Urban Secondary Teachers' Perceptions of a Standards-Based Teacher Evaluation System

Author: Jonathan David Pizzi

Persistent link: http://hdl.handle.net/2345/664

This work is posted on eScholarship@BC,
Boston College University Libraries.

BOSTON COLLEGE

Lynch Graduate School of Education

Department of Educational Administration

And Higher Education

School Leadership

URBAN SECONDARY TEACHERS' PERCEPTIONS OF A

STANDARDS-BASED TEACHER EVALUATION SYSTEM

Dissertation

by

JONATHAN D. PIZZI

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

May, 2009

ABSTRACT

URBAN SECONDARY TEACHERS' PERCEPTIONS OF A STANDARDS-BASED

TEACHER EVALUATION SYSTEM

Author: Jonathan David Pizzi

Advisor: Irwin Blumer, Ed.D.

This study was conducted within a large urban high school in the northeastern United States. This school resides in a district that utilizes a standards-based system of teacher evaluation. The purpose of the study was to determine teachers' perspectives on the effectiveness of standards-based teacher evaluation systems as a means of improving instruction and student achievement. The question for research, "What are teacher perceptions of evaluation in an urban high school that uses a standards-based evaluation system?" served as the framework for development and implementation of this study and for interpretation of the study findings. The body of related literature provided the theoretical rationale for the data collection, analysis, and interpretation. Major sections of the literature review included: a historical perspective on performance standards and of teacher evaluation from the mid-nineteenth through the early twenty-first centuries; an examination of contemporary teacher evaluation systems; a review of recent theoretical and empirical research concerning best evaluation practices; and a review of the research regarding teachers' perceptions of supervision and evaluation. Data were gathered from thirty participants through a thirty-one question Likert-based attitude scale. Key findings included teachers' perceptions that: they were not familiar with the teaching performance standards used in their district; they did not believe that their evaluations were rooted solely in the these standards; they believed their evaluation conferences with supervisors helped then to improve their practice; their evaluations were not tied to annual goals or individual and school

professional development plans; supervisors did not use multiple sources of data as a part of teacher evaluation; supervisors do not spend enough time in classrooms to carry out expectations for effective evaluation; and that peer observation aided teachers in improving their practice. While current research asserts that standards-based teacher evaluation systems coupled with an array of activities that foster teacher ownership for the process can be used as a powerful vehicle for improving student achievement, study findings were inconclusive regarding any perceived association between standards-based performance evaluation structures and practices and the improvement of student achievement.

Dedicated to my soul mate, my rock, and my best friend.
Paula, I love you more than I can ever express.

ACKNOWLEDGEMENTS

I embarked upon this academic endeavor nearly ten years ago in the fall of 1999. In the amount of time it has taken me to complete this seemingly insurmountable task, tremendous, life-changing events have come to pass within my personal life and within in the world around me: Paula and I have purchased a home and brought into the world our two dearest treasures, Annabella and Giovanna, and anxiously await the arrival of their sister/brother; I have realized my career ambitions of becoming a high school and district administrator within a very vibrant school and district; I have traveled to Belize, Canada, China, Ireland, Italy, and Mexico; my life - indeed all of our lives - have been forever shaken and our collective perspective of the world inexorably altered by the tragedies of September 11, 2001; the Red Sox have swept in and won two World Series Championships; the Patriots have been to five and have won three Super Bowls; and the Bruins have, well, they're the Bruins; and most recently, the nation has elected, with a mandate, mind you, its first African American president, Barack Obama. Throughout all of this, my family, friends, colleagues, and instructors have been unwavering in their support of my pursuit.

First, I would like to acknowledge my dissertation chair, mentor, and the greatest teacher I have ever had, Dr. Irwin Blumer. Under his careful and deliberate tutelage, my academic and practical understanding of public education has grown exponentially and I can say with confidence that I now know what it means to be a life-long learner. I thank Drs. Robert "Jerry" Starratt and Joseph Pedulla for agreeing to serve as readers on my panel. Their time, input, and guidance were invaluable to the completion of this work. I am also grateful for the time, assistance, and support of the following members of the Boston College community: Dr. Diana Pullin; Director Fran Loftus; Dr. Sandy Morse; Drs. Danné Davis and Janna Jackson for their

much needed camaraderie and early assistance with my work; Leslie Ciampi and Theresa Lungu of the Office of Educational Administration for their administrative aid; and Terrence Lee St. John, the hard-working and talented statistics student who provided me with invaluable consultancy. I would also like to thank Dr. Janice Jackson of Harvard University for her ongoing support; and Dr. Nicole Guttenberg of the Boston Public Schools for serving as editor, cheerleader, and as a superb colleague.

I owe a debt of gratitude to my life-long friends, Mark Coupe, Austin Kairnes, and Marc Cobb for their encouragement and unspoken motivational cues. Though they never uttered the phrase, I was spurred on by the pregnant pauses in conversation and frustrated looks that screamed, "You're not done yet?"

Throughout the entire experience I have relied on the support and sacrifice of my family. Paula's support in my pursuit of this degree was unwavering from the beginning and her love and determination have helped me to see the process through to fruition. Annabella and Giovanna have sacrificed many nights and weekends without me to this endeavor. Though the time cannot be replaced, know that my hard work was spurred on by the love that I have for you both and that my Saturdays are yours again. Robby and Kathy, Alex and Justin, Sheri and Berry, my in-laws Toni & Tony DiMichele, thank you for supporting my career, for never doubting my ability to complete this degree, and for being tenacious in your constant inquiries regarding my progress. I thank you and love you all very much.

There are no words to express the undying love, admiration, and gratitude that I have for my parents, Robert and Donna Pizzi. Their love was always unconditional and their expectations for my success never second best. My parents believed in the power of and necessity for rigorous formal education and sacrificed tremendously to provide me with the best instruction possible.

To this end, my mother (who always told me that I would wear a suit) worked tirelessly as a hair dresser then as an insurance actuary and underwriter. My father - still a dynamo at 64 and the very model of the Puritan work ethic – "grounded" my formal education with real-world experience in the family electrical business. Thanks, Dad, for doing this for me "one plug at a time." As a parent myself, I am learning just how difficult it must have been for my own parents to invest so much of their love, caring, and very lives into my development while at the same time providing me with the freedom to choose my own destiny. This, I am beginning to see, is the essence of parenthood and I thank both of my parents for setting such a tremendous example.

