indicated that:

1. Significantly stronger relationships were built between the teacher and administrator due to the frequency of classroom visits and the reflective conversations that quickly followed each observation;

2. The evaluation process for both teachers and administrators significantly improved the authenticity of supervision and evaluation. This improvement was accomplished primarily through the reflective professional conversations that were frequent, focused, real, and rich;

3. The mini observations were viewed by administrators and teachers as a low-stakes event where stress was significantly reduced due to the frequency of visits, collaborative reflective conversations, and the elimination of the “dog and pony show” where teachers were only observed a few times during the year;

4. The immediate feedback provided by the administrators after a mini observation was viewed as highly desirable as teachers could implement the recommendations
immediately and administrators could observe those changes during their next observation; and

5. The mini observation model and reflective conversations led teachers to think far more deeply and pedagogically about their practice. When asked, 92% of teachers and 100% of administrators indicated they preferred the mini observation model of evaluation to the traditional evaluation model.

As a result of its positive reception and results, the mini observation program will be implemented next year in all Massachusetts schools as required by the Department of Elementary and Secondary Education. The 5-step cycle in the Department’s model evaluation relies heavily on data gathered from mini observations conducted throughout the school year. This research project has been instrumental in understanding and “living” the process as the district prepares for full implementation in September 2013. As one administrator commented about the mini observation model, “I think this has been a great transformation.”

Chapter V: Discussion of Research Findings

This chapter will review the problem of practice and methodology for the study and includes the major findings of the research. Findings will also be presented in the contexts of theoretical framework and literature review. This presentation will be followed by final analysis and recommendations, significance of the study, limitations, issues of validity, conclusions, and finally, recommendations for future research.

Revisiting the Problem of Practice

This study explored teachers’ and administrators’ perceptions of a different approach to supervision and evaluation. For the last 30 years, public school teachers in Massachusetts, and throughout the nation, were evaluated using a system that relied on a small number of classroom
observations lasting approximately 45 minutes. Often referred to as dog and pony shows (Marshall, 2009), these observations were announced and allowed the teacher to create a special lesson plan that was unlikely to be indicative of typical behaviors, patterns, and routines in the classroom. Because these observation were conducted just two or three times per year by the evaluator, the stakes were extremely high for teachers and forced them to think, plan, and execute their lessons in a way that showcased their work for the purposes of evaluation. These infrequent observations were not representative of the teacher’s daily practices. Teachers and administrators agree that the traditional observation process has significant problems: the amount of time required for administrators to complete the process (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984); the limited number of classroom observations on which administrators base their perceptions and decisions about teacher performance (Marshall, 2009); the often rote-like quality of written feedback to teachers (Marshall, 2009); and a lack of evidence that the process improves professional practice or leads to improvement in student achievement (Wise et al., 1984).

This study explored a model developed by Kim Marshall (1996, 2005, 2006, 2009), which is based on increased classroom observations and reflective conversations with teachers. Through a series of unannounced mini observations, lasting no more than 10–15 minutes and conducted multiple times throughout the school year, the evaluator gathers a fuller understanding of student learning, the teacher’s strengths, and skills that require improvement. The purpose of this research was to investigate teachers’ and administrators’ perceptions of the relationship between mini observations and teacher performance, in order to determine what effect, if any, a system of mini observations has on teacher performance.
The use of mini observations has the potential to fundamentally change the meaning of teacher evaluation. It enables teachers and administrators to develop strong relationships, to engage in non-threatening, frequent, and collaborative professional conversations, and allows teachers to think more deeply about their practice.

**Review of the Methodology**

The purpose of this study was to understand how teachers and administrators perceived an evaluation system based on mini observations and reflective conversations conducted throughout the school year. This study was designed to specifically address the following research questions:

1. What is the experience of teachers and administrators who have been exposed to mini observations?
2. How does the mini observation model impact teacher’s instructional practices as perceived by both teachers and administrators?
3. What are teachers’ perceptions about the reflective feedback and its ability to lead them to modify their practice?
4. Do teachers and administrators perceive any change in their relationship as a result of participating in mini observations?

To answer these questions, a qualitative case study was designed and conducted with 47 teachers and administrators in a suburban school district in Massachusetts. Two schools were selected to participate in the study including a K–5 elementary school of 450 students and a middle school with 430 sixth-grade students. Data was collected through in-depth interviews with three teachers and three administrators, an anonymous online survey completed by 33 participants, and three focus groups conducted throughout the school year. Transcripts, survey
responses, and focus group transcripts were read carefully and coded according to common themes.

**Discussion of Major Findings**

Through careful data analysis involving first and second cycle coding (Saldaña, 2009) of 12 documents (in-depth interviews, survey data, and notes from three focus groups), a number of very clear themes emerged from the sources. The following themes were most salient as the data was coded and analyzed:

1. Significantly stronger relationships were built between the teacher and administrator due to the frequency of classroom visits and the reflective conversations that followed each observation.

2. The evaluation process for both teachers and administrators significantly improved the authenticity of supervision and evaluation. This improvement was accomplished primarily through the reflective professional conversations that were frequent, focused, real, and rich.

3. The mini observations were viewed both by administrators and teachers as a low-stakes event where stress was significantly reduced due to the frequency of visits, the collaborative reflective conversations, and the elimination of the “dog and pony show” where teachers were only observed a few times during the year.

4. The immediate feedback provided by the administrators after a mini observation was viewed as highly desirable as teachers could implement the recommendations immediately and administrators could observe those changes during their next observation.
The mini observation model and reflective conversations led teachers to think deliberately and deeply about practice, pedagogy, and student learning.

The mini observation process strengthened the relationships between teacher and administrator. A clear and consistent theme expressed by both teachers and administrators was the perception the mini observation model fostered deeper connections and enhanced relationships between the two groups. The frequent one-on-one, face-to-face reflective conversations were felt to be more collaborative than the traditional model because the teacher was more actively engaged in the conversation. Some teachers stated they felt comfortable disagreeing with administrators due to the feelings of trust that developed from frequent interactions facilitated by the mini observation model.

The teachers reported the mini observation model did not place undue stress or pressure on the supervisory relationship. Conversations were often non-confrontational with a focus on teaching and learning. Many administrators and teachers reported that the conversations felt more professional and focused on the craft of teaching. Structured interactions around student growth fostered the development of authentic, collaborative professional relationships.

In summation, it was reported that increased visits by the administrators encouraged nurturing relationships and contributed to a culture of risk taking that was condoned and not condemned. In turn, the administrator was viewed more as a coach with an interest in instructional improvement. As one administrator commented, “This nurtured a professional relationship between teacher and administrator. A relationship built around improvement.”

The professional conversations between teachers and administrators became more authentic. Teachers and administrators agreed effusively that the conversations about teaching and learning were far more authentic with the mini observation model. Because trust had been
built between the administrator and teacher, the conversations were real, direct, honest, focused, and brief. Unlike the traditional model, where the administrators tried to capture every detail in a 45-minute observation a few times a year, the mini observation reflective conversations were concentrated on a limited number of details. This level of authenticity led teachers to feel safer to try new things or to share areas of weakness or concern. Additionally, administrators felt they could provide direct feedback but felt it was less challenging to share than in the traditional evaluation model.

Authentic conversations and the direct feedback allowed teachers to implement recommendations more quickly. Administrators could then see the employed changes during the next observation. Administrators also overwhelmingly preferred verbal feedback to written feedback because it was immediate, unscripted, and genuine. As one administrator stated, “There’s a lot of clarification that can happen [verbally] that you can’t have in long, copious written reports. The feedback can be given right then and there at the reflective conversation. You can clear the air [and] the teacher is right there to ask the clarifying questions.”

The theme of authenticity extended beyond the verbal conversations. Because of the frequency of the unannounced visits, teachers believed, and administrators agreed, that observations were “real” and a “true picture” of what each day looked like in the particular classroom. Teachers were no longer given advanced notice of the observation so planning and execution could not be managed differently. One teacher commented, “I think the person who was observing me saw more of what my classroom looks like and how it works on a daily basis, not how it would work if everything was perfect.”

**Low stakes, lower stress.** Administrators and teachers both talked about the mini observation model as being a less intimidating approach to supervision and evaluation. Teachers
viewed the model as less formal and much more inclusive than the traditional observation model. Because of the frequency of classroom observations, the stakes were lower as teachers felt they had many more opportunities to make up for any mistakes in a future observation.

Conducting the reflective conversation soon after the observation was a significant relief to the teachers. Instead of waiting a week or more for a post-observation conference, teachers reported that improved timing of the reflective conversation allowed them to hear the “news” sooner rather than later, and allowed them to make rapid adjustments to their pedagogy as recommended by the administrator or as a result of thinking more deeply about their practice.

While teachers experienced a decrease in stress, many administrators experienced increased stress when trying to complete 12–15 mini observations. There were also stressful barriers associated with finding time for the reflective conversations. However, stress was significantly reduced because of the elimination of written reports that followed formal observations under the traditional observation model. Administrators reported feeling considerable relief due to the elimination of producing extensive written reports. Most administrators believe that these written reports summarizing the observations do not lead to improved teaching.

Teachers reported feeling less stress with the mini observation because having an administrator in the classroom become commonplace as opposed to an abnormal event. The mere frequency of the visits, coupled with the sense of trust, helped teachers to perceive the administrator more like a coach than an evaluator. As one teacher shared in the survey,

I just taught lessons day after day sort of organically as they would have happened anyway, which to me is a benefit because I don’t really think that the former method of observation is helpful . . . . No matter where you are at on the day of the observation, it is real life and you have to do your performance.
Another teacher reported feelings of lower stress when the administrator walked into the room. As she stated, “Because it wasn’t a planned visit, I didn’t stress out the night before. He just showed up and that was it, and so it really was a true reflection of me as a teacher in the classroom.”

Teachers also frequently reported feelings of lower stress at home either before a scheduled observation or while waiting for the post-observation conference. Speaking specifically about having a prompt reflective conversation, one teacher commented, “I mean the good thing was that you know, we could debrief . . . in 10 minutes, and it wasn’t like I was losing sleep the night before a lesson.”

**Immediate and frequent feedback.** Both teachers and administrators agreed that immediate feedback played an important role in the mini observation model. Teachers frequently cited the benefit of hearing feedback within a day or two of the mini observation, as opposed to several weeks later, which was common with the traditional observation model. One teacher expressed, “The immediate conversations after [the mini observation] were helpful because the lessons were fresh in my mind. I was also able to apply the advice given to me right away.”

While administrators found it a challenge to find the time for a reflective conversation, one stated, “I think that there is a direct impact in going in and giving feedback or going in and having a conversation and an immediate response . . . immediately transferring to what they are doing on a daily basis with kids.”

**Deeper thinking about practice.** Perhaps the most important finding in this research study was the discovery that teachers think more deeply about their practice with the mini observation model. Overwhelmingly, both teachers and administrators believed that the
reflective conversations after the mini observation caused teachers to think carefully about content and pedagogy. Teachers reported they spend more time thinking about their students, analyzing what they were teaching, and asking why the subject matter was important. Teachers also reported being more focused and reflective about their practice leading to the students having a clearer understanding about what was being taught and how it would help their learning. Administrators reported that, because of reflective conversations, teachers were able to make changes and adjustments and constantly think about what was working and not working in their classrooms.

This researcher has not searched the literature with regard to the relationship between deep reflection and a teacher’s overall effectiveness. However, my anecdotal accounts of hiring and supervising teachers as a principal and human resources director over the last 18 years has led me to believe that deep reflection is a common attribute of high performing and highly effective teachers. Constant introspection and personal reflection are practices that separate the very good from the exceptional teacher.

**Findings in Relation to the Theoretical Framework**

This study was informed by leadership and organizational theory, which provided the framework to evaluate the efficacy of mini observations. Leadership and organizational theory details how individuals and groups of people impact organizations.

**Evaluating effectiveness.** Schmoker (2006) proposed that the most important work in schools is what teachers are teaching students and how effectively they get students to learn. The mini observation model is an instrument focused on these two points. Teachers report that the model promotes a deeper reflection on what they are teaching, how they are teaching it, and why it matters. Many teachers reported in the in-depth interviews, surveys, and focus groups
that frequent visits by administrators and subsequent reflective conversations were unlike the traditional post-conference when they were primarily passive. The reflective conversations caused them to thoughtfully evaluate their practice and to question their craft unlike ever before. Teachers report making immediate changes based on the conversations and practicing the new learning.