If time and policy permitted, I would have all of your names listed along-side of mine on my degree. You all have helped me to realize a dream that I had, at times, convinced myself that I had no business pursuing. Your alternating patience and impatience and constant love and support were crucial to this realization, and for that I am indebted to you all.

TABLE OF CONTENTS

Chapter One: The Research Problem

As part of its revolutionary effort to promote educational equity and to spur ongoing

instructional improvement within Massachusetts's public schools, in 1996 the Commonwealth of

Massachusetts adopted and distributed the "Principles of Effective Teaching" to all public

schools as part of the Education Reform Act of 1993 (MERA). The Commonwealth intended

that these principles serve as the basis for standards-based teacher evaluation systems within all

Commonwealth public school districts (Department of Education, 2001). Through survey

research using qualitative and quantitative methods, the proposed study seeks to investigate

urban secondary teachers' perceptions of teacher evaluation systems that are based upon these

aforementioned standards.

A History of the Problem

At the beginning of the twentieth century, the efficiency of American industry was

revolutionized by the Time and Motion Studies that Frederick W. Taylor performed for the

automotive industry. Through his "scientific management" approach to maximizing industrial

worker proficiency, "Taylor broke tasks into minute parts and retrained workers to get the largest

payoff from each motion and second spent at work" (Bolman & Deal, 1997). As individual

worker productivity soared, expected worker output became highly quantified and resulted in

worker performance ratings based upon "scientifically" derived projections for output. As with

many management techniques originating from the American business-industrial complex,

Taylor's studies greatly influenced the organizational structure of American public schools. As

with autoworkers, teacher performance was assessed by means of readily observable and

quantifiable classroom behavior (Bolman & Deal, 1997). Consequently, until the publication of

A Nation at Risk in the mid 1980's, teacher evaluation tools reflected this thinking and included

such superficial and readily noted criteria such as teacher dress, the efficiency and thoroughness with which teachers performed administrative tasks, and disciplinary climate of the classroom. Student performance and learning rarely factored into a supervisor's determination of a teacher's effectiveness. Arguably, this trend continues in practice as "current systems rely heavily on the documentation of a small number of 'observable behaviors,' such as 'writing the learning objectives on the board,' 'smiling at students as you greet them,' and the like" (Danielson & McGreal, 2000).

The Commonwealth's Response

In 1996, the BOE (Massachusetts Board of Education) adopted a set of standards for teacher instructional performance labeled "The Principles of Effective Teaching." Developed by the Massachusetts-based educational firm, Research for Better Teaching, Inc. (RBT), the original document outlined seven standard areas of effective teaching:

- Currency in the Curriculum;

- Effective Planning and Assessment of Curriculum and Instruction;

- Effective Management of the Classroom Environment;

- Effective Instruction; Promotion of High Standards and Expectations for Student Achievement;

- Promotion of Equity and Appreciation of Diversity;

- Fulfillment of Professional Responsibilities.

The adopted document also contained examples and "descriptors" that described general teacher behavior within each of the seven areas (see Appendix A). The overall intent of the authors was to create a set of standards that would serve as an overarching common language for teacher evaluation and instructional performance for all Commonwealth public schools; the

language of these standards (or principles) was supposedly more closely linked to student

performance and learning outcomes: "All performance standards established for teachers shall

be consistent with and meet the Principles of Effective Teaching adopted by the Board of

Education (Department of Education, 2001).

In adopting this common language, the DOE illustrated its belief that a primary means of

instructional quality control in education is teacher evaluation. Commonwealth of Massachusetts

Regulation 603 CMR 35.00, § 35.02 defines teacher evaluation as "the ongoing process of

defining goals and identifying, gathering, and using information as part of a process to improve

professional performance and to assess total job effectiveness and make personnel decisions"

(Department of Education, 2001). In § 35.05 of the same regulation, the DOE mandates that

each district submit their new standards-based teacher evaluation process to the DOE: "Prior to

September 1, 1996, each school committee shall file its evaluation procedures and performance

standards for teachers … with the Department of Education unless such school committee is then

engaged in the interest arbitration process …" (Department of Education, 2001).

Thus, the Principles of Effective Teaching were distributed to every Commonwealth

public school district. The districts were subsequently instructed either to adopt the principles as

distributed or, as previously stated, to develop a unique set of standards for teacher instructional

performance based upon the Commonwealth's Principles of Effective Teaching. Consequently,

each public school district adopted the six standard performance areas as authored by RBT.

Their respective descriptors, however, were bargained at the local level. As a result, while the

headings are the same across the Commonwealth, the instructional behavior that reflects each

standard area differs from district to district. In the case of the district which participated in the

present study (District A), the standard areas for teaching performance differ from those

disseminated by the Commonwealth DOE in 1996. While the essence of the Principles of

Effective Teaching are captured in this district's process, the performance evaluation instrument

for District A's teachers is expanded to include seven areas for evaluation. District A added a

seventh competency area of "Professional Responsibilities" to its evaluation instrument (see

Appendix B).

Statement of the Problem

A comprehensive search of multiple databases (ERIC, ERIC Webspurs, OCLC First

Search, etc.) indicates that no quantitative dissertation has been conducted that solely examines

perceptions of Massachusetts' secondary teachers toward the Principles of Effective Teaching

and standards-based evaluation systems.

The problem exists, then, that teacher perceptions of standards-based teacher evaluation

systems have not been adequately examined quantitatively. Consequently, an essential set of

data that could aid in the overall and ongoing improvement of the standards-based teacher

evaluation in the Commonwealth has yet to be explored. Using a research protocol adapted from

the *Study of Systems for Evaluating Massachusetts Teachers* (MassPartners, 2000) which serves

as the pilot to the present study and will heretofore be referred to as such, it is the aim of the

present study to unearth some of these data in one urban secondary public school in the

Commonwealth. The protocol for the present study includes an initial focus group, distribution

of a survey instrument to most, if not all full-time teachers within one large secondary urban

school, analysis of this survey data, and use of a follow-up focus group to clarify quantitative

responses.

Question for Research

The primary research question for consideration in this dissertation is, What are teacher perceptions of evaluation in an urban high school that uses a standards-based evaluation system? The researcher sought to answer this question in terms of teacher familiarity with the stated standards and their indicators and the extent to which the standards are utilized as the common language for evaluation activities and instructional improvement within a school.