DuFour (2002) presented the need for schools to view the principal as the learning leader and not just as an instructional leader. Principals must carefully evaluate student learning and not just teacher performance. In this new role, the principal can focus on a teacher’s instructional strategy as a way to improve student outcomes. Through this research study, it is clear that both administrators and teachers believe that the mini observations concentrate more on what students are learning. This concentration corresponds to a decrease in administrators having to watch and record teacher behavior (incredibly time-consuming activities). Administrators report significantly more interactions with the students during the 10–15 minute observations and are able to engage students in more substantial conversations prompted by questions such as “Tell me about what you are learning?”

**Improving instructional capacity.** Elmore (2008) described the need for teachers to operate less as independent practitioners and instead to increase collaboration to improve instructional capacity. Such collaboration will lead to improved outcomes for students. The mini observation model has led to deeper, more collaborative conversations between teachers and administrators. Teachers report that the frequent and immediate conversations help them institute changes in instructional approaches more quickly than they have done with the inefficient traditional evaluation model. In addition, both teachers and administrators have seen evidence of the notion of “cross pollination” as a result of the mini observations. “Cross
“pollinated” conversations are not just between administrators and teachers, but are more frequently between and among teachers. Like a bee, administrators witness exemplary practices in one classroom (or flower) and share those practices (pollinate) with other teachers more frequently than before. Administrators and teachers see the evaluator’s role as shifting from judging to coaching.

**Power shift.** Bush (2003) articulated that in the collegial model decision-making power shifts and is shared among various stakeholders. The three leadership models that promote collegiality include transformational leadership, participative leadership, and interpersonal leadership. These models focus on school outcomes, bringing staff together, and improved collaboration and interpersonal relationships (Bush, 2003).

The mini observation model has significantly affected the relationship between the teachers and administrators. A majority of teachers and administrators report that a more trusting, deeper relationship was cultivated due to the frequent visits to the classroom and the “real” conversations that took place during the reflective conversations.

**Leadership focus.** Leithwood (1994) identified four specific areas that leaders must focus on in schools: personal attention, helping others to solve problems in new ways, communicating high expectations, and modeling behavior. Participants report the mini observation model of evaluation provides a deep connection between administrator and teacher as well as a collaborative process for exchanging ideas and views on professional practice.

**Second-order change.** Marzano (2005) defined second order change as a “dramatic departure from the expected, both in defining a given problem and finding a solution” (pp. 65-66). The implementation of the mini observation model represented a dramatic shift from the traditional observation model, which had been in place in the Needham district for decades. The
collaborative effort of the administration and the teacher’s union was three years ahead of the Massachusetts Department of Elementary and Secondary Education’s efforts to bring meaningful change to the teacher evaluation process. Through a carefully crafted pilot with sound data collection methods to evaluate the outcomes, the Needham Public Schools is ahead of the curve in understanding the benefits and challenges of mini observations.

**Reframing.** Bolman and Deal (2008) discussed leadership in the context of reframing or “deliberately viewing a situation from multiple perspectives” (p. 37), and offered the human resource frame as one perspective: “Human resources theorists typically advocate openness, mutuality, listening, coaching, participation and empowerment” (p. 41). These qualities are at the heart of mini observations. Many teachers and administrators believed that the mutuality of reflective conversation was best displayed in real, genuine and focused discussions where both parties were responsible for the conversation. Listening was clearly on the minds of administrators, as one remarked, “I really like hearing teachers talk about their students and their thinking about their students and the learning.” The idea of coaching was clearly articulated by a number of teachers who believed that the administrator-teacher relationship had shifted from supervisor to coach. Finally, empowerment was a major theme as teachers reported a sense of control when they could immediately implement the recommendations of the administrator. In addition, teachers reported that they thought more deeply about their practice, their students’ learning, and often questioned the “whys” of their practice. With the mini observation model, the supervision process shifted from evaluation and documentation of performance to that of reflective conversation where the supervisor asks questions and listens, in an effort to help the teachers in their professional growth.
Bolman and Deal (2008) articulated that “*Human Resources Leaders Are Visible and Accessible*” (p. 41). The mini observation model required administrators to observe teachers 12–15 times per year, and follow up each classroom visit with a personal conversation with the teacher. A clear and consistent theme expressed by the teachers was how much they valued administrators in their classrooms, whether for validation of their work or to better understand students; teachers valued the time administrators were present in their classrooms.

**Making Sense.** Gallos (2008) noted that effective leaders make sense of everyday situations. She states, “Sense making involves three fundamental steps: noticing something, deciding what to make of it, and determining what to do about it” (p. 163). Marshall (2010) has reformulated a 30-year old tool and simplified the model to the most important aspects of supervision and evaluation. Marshall’s (2010) model focuses on what Schmoker (2006) identified as the heart and soul of our classrooms: What is being taught and how well it is taught. With a focus on frequent, brief classroom observations and focused reflective conversations with teachers, this alternative method has enhanced the professional growth of teachers.

**Discussion of the Findings in Relation to the Literature Review**

The literature review in this study examined the past 300 years of teacher supervision and evaluation, with a focus on the last three decades including accountability, public perception, teacher preparation, and a focus on standards. While much of the work on mini observations has been groundbreaking over the past four years, the findings from this study have a relevant connection to the work presented in chapter II. The following section will summarize the data in this study in relation to the last 30 years of teacher evaluation research.

**Accountability, public perception, and teacher preparation.** From 1980 to 1990, the job of teaching was subject to scrutiny from the public largely due to the publication of
documents that were widely disseminated to the public (A Nation at Risk, 1983; United States National Commission on Educational Excellence, 1983). With the notion that the American education system was broken, reforms for teacher preparation and accountability were demanded. Helping teachers grow and improve were the new ideas for the decade, and teacher evaluation through formative and summative assessments was growing (Harris, 1986; Darling-Hammond, 1988). Evaluation systems focused on objective standards and documentation. Wise et al. (1984) found that the most significant positive outcomes during this time period were the enhanced communication between administrators and teachers and improvement in teachers’ instructional skills and feelings of less isolation. In this current research study, communication between administrators and teachers was enhanced considerably and most participants shared that a stronger relationship had developed as a result of frequent mini observations and reflective conversations that focused on improvement. Teachers believed their instructional skills had improved as they began to think more deeply about what they were teaching and why. They were able to make instructional improvements more quickly as feedback was regular and focused. Teachers valued having administrators in their classrooms, and teachers shared the results of reflective conversations with each other. One administrator described herself as a “cross-pollinator” because she was able to share insights from one classroom with teachers in other classrooms.

In 1986, Harris recommended an eight-step approach to teacher evaluation because existing models provided very little feedback to the teacher. A consistent theme of this mini observation research was the ongoing, focused, and regular communication about teaching and learning that “lifted the learning” for many students. Harris and Hill (1982) strongly recommended formative assessments as part of the evaluation process. Teachers should learn
about both their strengths and weaknesses. The current research found that reflective conversations revolved around the positive aspects of the teachers’ practices, enhancing their understanding of other teachers’ pedagogies and curriculum, and identifying weaknesses to help teachers to improve their craft.

Manatt (1982) and Stiggins and Duke (1988) advocated for an evaluation model with frequent observations and increased professional conversations with teachers. Manatt (1982) believed that the conversations with the teacher held after an observation wielded the most impact on teacher improvement. Based on the data from in-depth interviews, surveys, and focus groups, mini observation and reflective conversations were identified as having a positive impact on teaching.

**Focus on standards.** Over the past twenty years, a focus on standards proliferated in the literature related to curriculum and teacher evaluation. Standards-based teacher evaluations were often complex and complicated. Danielson’s (1996) model included 65 different elements across four domains of teaching. The domains included planning and preparation, the classroom environment, instruction, and professional responsibilities. Research has shown that teacher evaluation models built on standards have resulted in improved student achievement (Taylor & Tyler, 2011; Darling-Hammond, Amrein-Beardsley, Hartel & Rothstein, 2012). In 2005, Tucker and Stronge’s model of standards-based evaluation linked teacher performance to student performance.

**Professional conversations.** In 2000, Danielson and McGreal developed teaching standards focused on strengths, areas needing improvement, helping teachers reflect on their practice, and professional conversations. Danielson (2009) continued to state that professional conversations about practice are the most powerful approach to improve student learning. When
teachers are not passive in the evaluation process and must participate collaboratively, student
learning improves. Teachers must engage in the analysis of classroom practices and own the
instructional decisions they make. Papay (2012) advocated for the use of professional
conversations in the context of professional development to help teachers focus on improving
student learning by “driving instructional change” (p. 133).

The central focus of this research has been on the efficacy of reflective conversations.
Through careful data collection and analysis, it is clear than mini observations followed by
reflective conversations support the findings of both Danielson (2009) and Papay (2012).

**Shared expertise.** Nolan and Hover (2010) suggest that neither teachers nor
administrators hold all of the knowledge, and each person contributes an understanding and
expertise to content and pedagogy. Data collected for this research suggests that both
administrators and teachers were engaged in a collaborative process because a relationship had
been built through the professional conversations. Teachers and administrators also commented
on their belief that mini observation was more like coaching than evaluation, because the
conversations were collaborative and engaging and focused on teaching and learning.

**Mini observations.** The use of mini observations as a form of teacher evaluation has
and Milanowski (2011) advocated for the use of short, focused, and frequent observations as a
observations are an effective method to improve teaching and learning. Based on the results of
this study, mini observations appear to be a powerful method to help teachers to grow in their
practice. However, this research was not able to measure the effects of mini observations on
student learning. Future research will need to address this issue, likely through quantitative analysis.

**Final Analysis**

This research attempts to gain a better understanding of how teachers and administrators perceive the relationship between mini observations and improved instructional practice in the classroom. The researcher conducted a qualitative case study to find answers to four research questions: (1) What is the experience of teachers and administrators who have been exposed to mini observations? (2) How does the mini observation model impact a teacher’s instructional practices as perceived by teachers and administrators? (3) What are teachers’ perceptions about the reflective feedback and its ability to lead them to modify their practice? (4) Do teachers and administrators perceive any change in their relationship as a result of participating in mini observations? The research questions guided this study, and data was collected through in-depth interviews, a survey with 33 teachers and administrators responding to 21 questions, and three focus groups conducted over the course of the school year. For this research study, data was coded and categorized to help develop comparisons (Maxwell, 2005) and interpret meaning.

Overall, the results of this study show that the participants in the mini observation model of evaluation perceived this approach to teacher evaluation as more effective than the traditional model of evaluation. Using the mini observation as a model of teacher supervision and evaluation, participants report developing significantly stronger relationships between the teacher and administrator, an increase in authentic conversations, lowered stress, an increase in collaborative reflective conversations, the value of immediate of feedback, and thinking more deeply about their practice. While there were some negative responses to the model, such as the
difficulty for administrators to complete 12–15 observations, the overall results demonstrate a significant positive perception of the model.

**Significance of the Study**

Teacher evaluation in the United States is undergoing significant reform due the Federal Race to the Top program. Evaluation systems must provide ratings of teacher performance, and student growth is required to inform those ratings. Decisions about compensation, promotion, removal, tenure, and relicensing will be part of the teacher evaluation system. As a Race to the Top State, Massachusetts has developed a comprehensive five-step cycle of evaluation that includes classroom mini observations.

This study was conducted to gain a fuller understanding of how a system of mini observations affects both teachers’ and administrators’ practices in the classroom. While this small study was conducted in a suburban, middle class school district, the information learned may be similar to the results in similar communities within the Commonwealth of Massachusetts, and perhaps the nation.