Theoretical Rationale

The review of related literature for this study is organized into two overarching areas. These are: 1. an overview of the development of modern teacher evaluation standards and the associated systems for teacher supervision and evaluation and 2. teacher and administrator perceptions of these standards and systems, including obstacles to effective supervision and evaluation. These general themes are reflective of the research question and will provide further insight into the purpose of, need for, and obstacles to effective teacher evaluation.

Based upon findings of the pilot study, the researcher hypothesizes that teachers do not believe in the efficacy of applied standards-based teacher evaluation as a means of improving their practice. While evaluation instruments have changed in order to accommodate the aforementioned need, implementation of evaluation and professional ideologies, i.e. those of teachers, administrators, and researchers, professional behaviors have not, according to much of the current literature. For these reasons, an examination of relevant literature on the subject of formal teacher supervision and evaluation has been conducted. This includes significant treatment of teacher evaluation during the standards-based education movement, as well as the different ideas and practices that have emerged from it. The concepts of differentiated teacher evaluation (i.e. specialized systems of teacher evaluation and relative support based upon

individual teacher competence) and collegial observation and coaching, for example, have emerged as leading vehicles for the improvement of teacher practice in many standards-based districts. Major contributors to this field of thought include Robert J. Starratt (*Transforming Educational Administration: Meaning, Community, & Excellence*, 1996), Thomas Sergiovanni and Robert J. Starratt (*Supervision: A Redefinition*, 1993), P. Rettig ("Differentiated Supervision: A New Approach", 1999), and Danielson and McGreal (*Teacher Evaluation: To Enhance Professional Practice*, 2000). The principal researcher also includes authors of multiple works and articles that examine the subject of teacher evaluation and professional development within the context of standards-based reform. Contributors to this field include, but are not limited to, Danielson and McGreal (2000), Glanz (2007), Marshall (2006), Starratt (1996), and Vandeweghe (2006).

The second part of this review provides an examination of professional literature on teacher perceptions of evaluation, including those obstacles that inhibit the use of teacher evaluation as a means of improving teaching and learning. Here, the principal investigator found that a lack of time spent in the classroom by administrators, failure by administrators and teachers to set priorities around instructional improvement, as well as a lack of targeted professional development for teachers and administrators in instructional supervision and evaluation all serve as foremost hindrances to effective teacher evaluation. For example, the research of Paul Bredeson ("Superintendents' Roles in Curriculum Development and Instructional Leadership: Instructional Visionaries, Collaborators, Supporters, and Delegators", 1996) illustrates that while administrators believe that setting priorities around instructional improvement is crucial, very few of them actually do. A study conducted by the Massachusetts Elementary School Principals Association (1999) also found that principals believe supervision

and evaluation to be their most important function. However, they also cite lack of time as a major obstacle to their use of supervision and evaluation as a key means of improving teaching and learning. In addition, DeMitchell ("Competence, Documentation, and Dismissal: A Legal Template", 1995) claims that administrators do not seek to improve the performance of mediocre and marginal teachers due to a lack of administrative time to carry out evaluation, an unwillingness to document incompetence when faced with subsequent "incomplete information," lack of confidence in their own skills as supervisors and evaluators, and the human predisposition to avoid conflict in social interactions (in this case, "withholding of negative information regarding performance from ineffective teachers", 1995).

An in-depth discussion and examination of these works is presented in Chapter 2 of the present work. These and many other works served as the theoretical framework around which the principal investigator structured the present study

Significance of the Study

The primary intent of promulgation of the MERA of 1993 was to promote educational equity for all public school students in the Commonwealth. As the Massachusetts Supreme Court held in it's 1992 decision in *McDuffy v. Executive Secretary of Education, et. al.*, students of lower socioeconomic school districts were more likely to receive substandard education due in part to the property tax based formulas by which schools in those districts were funded. The Court held that the resultant lack of masters-level subject area teachers and inadequate instructional resources directly impacted student achievement within these districts. Consequently, reform of the Commonwealth's policies concerning the funding of public education was a key component of the MERA. In addition to bolstering financial support for education, the Commonwealth also demonstrated its specific commitment to improving the

quality of instructional practice by adopting the state Curriculum Frameworks as content and

skill standards for student learning (not to be confused with standards for teacher/instructional

practice). The Massachusetts Comprehensive Assessment System (MCAS) examinations were

adopted as summative assessments through which educators could gauge student learning. For

licensed public school teachers, however, the teacher evaluation system at the district level

remains the primary means through which the quality of instructional practice can be formally

assessed and evaluated directly, as standardized tests such as MCAS, SRI, and SAT test

cumulative student performance that historically had been difficult to attribute to specific

teachers and instructional practices. As aforementioned, the Commonwealth has determined the

standards for instructional practice of all Massachusetts teachers through its adoption of the

Principles of Effective Teaching.

The significance of this study lies herein; it seeks to determine teachers' perceptions of an

evaluation system based upon the performance standards set forth by the Commonwealth, i.e.

"The Principles of Effective Teaching."   Overall, the principal researcher has tailored the

research design and question for research to determine how teachers in one urban school

perceive evaluation of their own practice in this standards-based model.

Research Design

The purpose of this investigation is to determine secondary teachers' perceptions of a

standards-based teacher evaluation system. The basic design of this study was developed and

implemented with that end in mind. Although the primary medium for data collection was a

Likert-based survey, the first piece of the design was qualitative in nature, i.e. an initial focus

group to facilitate survey development. This focus group was conducted in a small public

secondary school within an urban school district in the Commonwealth. Survey distribution was

conducted in a second large secondary school within the same district. These schools will be heretofore referred to as School A and School B respectively The researcher developed and utilized the "The Standards-Based Evaluation Attitude Scale," which contained thirty-one items in five sections (see Appendix D).

Study Limitations

The present study is limited by several factors. These include but are not limited to: the general limitations of survey research; the small number of study participants (30); and the restricted number of statistical analyses that could be conducted due to the small number of participants. These limitations and others are examined in more depth within Chapter Four of the present study.

Definitions of Terms

The following definitions of terms apply to the present investigation:

- Pilot study: the study conducted in the summer of 2000 for the Massachusetts Partnership for schools.

- Present study/present investigation: the present study is the piece of original research conducted and examined in this dissertation.

- Teachers/participants: secondary teachers of math, science, English, social studies, foreign language, and special and bilingual education in District A (teachers of students in grades 9-12).

- Administrators: principals, assistant principals, and other evaluators who are responsible for conducting formal teacher evaluation.