The data was conclusive that teachers and administrators prefer a system of mini observations as compared to a traditional evaluation model. Through this study, the researcher learned that mini observations deepen the relationships between teachers and administrators, lowers the stress of both evaluator and teacher, increases the focus on student achievement through authentic conversations, and deepens the teacher’s thinking about content and pedagogy. While the outcomes are positive and may have implications for the future, this study did not intend to measure how the outcomes affect student achievement. Further study is required to determine if student achievement is different in schools and districts using the mini observation model.
This study will be helpful to those Massachusetts school districts, administrators, and teachers who are implementing mini observations as part of the new evaluation model developed by the Department of Elementary and Secondary Education. An important finding from this study includes the challenges administrators face in implementing 12–15 mini observations for each teacher and engaging in a reflective conversation within 48 hours of the observation. It will be important for administrators to develop a schedule from the beginning of the year that allows for several mini observations each day. Secondly, both administrators and teachers found it challenging to find the time to engage in the reflective conversation. Careful planning is advised to ensure those challenges do not become handicaps. The entire model is developed on the premise that frequent observations and immediate reflective conversations leads to teacher improvement.

Limitations

The ability to generalize the findings of this study will be limited, based on the small number of participants (Mears, 2009). Yin (2009) stated, “case studies, like experiments, are generalizable to theoretical propositions and not populations or universes” (p.15). While wide generalizations cannot be applied from this study, the results are very promising given the teachers’ and administrators’ positive perceptions of mini observations. Ninety two of the teachers and 100% of the administrators stated they preferred mini observations to the traditional evaluation model, largely because the model led to deeper reflections about practice. These findings may lead other researchers to investigate mini observations across a wider school population as well as the question: “Do students who have teachers participating in a mini observation model of supervision and evaluation achieve at higher levels than the students of
teachers who are part of a traditional evaluation?” While teachers’ and administrators’ perceptions are important, the real issue is how the model affects students and their learning.

Validity

Throughout the course of this research study, the researcher relied on protocols and procedures developed by prominent qualitative case study researchers including Yin (2009), Merriam (2009), Maxwell (2005), Mears (2009), and Creswell (2007, 2009). These protocols and procedures included the following:

- Multiple case design provides the opportunity for replication while conclusions are derived from different sources (Yin, 2009).
- Inclusion of multiple cases to expand the possibility of wider generalization and enhancing external validity (Merriam, 2009).
- Strong ethical conduct of the researcher (Merriam, 2009).
- Collection of solid evidence to mitigate possible concerns with validity (Maxwell, 2005).
- Consistency of design and process that link the stories together (Mears, 2009).
- Allowing the participants to review the transcripts for accuracy (Mears, 2009). For this research study, each participant was emailed a copy of the recorded transcript and encouraged to provide feedback on accuracy.
- The strength of the relationship developed between the researcher and participant and using multiple sources of data (Creswell, 2007, 2009).
- Openly disclosing the researcher bias (Creswell, 2007, 2009).
- Maintaining objectivity when conducting the interviews and not leading the participants (Mears, 2009).
• Maintaining a “chain of evidence” (Yin, 2009).

• Understanding the “power differentials” as the researcher also held a position of authority in the school district, as Director of Human Resources (Mears, 2009). In addition, the participants were aware that this study was not only for research purposes, but also for the district’s desire to develop and pilot a new model for teacher evaluation.

• For the purposes of this study, the triangulation of data reduced the “potential problems of construct validity . . . because multiple sources of evidence provide multiple measures of the same phenomenon” (Yin, 2009, pp. 116-117).

• A “case study database” (Yin, 2009, p. 119) was developed to secure and record interview transcriptions, notes, survey responses, and focus group records. All data resides in files on the researcher’s computer and is backed up using four external resources. Because the researcher was able to access the data in the case study database, the “reliability of the entire case study . . . markedly increases” (Yin, 2009, p. 119).

Conclusions

This study involves 43 teachers and administrators in the Needham Public Schools and is viewed as having positive results based on teachers’ and administrators’ perceptions of mini observations. The data consistently illustrates that the experience was very positive, and that both teachers and administrators believe the mini observation model offers far more benefits than the traditional model of supervision and evaluation. As stated by Marshall (2009):

The ultimate goal of supervisory feedback, whether it’s the principal’s mini observations, one-to-one conferences, or a more extensive visit by a coach or peer, is to install a supervisory voice in the teachers’ heads and foster an acute consciousness of whether students are learning what’s being taught. Achievement will really soar when individual
teachers and teacher teams are constantly puzzling and theorizing and debating about how students are responding and how teaching can be improved (p. 84).

Future Research

An abundant amount of literature exists on teacher supervision and evaluation, particularly since the 1980s. In the last two decades, the literature has focused primarily on standards-based teacher evaluation. This approach to supervision and evaluation has provided a framework for teachers to understand better what a high performing teacher looks like and how they can become one. Through the use of rubrics, standards, and indicators, teachers have—for the first time in the history of teaching—a clear indication of how to improve their instructional practice. Benchmarks, in all domains of teaching, have been established that clearly determine whether a teacher’s practice is unsatisfactory, improving, proficient, or exemplary.

Further study of the mini observation model should be conducted in school districts that have implemented the mini observation model, including the Westwood, Groton-Dunstable, Sudbury, and North Andover school districts. Additional research should include teachers and administrators in kindergarten through grade 12. In addition, research should be conducted in urban and rural school districts.

Further research should also explore the relationship between mini observations and teacher retention and the degree to which the major finding of the current study impacts a teacher’s decision to remain in a district or in the field of teaching. Additional studies should also explore the impact of mini observations when teachers are supervised by reluctant or unenthusiastic administrators. In addition, the use of technology for the purposes of organization and efficiency should be explored to aid in the implementation of the mini observation model.
Finally, future research must focus on achievement outcomes for students. For this research to be of any relevance, a positive correlation must be identified to show that students who are taught by teachers who are supervised using a mini observation model achieve at higher levels. The case study design could be replicated by interviewing students who have teachers supervised using mini observations compared to students whose teachers are evaluated using the traditional model.
References


Massachusetts Education Reform Act (1993). Massachusetts General Laws, Chapter 69, Section 1J, Chapter 71, Section 41, 42.


MAXQDA, software for qualitative data analysis, 1989-2012, VERBI Software - Consult -
Sozialforschung GmbH, Berlin, Germany.


CA: Sage.


Appendix A

Invitation to Participate

Dear Hillside and High Rock School Faculty,

As you know, the Needham Public Schools will be piloting the mini-observation teacher evaluation program during the 2011-2012 school year. I hope that you have learned about the model through the training conducted by Kim Marshall as well as having conversations with each other about the protocol.