- Instructional performance standards, performance standards, standard areas, or standards: Research for Better Teaching's Principles of Effective Teaching and/or the variation of these teacher performance standards as defined and adopted by District A.

- Formal teacher evaluation: the system of assessing teacher instructional performance as defined and utilized within District A. While teacher supervision is an integral part of formal teacher evaluation, the two terms are not interchangeable.

Overview of the Study

This study is presented in five chapters. In Chapter 1, the researcher provides a basic introduction to the study. This includes an introduction to the topic of teacher evaluation and the purpose of the study, the history and background information of the topic, as well as an overview of the fundamental design of research. Chapter 2 presents the reader with a two-part review of relevant literature on the topic at hand. The first section features a discussion on the history and development of contemporary evaluation systems and standards and current schools of thought on the subject of teacher supervision and evaluation. In part two, the researcher examines both theoretical and research literature on teacher and administrator perceptions of the teacher evaluation and its associated activities.

Chapter 3 outlines the research design of the present study. This includes a brief review of the pilot study conducted in the summer of 2000. It also includes a description of the mixed methods that are employed in the present investigation, i.e.: the initial focus group and the protocol used; survey development and distribution; participant selection; and means of data collection and analysis. All major findings of the present investigation are offered to the reader in Chapter 4. Chapter 5 communicates the researcher's interpretation of these findings, as well as their implications and impact upon public policy and educational practice, as well as

suggestions for further investigation and research. In addition, the primary researcher compares

and contrasts the findings of the present study with the body of professional literature as outlined

in Chapter 2.

Chapter 2: Review of Related Literature

Introduction

Chapter 2, the Review of Literature, is centered upon the research question: What are teachers' perceptions of evaluation in an urban high school that uses a standards-based evaluation system? The researcher sought to answer this question in terms of teacher familiarity with the stated standards and their indicators and the extent to which the standards are utilized as the common language for evaluation activities and instructional improvement within a school. The first area of investigation provides a historical view of teacher evaluation and the standards used therein. Within this area, the principal researcher examines the historical roles and relationships between teachers and administrators as well as the evolution of standards and criteria for supervision and evaluation (Abbott, 1969; Beck, 1993; Blumberg, 1985; Dreeben, 1965; McGregor, 1960; Millman, 1990; Postel, 1955; Schott, 1956; Stanley, 1988; Travers, 1981). It is important to note that the principal function of this section is to explore the evolution of standards for teacher evaluation and not solely to outline the aforementioned practices, procedures, and roles.

In the second part of this chapter, the researcher reviews literature that advocates for teacher ownership for the evaluation process. Here, authors discuss the importance of teacher engagement in the creation and implementation of evaluation processes, teacher reflection and highly structured teacher-supervisor conversations as critical components of teacher-owned supervision and evaluation. The integration of professional development into teacher-owned evaluation activities also emerges as a prominent theme in this section (Beck & Murphy, 1993; Black, 1993; Danielson & McGreal, 2000; Daresh, 2001; Glatthorn, 1984; Rettig, 1999; T. J. Sergiovanni & Starratt, 1993).

The main purpose for this investigation is to uncover urban teachers' perceptions of evaluation. Professional literature, however, provides significant treatment of administrators' perceptions of teacher evaluation as well. Consequently, the second area for review highlights literature surrounding both teachers' and administrators' perceptions of teacher evaluation and associated activities. Here, teachers' perceptions of administrator/supervisor competence are examined, as well as teachers' ownership for their evaluation and its related activities and their familiarity with these processes and related standards (Howard, 1996);(MassPartners, 2000); (Rooney, 1993); (Shen, 1998); (Wang, 2002); (Zimmerman & Deckert-Peton, 2003)). In addition, the literature provides a significant treatment of administrators' perceptions of teacher evaluation both at the district and school levels (Bredeson, 1996; Dawson & Acker-Hocevar, 1998; MassPartners, 2000; MESPA, 1998; T. Sergiovanni, 1985).

Development of Supervision & Evaluation

Instructional Performance Standards: A Historical Perspective. The presence of instructional standards for teachers in the United States can be traced as far back as the 1840s *(Blumberg, 1985)*. During this era, traveling circuit superintendents were responsible for the direct supervision and evaluation of local teachers. Teacher ratings were communicated by the rudimentary policy of issuing first, second, and third class teacher licenses by these circuit superintendents and criteria for issuing these licenses were arbitrary. A teacher who lacked adequate "energy" for the teaching experience, for instance, could be assigned a second class license while a teacher determined to hang "like an incubus upon the schools in which they are employed" would most assuredly receive a third class license (Blumberg, 1985, p.60). According to Blumberg, however, many circuit superintendents wrote of explicit criteria for competent teacher performance that were derived from their fundamental beliefs in the role of

the public schools in the preservation and growth of democracy.   Superintendents characterized

the best, or "first class teachers" as being able to relate to student interests, to modify

instructional strategies based upon student skills and needs, and to creating logical, analytical

thinkers *(Blumberg, 1985)*.   In this case, the standards for teacher practice were not prescribed;

in fact, the aforementioned characteristics of "first class" teachers were gathered by the

Superintendent of Common Schools of the State of New York in 1845 and were representative of

comments from circuit superintendents in 59 counties *(Blumberg, 1985)*. Despite these

classifications, American educators largely held that student learning was solely the

responsibility of the student. This belief was not challenged until the twentieth century *(Travers,*

1981).

Thus, the seeds for using instructional standards as a means of evaluating teacher practice

were present within the educational paradigm within pre-industrial America.   According to

Beck and Murphy (1993) however, the place of teaching performance standards can be seen

through the changing metaphors for the principalship, throughout the twentieth century. The

American principal of the 1920s, for example, is viewed as the "values broker" within society,

one who is meant to extol and promote the certain virtues, i.e. academic truth, justice, and

aesthetics (Beck & Murphy, 1993). By the 1930s, the tenets of Frederick Taylor had filtered

through to educational practice and the principal was seen as the "scientific leader", an effective

one attending to all things practical with competence, expediency, and diligence (Beck &

Murphy, 1993; Bolman & Deal, 1997). In the 1940s and 1950s, principals' authority was often

compared to that of military commanders, as school leaders were "encouraged not to share their

authority, but rather to exercise it-working through established hierarchical structures" (Beck,

1993, p. 204).

Clearly, this latter metaphor for school leadership does not allow for teacher ownership of their practice and its improvement. In fact, effective principals were expected not only to be efficient micromanagers of staff time and school operations, but also to be master educators within their schools, providing teachers with the latest instructional theory and subsequently overseeing teachers to ensure proper implementation (Beck & Murphy, 1993). While focused on instruction, the principal's supervision of teachers was very didactic in nature, and required her/him to be the expert on specific instructional techniques. Rather than leading teachers in reflective conversation regarding their practice, principals were expected to insist upon the use of specific resources and techniques that they knew to be effective and to develop instructional programs and curricula for teacher use (Postel, 1955; Schott, 1956).   Hence, the criteria for successful teaching would, then, vary widely from school to school based upon individual principal's beliefs about education and their knowledge of effective and current instructional theory during the decades in question.

For school leaders, the 1960s and 1970s brought with them new definitions of accountability and consequent changes in expectations for teachers. During these decades, principals still maintained the role of master teacher, i.e. the instructional expert of the school (Beck & Murphy, 1993). With the political and consequent social turmoil of the 1960s, the principal was viewed as the protector of all that is stable and organized, i.e. educational bureaucracy. This overarching theme reinforced the predominant metaphor for twentieth century education as a hierarchical organization akin to an industrial plant or a military division. Here, the primary goal of the principal was the maintenance and efficient management of the educational bureaucracy within the school as a means of maintaining social stability (Abbott, 1969). Within the school hierarchy, the principal still served as the expert on all matters

instructional who viewed pupil instruction as a science that could be mastered (Beck & Murphy, 1993).

Though not treated extensively within this review, the present study, the aforementioned paradigm was supported by a new model of teacher supervision and evaluation. Though tenets of the process had been discussed prior to its emergence, clinical supervision model was formally codified by Morris Cogan, Robert Andersen, and Robert Goldhammer from the mid 1950s through to the early 1970s (Neville and Garmen, 2000). The overarching objective in this model reflected that of medical practice and is centered on the planning for, supervision, and treatment of teachers' instructional practice. This is all predicated upon the formation of a respectful and trusting individual relationship between the teacher and supervisor. In practice, Cogan and Goldhammer characterized clinical supervision through the following "methods" or "stages": "establishing a relationship with the teacher, observing classroom teaching, analyzing classroom scenarios, and holding a conference with the teacher" (Neville and Garmen, 2000, p. 230). Underlying these objectives and practices, however, remained the underling belief and expectation that principals were still expected to be the "master teacher" and "professional developer" within the school building. Even as teacher supervision and evaluation developed toward more reflective practices, the basic tenets of clinical supervision as mentioned above have permeated most subsequent models of teacher supervision and evaluation, many of which are outlined below.

This perception of the principal's role and instruction, predictably, had a marked impact on the supervisory relationship between teachers and their principals. In terms of professional interaction, this perception resonates well with the "X Theory of Personnel Management" (McGregor, 1960). Within this organizational paradigm, teachers are viewed as cogs in the

machine of the aforementioned bureaucracy. These cogs must be constantly observed and maintained for the machine to function properly. If one were not working, it had to be fixed or replaced at once. Hence while a focus upon instruction does exist at this time, as head of the organization the principal was charged not with creating an atmosphere and culture of change, but with promulgating particular instructional techniques. Teachers were to be constantly supervised and those who failed to implement these techniques were viewed as non-compliant and unreliable and had to be replaced (Abbott, 1969; Beck & Murphy, 1993; McGregor, 1960; T. J. Sergiovanni & Starratt, 1993).

The aforementioned paradigm for teacher supervision and evaluation was to be thoroughly challenged by an emerging teacher evaluation ideology. In the mid nineteen-seventies, the Joint Committee on Standards for Educational Evaluation was developed as a consortium of sixteen professional educational associations. The committee's initial goal was to create standards for educational program evaluation, which it released initially in 1981 and subsequently in 1994 as the Program Evaluation Standards (Stufflebeam & Sanders, 1990); (Worthen, 1987). At their initial meeting in 1975 the committee intentionally decided not to develop teacher evaluation standards (Stufflebeam & Sanders, 1990). This position would change for three reasons, however, in 1984 when the Joint Committee saw a demonstrated need for personnel evaluation standards that would aide state and local education authorities in both analyzing and developing teacher evaluation systems (Worthen, 1987).

As stated above, when the Joint Committee was created, it intentionally avoided an exploration into the development of such standards because its members believed the topic too controversial and conflict- ridden. By 1984, however, the Joint Committee believed that it had matured enough as a functioning professional body to initiate the enormous and seemingly

contentious task of developing personnel evaluation standards, i.e. standards for creating

effective teacher evaluation systems (Stufflebeam & Sanders, 1990). Second, and perhaps most

important, the committee examined contemporary research on supervision and evaluation of

education professionals to find that personnel evaluation systems were in dire need of reform.

According to research at the time, teacher evaluation systems were viewed as ineffective and

fruitless and performance evaluation criteria were seen as arbitrary, biased, irrational, and

specious at best (Darling-Hammond, Wise, & Pease, 1983; NCEE, 1983). Thus the Joint

Committee sought to ameliorate the conflicts caused by the lack of coherent systems for teacher

evaluation that did not articulate clear standards for performance. Through a series of analyses,

drafting, peer review, public hearings, and field testing both on the national and international

level, the Personnel Evaluation Standards were finalized by the Joint Committee on Standards

for Educational Evaluation in 1988 (Worthen, 1987). The Joint Committee developed four

standards by which teacher evaluation systems should be created or revised. As Sanders (1997)

indicates, the standards of Propriety, Utility, Feasibility, and Accuracy should be the

quintessential tenets used by workgroups charged with devising and/or revising teacher

evaluation systems in order that they are fair and accurate and provide teachers and

administrators with clear indicators of good teaching as well as time for discussion and reflection

upon teacher practice (Worthen, 1987).

Essentially, the Joint Committee argued that teacher evaluation systems should adhere to

the guidelines outlined within each standard. Through the five Propriety standards (service

orientation; formal evaluation guidelines; conflict of interest; access to personnel evaluation

reports; interactions with evaluates), for example, system developers could ensure that

evaluations were ethically and legally sound instruments. These standards were created to make

certain: that teacher evaluation instruments protected both the privacy and educational rights of students and their families; that evaluation instruments and processes were applied with integrity and consistency; that evaluation policies and procedures reflected contractual obligations and stated school district policies; and that teacher professionalism and dignity were upheld and valued throughout the evaluation process (Worthen, 1987). The five Utility standards (constructive orientation; defined uses; evaluator credibility; functional reporting; follow-up and impact*)* were written to provide guidelines to guarantee the usefulness of evaluation systems. Here the Joint Committee stressed: that the objectives of teacher evaluation be clearly outlined and explained fully prior to the beginning of the process; that evaluators be fully trained, knowledgeable, and competent as such; and that feedback be timely, accurate, and useful to the educator in improving his/her instructional performance (Worthen, 1987).