Currently, I am a doctoral student at Northeastern University. I have decided to study the mini observation teacher evaluation model for my dissertation. Specifically, I want to learn about the teacher and administrator's perceptions about the model.

I will be interviewing ten teachers and administrators for this research project. I will be asking questions such as "Tell me about your experience being observed using the mini observation model" and "Do you think the face-to-face conversation after the mini observation caused you to reflect more or less on your teaching than the traditional post observation conference?" There are no correct or incorrect answers to the questions I will ask, and your responses will be kept confidential.

All of the interviews will take place at a time and place that is convenient for you. I will begin interviews in October and continue through March 2012. I will use a device such as an ipad, laptop or other equipment so that I the responses can be transcribed later. I will use all staff responses as part of my dissertation, but I will not use names or identify them in any way.

If at anytime you wish to withdraw from the study, I will end your participation immediately.

I would like to meet with interested participants to explain the research in more depth, and to answer any questions may have. To that end, I would like to invite you to meet with me on either Tuesday, Nov. 1 at the High Rock School #1 2:35 pm and Wednesday, Nov. 2 at the Hillside School at 3:00 pm in the library.

I appreciate your consideration and support.

Sincerely,

Thomas F. Campbell
Director of Human Resources
Appendix B

Informed consent

Northeastern University, College of Professional Studies
Principal Investigator: Claire Jackson, Student Researcher: Tom Campbell
Title of Project: Teacher Supervision and Evaluation: Investigating the Efficacy of Mini Observations

Informed Consent to Participate in a Research Study

I am inviting you to take part in a research study. This form will tell you about the study, but the researcher (Tom Campbell) will explain it to you first. You may ask this person any questions that you have. When you are ready to make a decision, you may tell the researcher if you want to participate or not. You do not have to participate if you do not want to. If you decide to participate, the researcher will ask you to sign this statement and will give you a copy to keep.

Why am I being asked to take part in this research study?

I am asking you to be in this study because you will be part of the teacher supervision and evaluation program pilot at the High Rock and Hillside schools using mini observations (as opposed to the traditional model.)

Why is this research study being done?

The purpose of this research is to better understand if the mini observation model is more effective in helping you to improve as a teacher and think more reflectively about your practice, leading to improved student achievement.

What will I be asked to do?

If you decide to take part in this study, I will ask you to meet with the researcher several times during the school year for an in-depth interview. A mechanical device will record the interview, such as an ipad or laptop computer so that the researcher can keep an accurate account and details about your responses. Approximately ten teachers and administrators will be interviewed and asked the same questions. At the conclusion of the interview process, the researcher will code all of the responses and group them by themes. A final report will be written in the form of a dissertation explaining what was learned about mini observations based on the information you provided in the interviews.

If there are more than 10 teachers and administrators who would like to participate in the study, participants will be randomly selected but ensuring there are equal numbers of teacher with and without professional teacher status.

Where will this take place and how much of my time will it take?

You will be interviewed in your classroom or office or at a time and place that is convenient for you. The interview will take about one hour. It is possible that you may be interviewed two or three times from September 2011 through March 2012.

Will there be any risk or discomfort to me?

APPROVED

NU IRB# ______________
VALID ____________
THROUGH ____________
The only known risk to you may be fatigue experienced during the interviews. At any time, the interviews can be stopped and rescheduled at a different time. It is important to remember that your decision to participate or not to will have no bearing on your employment.

Will I benefit by being in this research?

There will be no direct benefit to you for taking part in the study. However, the information learned from this study may help the Needham Public Schools and other school districts to better understand the strengths and weaknesses of using mini observations as part of the teacher supervision and evaluation process.

Who will see the information about me?

Your part in this study will be confidential. Only the researcher on this study will see the information about you. No reports or publications will use information that can identify you in any way. It is expected that a second person will transcribe your recorded interviews into a word document solely for the purpose of the researcher to read, code and categorize the information. The transcriber will sign a statement of confidentiality and will be bound to maintain the transcriptions completely confidential.

Your personal information will be protected at all times and the researcher will not disclose who participated in the study or individual responses at any time. The interview data will be used to help understand if mini observations provide any advantages over Needham's traditional evaluation system. The interview data will be transcribed and coded for the purpose of summarizing your opinions and perceptions. Johnny Saldana wrote a book entitled The Coding Manual for Qualitative Researchers (2009) in which he describes coding as "a word or phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data...just as a title represents and captures a book or film or poem's primary content and essence, so does a code represent and capture... [the interview's] primary content and essence" (p. 3).

The data will be maintained in a locked file cabinet should the data need to re-analyzed or used in a future study.

Confidentiality will only be compromised if issues of safety are of concern. The researcher in his role in the Needham Public Schools is a mandated reporter under MGL 119, section 51A and will report suspected child abuse or neglect. 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. I will only permit people who are authorized by organizations such as Northeastern University to see this information.

Can I stop my participation in this study?

Your participation in this research is completely voluntary. You do not have to participate if you do not want to. Even if you begin the study, you may quit at any time. If you do not participate or
if you decide to quit, you will not lose any rights, benefits, or services that you would otherwise have as an employee.

Who can I contact if I have questions or problems?

Please contact Thomas Campbell at 781-455-0400 ext. 208 for any questions or problems.

Who can I contact about my rights as a participant?

If you have any questions about your rights as a participant, you may contact Nan C. Regina, Director, Human Subject Research Protection, 960 Renaissance Park, Northeastern University Boston, MA 02115 tel. 617-373-7570, email: irb@neu.edu. You may call anonymously if you wish.

Will I be paid for my participation?

No payment is offered to participants.

Will it cost me anything to participate?

There is no cost to participate in this study.

I agree to take part in this research.

Signature of person agreeing to take part

Date

Printed name of person above

Signature of person who explained the study to the participant above and obtained consent

Date

Printed name of person above

APPROVED

NU IRB # 11-043
VALID: 6/21/11
THROUGH: 6/21/12
Appendix C

Online survey for teachers

Dear Haggerty and Jonas Stone Teachers,

This survey is designed to help us understand your opinion about the mini observation model piloted at Haggerty and Jonas Stone this year. Please take some time to share your insight and ideas to help us make a decision about the pilot for the next school year. Your responses are confidential and your name will not appear anywhere on the results.