The Feasibility standards (practical procedures; political viability; fiscal viability) were authored to promote the practicality of implementation of teacher evaluation systems. For example, in determining evaluation system feasibility, in 1988 the Joint Committee suggested that instrument developers ask the questions:

1. Are the procedures easy to implement, with minimal disruption and cost?

2. Were stakeholders in the teacher evaluation involved in the development of goals, policies, and procedures of teacher evaluation? Are the goals, policies, and procedures acceptable to all?

3. Are adequate time and resources provided for teacher evaluation (Worthen, 1987, p. 96)?

Last, the Joint Committee outlined seven Accuracy standards (defined role; work environment; documentation of procedures; valid measurement; reliable measurement;

systematic data; bias control; monitoring evolution systems) to mitigate subjectivity inherent in

evaluation instruments, evaluation processes, and the evaluators themselves (Worthen, 1987).

The intent of the Joint Committee here was to ensure that evaluation system creators and

reviewers accounted for the multitude of variables that could affect student learning and teacher

performance. For instance, evaluators and teachers should be certain that the professional roles

and responsibilities of those being evaluated are clearly outlined and understood by all

stakeholders. Furthermore, evaluation systems should uphold strict standards for due process

and confidentiality and should measure teacher performance with accuracy and consistency

(Worthen, 1987). Ultimately, the Joint Committee intended that teacher evaluation systems be

reviewed in their entirety on a periodic basis (every 1-2 years) to make certain that the

aforementioned standards were being met within teacher evaluation instruments (Worthen,

1987).

New Influence in Supervision. Prior to the development of uniform and mandated

instructional performance standards, Madeline Hunter developed a set of standards for effective

instruction generated with the end of improving and measuring the effectiveness of teacher

practice. In the 1970s, Hunter outlined the Elements of Effective Teaching as manifest in the

Lesson Design Model (Shinkfield & Stufflebeam, 1995). These elements are: (1) the

anticipatory set, or an initial activity designed to introduce a new topic or to activate prior

knowledge of a concept or skill from the previous lesosn; (2) the lesson objective and purpose,

i.e. a statement of lesson importance and what the student should be able to do or to know at the

end of the lesson; (3) instructional input, or the information or skills needed for students to

successfully complete the lesson, including skills or concepts that they already know or new

skills and concepts; (4) modeling, where the teacher models the skill or concept to be mastered

by the students; (5) checking for understanding, in which the teacher utilizes a variety of

instructional techniques and questioning strategies to ascertain the level of student understanding

of the new skill of knowledge; (6) guided practice, in which students attempt to model the new

concepts under the direction and with the assistance of the teacher; and (7) independent practice,

where the teacher allows students to demonstrate their mastery of the subject or skill without

mistakes or failures being corrected by the teacher (Shinkfield & Stufflebeam, 1995). Relative to

supervision and evaluation, as early as 1979, Hunter called for "a diagnostic prescriptive

instrument that provides knowledge of results in professional performance" (Hunter, 1979, p.

67). Her belief was that by training teachers to model effective planning and instruction and by

training supervisors to identify these elements, all parties speak in sophisticated ways about

teaching, learning, an ongoing instructional improvement. The seven elements of effective

instruction and their indicators would serve as the common vocabulary for these conversations.

The lesson development model itself consists of teacher-centered competencies for

planning and practice that Hunter identified synonymously with the concept of direct instruction.

In regard to the latter, she was clear that her model called for observation of specific teacher

skills and practices, not of student learning or outcome. Shinkfield indicates:

> The essence of Lesson Design is that teachers learn to spend instructional time in areas
>
> where there is reasonable support for lesson plans having a direct impact on student
>
> learning. Although the focus is placed on the teacher and the teacher's decision-making,
>
> the approach provides an opportunity for leaders to work closely with the teacher to
>
> achieve planned ends (1995, p. 196).

Hence, the underlying belief was that by training teachers to model effective planning and

instruction and by training supervisors to identify these elements, all parties speak in

sophisticated ways about teaching, learning, an ongoing instructional improvement. The aforementioned seven elements of effective instruction and their indicators would serve as the common vocabulary for these conversations. To this end, both the Association for Supervision and Curriculum Development (ASCD) and the National Assessment of Educational Progress (NAEP) both adopted and supported Hunter's approach to lesson planning an supervision.

Hunter's norms differ from the more modern state-mandated evaluation standards, e.g. The Principles of Effective Teaching (see Appendix A), in several ways. First, the Elements of Effective Teaching and Hunter's associated Lesson Design models act as a broad set of norms for teacher planning and instructional effectiveness from which teacher practice can be assessed in formative ways. The seven elements are intrinsically aimed at improving instruction through observation of and subsequent dialogue about teacher practice, though not specifically about student learning. Likewise, the Principles of Effective Teaching serves as a common language for teaching and learning within participating schools of the present study. These standards, however, contain very detailed descriptors that outline expected teacher behaviors both inside and outside of the classroom. More important, these standards (which serve as the basis for formative and summative teacher evaluation) outline contain indicators that assess the effectiveness of teacher practice through observations of specifically defined student behavior and performance. Thus, while Hunter sets a construct for how teachers should develop and implement lessons and make instructional decisions, these elements are only part of more modern standards which are used to assess and judge teacher knowledge, behavior, and decision-making by looking at anticipated and actual student learning outcomes, not just the teacher planning and decision-making processes in and of themselves.

As indicated, Hunter was clear that prior to the training of teachers, school leaders must receive in-depth training on effective models of instruction in order that effective instruction is properly supported and understood in the classroom (Shinkfield & Stufflebeam, 1995). While Hunter did outline and develop the Elements of Effective Instruction, as examined later in this chapter, only in the 1990s was the effectiveness of her work upheld through the research of McGreal (Shinkfield & Stufflebeam, 1995).

As the literature indicates, development of the Personnel Evaluation Standards marked significant paradigm shift concerning the supervision and evaluation of teachers. While the standards did not provide an outline of effected and expected instructional behavior, they did provide a framework for the standards-based systems teacher evaluation systems that were subsequently implemented throughout the United States.   This is also true concerning multi-state adoptions of teacher competency assessments and standards for both initial and recertification, e.g. the Praxis series of teacher assessments as developed by Educational Testing Services beginning in the late 1980s (Danielson, 1996). In addition, Hunter's Elements of Effective Instruction provided criteria for teacher-centered decision making within the classroom regarding the instruction of individual students and her work set the foundation for the development of teacher evaluation standards in the United States.

Supervision and Evaluation: Contemporary Theory & Practice

Contemporary Evaluation Systems & Teacher Ownership. The advent of the Personnel Evaluation Standards underscored the beginning of a very different view of teacher supervision and evaluation from those of early to mid twentieth century.

Beginning in the early 1990s, teachers and teachers' unions became involved in the development of evaluation procedures for use in their respective districts (Black, 1993). Just as

the management, oversight, and inspection of teachers had dominated the routines of principals

and administrators as supervisors and evaluators in earlier years, the predominant features of this

new paradigm were teacher ownership of the evaluation process through development of

differentiated supervision and evaluation systems and professional development, thus allowing

for more reflective practice and teacher ownership of their own instructional improvement

(Ballard & Gullat, 1998; Black, 1993; Mertler, 1997; Rooney, 1993; Shen, 1998; Weiss &

Weiss, 1998).

Subsequent to the development of the Personnel Evaluation Standards, states such as

California, Texas, Vermont, and Massachusetts began to develop and adopt standards by which

teacher practice would be measured (Danielson & McGreal, 2000). Hence, rather than

supervising and evaluating teachers on readily observable managerial teacher responsibilities,

performance standards now provided the foundation for teacher supervision and evaluation

systems that focused upon teacher practice as it affected student learning (Danielson & McGreal,

2000). School districts were now primed for the task of developing evaluation systems that

elicited teacher ownership for student achievement and instruction improvement based upon

these standards (Black, 1993; Danielson & McGreal, 2000; Rettig, 1999).

For example, in the early 1990s, Black (1993) examined the transformation of evaluation

systems in one New York state school district (1993). She describes the original evaluation

processes in Elmira City School District to be arbitrary, punitive, and intimidating in nature.

Any instructional standards used in the processes were equally subjective. Black indicates that

the impetus for evaluation reform came from the teachers' union which sought to provide due

process, stability, and equity in the teacher evaluation process. In addition to outlining these

practices, the Elmira City district worked with members of the bargaining unit and administrators

to make instructional improvement the quintessential purpose and object of teacher evaluation. In terms of instructional standards, teachers and principals worked together to question the subjectivity of teacher performance ratings on summative evaluations and to determine the needs of professionals at different places in their careers.

While Black does not report on specific instructional standards, she does explain some processes used for reaching them. For principals, this meant outlining instructional expectations and areas for instructional improvement at the beginning of the school year. Ideally, teachers would be involved in setting these expectations as they would be a major focus of all evaluation activities, e.g. conferencing, observation, reflection, goal setting, for the entire school year. Thus an evaluation system which teachers and administrators once viewed as perfunctory and futile was transformed into one which promoted collegiality among teachers and administrators, teacher ownership for the process, and a strong emphasis on instructional improvement (Black, 1993). As aforementioned, the system as outlined by Black bear's is reflective that the Clinical Supervision model as promoted by Goldhammer in the 1960s.

On the contrary, contemporaries of Black sought teacher evaluation reform from a much different perspective (Rooney, 1993). Like Black, Rooney describes the evaluation system in the Community Consolidated School District in Palatine, Illinois as one in need of change. Ironically, the system she describes is nearly identical to the one to which Black (1993) aspired in Elmira City. Rooney sought to make the evaluation process one from which teachers gained valuable insight into their own practice, thus providing opportunities for teacher reflection and improvement. She outlines a different level of professional ownership for supervision and evaluation in which teachers have a much more active and collegial role in the definition and function of evaluative processes within collectively bargained parameters.   The resulting

teacher-generated system relied heavily on group pre and post lesson conferencing, peer

observation, peer coaching, and exclusively formal visits by administrators (Rooney, 1993).

While ownership for teacher evaluation was the desired objective here, the author and

administrator states a much more important by-product of this ongoing dialogue: "What

changed in our system – more than anything else-is the concept of the principal. I am no longer

the one responsible for a teacher's behavior. Teachers are now responsible for their own

professional growth –both individually and as a group" (Rooney, 1993, p. 4).

Mertler (1997), Rettig (1999), Shen (1998), and Weiss (1998) all describe observing

teacher evaluation systems similar to those Black (1993) and Rooney (1997) sought to transform.

Again, each of the aforementioned authors characterizes teacher evaluation systems by

describing many of the same elements. In each case, the principal served as the sole source of

feedback for teacher performance. In actuality, this lead to little or no feedback at all as

classroom observation was often not a priority for principals and administrators. Second, many

evaluation standards were based upon readily observable indicators which revealed little if

anything about student learning or teacher practice. Even more disconcerting is the fact that

from a single observation, principals were to offer teachers feedback upon: "teacher's

appearance, professional development, rapport with colleagues, and instructional ability" (Rettig,

1999). Conspicuously absent from these processes were peer observation and feedback,

effective pre and post conferencing, and the use of multiple sources of information in addition to

classroom observation that provide a more complete picture of a teacher's performance (Mertler,

1997; Rettig, 1999; Shen, 1998; Weiss & Weiss, 1998). The transformed evaluation systems

provided for the aforementioned collegial structures and methods implemented in New York and

Illinois (Black, 1993; Rooney, 1993) and resulted in improvement of teacher morale, practice

and collegiality and in a more sophisticated understanding of instruction.

More recently, a study utilizing mixed-methodology within the New York City Public

Schools yielded information about current supervision and evaluation systems which reflect the

aims of Black (1993) and Rooney (1997) and others. In this study, Glanz, Shulman, and Sullivan

(2007) seek a direct link between instructional supervision and student achievement. They

utilize a definition if instructional supervision that does not rely upon specified systems,

procedures, and logistical structures, but that calls for processes that foster change through

implementation of differentiated strategies, methods, and advances in order to improve both

classroom instruction and instructional leadership itself (2007).