* Required

1. Please select the number of years you have been in Needham. *
   a. First year
   b. Second year
   c. Third year
   d. I have professional status

2. To what degree did the mini observation model help you as a teacher to grow and improve student learning?
   a. Five-point Likert scale from not at all to significantly

3. To what degree did the traditional model help you to grow and improve student learning?
   a. Five-point Likert scale from not at all to significantly

4. How many mini observations do you feel are necessary to have an impact on teaching? *

5. What has the most impact on improving your teaching?
   a. Coaching
   b. Peer observations
   c. Reflective conversations
   d. Mini observation and short debrief
   e. Traditional 45 minute observation
   f. Written reports after observation
   g. Written final summative evaluations

6. Describe how the mini observations helped you to grow and improve your teaching.

7. The mini observation model has helped to enhance the relationship between teachers and administrators *
   a. Five-point Likert scale from strongly disagree to strongly agree

8. Should goal setting be part of the mini observation process? *
   a. Yes
   b. No

9. Was the feedback you received during the reflective conversation sufficient?
   a. Five-point Likert scale from not at all to yes, definitely

10. On average, how many minutes did the reflective conversation take? *

11. Compared to the traditional observation model, was your level of anxiety more, less or about the same with the mini observation model?

12. In what ways has the mini observation model impacted the teaching and learning culture in your school? *

13. Do you prefer the mini observation model or the traditional evaluation model? *
   a. Mini observation model
b. Traditional model  
c. No preference  

14. What changes would you recommend if the mini observation model were to continue? *  

15. Please rate your overall satisfaction of the mini observation model *  
   a. Five-point Likert scale from not satisfied at all to very satisfied  

16. Please rate your satisfaction with the traditional evaluation model  
   a. Five-point Likert scale from not satisfied at all to very satisfied  

17. Using the mini observation model, my evaluator has good understanding of my classroom culture  
   a. Five-point Likert scale from strongly disagree to strongly agree  

18. When is the best time for teachers to engage in the reflective conversations?  
   a. 15 minutes before school (contractual time)  
   b. Before school (non-contractual time)  
   c. During my prep time  
   d. In the hallway between classes or lunch  
   e. During lunch  
   f. 30 minutes after school (contractual time)  
   g. After school (non-contractual time)  
   h. Other:  

19. Do you think the length of the reflective conversation was too short, too long or just right? *  
   a. Too short  
   b. Too long  
   c. Just right  

20. Mid-way through the year and at the end of the year, a rubric was used to rate performance. What aspects of the rubric were positive, and what aspects were challenging? *  

21. Do you think this process benefits student learning? How do you know? *  

22. Please think about the rubric used mid-way through the year and at the end of the year. The rubric was: *  
   a. Very valuable to me as a teacher  
   b. Not valuable  
   c. Contained too many indicators  
   d. The number of indicators was appropriate  
   e. Other:  

**Online survey for administrators**  

Dear Haggerty and Jonas Stone Administrators,  

This survey is designed to help us understand your opinion about the mini observation model piloted at Haggerty and Jonas Stone this year. Please take some time to share your insight and ideas to help us make a decision about the pilot for the next school year. Your responses are confidential and your name will not appear anywhere on the results.  
* Required  

1. Please select the number of years you have been in Needham *  
   a. First year  
   b. Second year
c. Third year
d. Four years or more
2. In your opinion, to what degree do you think the mini observation model helped teachers to grow and improve student learning? *
   a. Five-point Likert scale from not at all to significantly
3. In your opinion, to what degree did the traditional model help teachers to grow and improve student learning? *
   a. Five-point Likert scale from not at all to significantly
4. How many mini observations do you feel are necessary to have an impact on teaching and learning? *
5. What has the most impact on improving your teaching?
   a. Coaching
   b. Peer observations
   c. Reflective conversations
   d. Mini observation and short debrief
   e. Traditional 45 minute observation
   f. Written reports after observation
   g. Written final summative evaluations
   h. Describe how the mini observations helped teachers to grow and improve their teaching *
6. The mini observation model has helped to enhance the relationship between teachers and administrators *
   a. Five-point Likert scale from strongly disagree to strongly agree
7. Should goal setting be part of the mini observation process? *
   a. Yes
   b. No
8. Do you think that teachers believe the feedback you gave during the reflective conversation was sufficient? *
   a. Five-point Likert scale from not at all to yes, definitely
9. On average, how many minutes did the reflective conversation take? *
10. Compared to the traditional observation model, do you think that the teacher's level of anxiety was more, less or about the same with the mini observation model? *
11. In what ways has the mini observation model impacted the teaching and learning culture in your school or department? *
12. Do you prefer the mini observation model or the traditional evaluation model? *
   a. Mini observation model
   b. Traditional model
   c. No preference
13. What changes would you recommend if the mini observation model were to continue? *
14. Please rate your overall satisfaction of the mini observation model *
   a. Five-point Likert scale from not satisfied at all to very satisfied
15. Please rate your satisfaction with the traditional evaluation model *
   a. Five-point Likert scale from not satisfied at all to very satisfied
16. The mini observation model gave me a good understanding of each teacher's classroom culture *
   a. Five-point Likert scale from strongly disagree to strongly agree
17. As an evaluator, how many hours per week on average did you devote to the mini observation process? *
   a. 1-3 hours
   b. 4-7 hours
   c. 8-10 hours
   d. 11-13 hours
   e. 14-16 hours
   f. 16+ hours

18. When is the best time for teachers to engage in the reflective conversations? *
   a. 15 minutes before school (contractual time)
   b. Before school (non-contractual time)
   c. During my prep time
   d. In the hallway between classes or lunch
   e. During lunch
   f. 30 minutes after school (contractual time)
   g. After school (non-contractual time)
   h. Other:

19. Do you think the length of the reflective conversation was too short, too long or just right? *
   a. Too short
   b. Too long
   c. Just right

20. Mid-way through the year and at the end of the year, a rubric was used to rate performance. What aspects of the rubric were positive, and what aspects were challenging? *

21. Do you think this process benefits student learning? How do you know? *

22. Please think about the rubric used mid-way through the year and at the end of the year. The rubric was: *
   a. Very valuable to me as an administrator
   b. Not valuable
   c. Contained too many indicators
   d. The number of indicators was appropriate
   e. Other:
Appendix D

Letter of Approval to use Data

May 31, 2012

Nan C. Regina, Director
Human Subject Research Protection
960 Renaissance Park
Northeastern University
Boston, MA 02115-5000

Dear Ms. Regina,

I am writing to support the research proposed by Mr. Thomas F. Campbell, doctoral student at Northeastern University entitled Teacher Supervision and Evaluation: A Case Study of Administrators' and Teachers' Perceptions of Mini Observations. I originally reviewed Mr. Campbell's Application for Approval for human subject research with the Institutional Review Board on April 19, 2011.

Please be advised that Mr. Campbell has permission to conduct an additional online survey of teachers and administrators using Google Survey. I have reviewed the survey questions and approve the survey to be used as part of the research project as outlined in the Application for Approval at the High Rock and Hillside Schools in the Needham Public Schools district.

Please free to contact me at 781-455-0400 extension 203 if you have any questions or require additional information.

Sincerely,

[Signature]

Daniel Gulakanski, Ed.D.
Superintendent of Schools

Needham Public Schools
Office of the Superintendent
A school and community partnership that:
creates excited learners • inspires excellence • fosters integrity.

1330 Highland Avenue, Needham, MA 02492
Telephone 781-455-0400 extension 203
Appendix E

Interview questions for teachers

1. How did the mini observation process affect your practice in the classroom?
2. Describe your experience with the mini observation process,
3. What are the strengths and weaknesses of the mini observation process?
4. Describe how you would compare the mini observation process to the traditional evaluation.
5. Do you believe the mini observations improved your instructional practice as a teacher?
6. Describe your experience with the old evaluation system (two 45 minute observations).
7. How did participation in the mini observations affect your teaching?
8. What are the benefits of the mini observation model?
9. What are the disadvantages of the mini observation model?
10. Tell me about your perceptions of the 10-minute observation?
11. How many mini observations did your supervisor complete?
12. Tell me about your experience with the reflective conversations.
13. How were the reflective conversations different from the old post conferences?
14. Tell me what it was like not to have a written report after an observation was conducted.
15. Did you find value in the post observation reports that your supervisor provided after a classroom observation in the old system?
16. Tell me about how you felt having only unannounced observations.
17. Do you prefer mini observations or full class observations?
18. Do the mini observations allow your supervisor to see more or less of your overall job responsibilities than the traditional observations?
19. Tell me about the use of rubrics as your summative evaluation

Interview questions for administrators

1. How did the mini observation process affect your practice as a supervisor/evaluator?
2. Describe your experience with the mini observation process
3. What are the strengths and weaknesses of the mini observation process?
4. Describe how you would compare the mini observation process to the traditional evaluation.
5. Do you believe the mini observations improved the teacher’s instructional practice?
6. Describe your experience with the old evaluation system (two 45 minute observations).
7. How did participation in the mini observations affect your role as an evaluator?
8. What are the benefits of the mini observation model?
9. What are the disadvantages of the mini observation model?
10. Tell me about your perceptions of the 10-minute observation?
11. How many mini observations did you complete?
12. Tell me about your experience with the reflective conversations.
13. How were the reflective conversations different from the old post conferences?
14. Tell me what it was like not to have a written report after an observation was conducted.
15. Did you find value in the post observation reports that you provided after a classroom observation in the old system?
16. Tell me about how you felt having only unannounced observations.
17. Do you prefer mini observations or full class observations?
18. Did the mini observations allow you to see more or less of the teacher’s overall job responsibilities than the traditional observations?
19. Tell me about the use of rubrics as your summative evaluation.
Appendix F

Focus Group January 13, 2010

Welcome and Purpose
Welcome to the teacher evaluation focus group. The Evaluation Study Committee comprised of administrators and teachers are working collaboratively to improve the teacher evaluation process with a focus on professional growth. This focus group will help the Evaluation Study Committee to gather information to better understand your perceptions of the teacher evaluation process and instrument.

The Evaluation Study Committee
Tom Campbell, Director of Human Resources
Laurie Sullivan, Jonas Stone Teacher
Merle Berman, Pollard House Administrator
Raylene Roberts, Haggerty Teacher
Michael Kascak, Haggerty Principal
Sherrill Neilson, High School Teacher
Jonathan Pizzi, High School Principal

Format
Working in small mixed groups, please consider the questions below. You will have approximately 7 minutes per question, so your group will need to be focused. Please choose a recorder to write responses on the large paper. A member of the Evaluation Study Committee will sit in on your group as an observer and to answer questions.

Your honest and candid responses are appreciated and will be collated with the other groups. Your anonymity will be protected.

Questions

1. What are the strengths and benefits of our current teacher evaluation system?
2. What are the weaknesses of our current system?
3. What suggestions and recommendations can you offer to improve our teacher evaluation system with a focus on professional growth?
Focus Group April 5, 2012

1. Welcome and introductions
2. Brief summary of the mini observation pilot
3. Working in small groups, discuss the strengths and challenges of the mini observation model
4. Develop questions for a survey for all teachers and administrators
5. Brief overview of the DESE model plan for educator evaluation
6. Planning Ahead: Recommendations for September
   a. Where do we go from here with the pilot?

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Don</td>
<td>Mark</td>
<td>Jamie</td>
<td>Dan</td>
</tr>
<tr>
<td>Kathy</td>
<td>Kris</td>
<td>Millie</td>
<td>Abigail</td>
</tr>
<tr>
<td>Shannon</td>
<td>Leslie</td>
<td>Karla</td>
<td>Darlene</td>
</tr>
<tr>
<td>Abbey</td>
<td>Mickey</td>
<td>Jack</td>
<td>Edda</td>
</tr>
<tr>
<td>Missy</td>
<td>Mal</td>
<td>Rana</td>
<td>Jimmy</td>
</tr>
<tr>
<td>Tara</td>
<td>Susie</td>
<td>Davio</td>
<td>Bonnie</td>
</tr>
<tr>
<td>Tabitha</td>
<td></td>
<td>Darren</td>
<td>Jeb</td>
</tr>
</tbody>
</table>