In accordance with the aforementioned definition, the researchers conducted "an

instrumental case study" of administrators in one Brooklyn public school in which they sought to

"make the connection between supervision and student achievement" (Glanz, Shulman, &

Sullivan, 2007, p. 4). To this end, the researchers developed the five major questions for

research when constructing their study. The research questions were:

1. What does effective teaching "look like?"

2. What impact do supervisors have on teachers' in-class teaching behaviors and
   attitudes toward promoting student learning?

3. What is the connection among instructional supervisory practices, teacher
   classroom behavior, and levels of student achievement?

4. What can we learn about making the connection between supervision and student
   achievement (Glanz, 2007, p. 5)?

To answer these questions, the researchers gathered and analyzed both qualitative and quantitative data. The quantitative data consisted of multiple years of student performance information during the 2003-2005 school years. In particular, the data reflected grade three and five student performance on the New York City ELA and Mathematics Standardized Tests. To gather qualitative data, researchers asked teachers and administrators within the school the following questions:

1. What do you believe are the most important factors leading to improved achievement in your school?
2. What does the supervisory process look like in your school?
3. What types of classroom supports have been provided in the implementation of DOE initiatives?
4. How has student information, such as standardized test scores, been used to support classroom instruction?
5. What does professional development look like in your school (Glanz, Shulman, & Sullivan, 2007, p. 13)?

Several themes emerged from this study that not only coincide with the aforementioned findings of Black, (Mertler, 1997; Rettig, 1999; Shen, 1998; Weiss & Weiss, 1998)but that also build upon and expand their definitions of instructional supervision and of the instructional leader. One major theme was grounded in the relationships among teachers and their supervisors. Researchers found that all professionals, both teaching and administrative, took responsibility for instructional improvement and instructional supervision, which has historically been viewed as the sole responsibility of the administrator. This resulted in much more collegial relationships which, in this particular school, were grounded in shared leadership, mutual

respect, and trust. When instruction and student learning became the focus of all professional

interactions, the power dynamic between teachers and administrators virtually evaporated in this

regard. As mentioned by researchers over a decade earlier, Glanz (2007) discovered that all

observations in the Brooklyn school were considered informal. While administrators did share

leadership and instructional/professional development facilitation with teachers, Glanz also notes

that they did consistently take to task those teachers who did not implement agreed upon

strategies and curricula. Thus the overall focus was not upon teacher adherence to administrative

directives, but upon professional growth and application of mutually agreed upon practices

(Glanz, Shulman, & Sullivan, 2007).

In the Glanz (2007) study, researchers present very clear preliminary findings. Through

examination of the aforementioned quantitative student performance data, student achievement

on the New York City tests in ELA and mathematics increased substantially over the three year

course of the study, from sixty nine percent to eighty percent in meeting the ELA standards and

seventy three percent to eighty eight percent meeting the mathematics standards. The

researchers conclude that there is a definitive correlation between effective and collegial

instructional supervision and augmented student performance. Glanz, et. al. do, however, point

out that their data is too preliminary to ascertain whether the link is a causal or a correlational

one and that many curricular and instructional initiatives outlined in the study existed within the

Brooklyn elementary school during the previous school administration. It is quite clear, though,

that this study reflects the initial beliefs and practices as reported by (Mertler, 1997; Rettig,

1999; Shen, 1998; Weiss & Weiss, 1998). This includes the necessity for a collegial, not a

hierarchical, relationship among teachers and their supervisors. It also necessitates a clear and

common understanding of good teaching practices, the expectation among colleagues that these

practices will be implemented and measured for their effectiveness, and a school culture that

engenders trust and that supports and demands collegial behavior.

Differentiated Supervision and Evaluation

Like the aforementioned authors, teacher evaluation theorists of the mid-to-late 1990s

intimate that many evaluation systems at that time still reflected educational belief systems and

research from the 1970s (Danielson & McGreal, 2000; Nevo, 1994; T. J. Sergiovanni & Starratt,

1993; Starratt, 1996). They cite some of the flaws inherent in evaluation systems of the time as:

a lack of effective collegial conferencing among teachers and administrators about instructional

practice due to a lack of trust; a limited understanding of effective instructional practices and

evaluative expertise by administrators; and lack of clearly stated criteria for teacher performance

ratings (i.e. "meets/does not meet standards"). These researchers look beyond uniform

evaluative structures for all teachers when seeking to improve teacher practice through

evaluation processes. As espoused by Danielson and McGreal (2000), differentiation among

evaluation structures and practices must exist:

> Teachers at different stages in their careers have different needs and different levels of
>
> skill. Novice teachers are engaged in rapid and intense learning and require a good deal
>
> of support in developing their basic teaching skills. Experienced teachers, on the other
>
> hand, have many of the routines of teaching under control and can devote energy and
>
> attention to refining their skill. Experienced, but struggling, teachers may have to
>
> reexamine their philosophy of education and their current practice in light of changing
>
> educational theory and changing student populations (Danielson & McGreal, 2000).
>
> To be certain, all teachers must take part in formal evaluation processes and using the

same standards. However, the needs of novice teachers may require that they are evaluated

according to a limited number of teaching competencies rather than upon all indicators within evaluation systems. For example, administrators and mentor teachers can work with novice teachers on establishing effective classroom routines and basic classroom management skills in addition to basic lesson planning and implementation skills. Although Danielson and McGreal believe that beginning teachers should work to understand all teaching competencies within their first year, competencies such as development of mastery learning objectives may be developed once these more basic teaching skills are ingrained.

Likewise, teachers at more advanced skill levels may take part in more reflective, self-directed formal and informal evaluation activities that take into account multiple sources of instructional data in addition to administrative classroom observations as a means of evaluating teacher performance. These may include teaching portfolios, reflective writing, peer observations, student feedback, and video taping of lessons (Danielson & McGreal, 2000; T. J. Sergiovanni & Starratt, 1993; Starratt, 1996; Vandeweghe & Varney, 2006). A major source of evidence of instructional improvement is student achievement data. These data must reflect teachers' knowledge of individual student learning, strengths, and areas for development (Glanz, Shulman, & Sullivan, 2007). In this case, instructional improvement strategies are derived from two sources of student information. The first source is the conferences that administrators hold with teachers. As Glanz (2007) observed, administrators from select New York City Public Schools help teachers to develop individual student success plans by having them reflect upon individual student assessments and their conferencing notes for individual students. The second source of student information is derived from educator team meetings which consist of teachers, administrators, counselors, therapists, and other educators who provide direct service to students. Teams meet periodically to discuss a student's overall socio-emotional and academic status and

from there to develop a plan of action or intervention for that student (Glanz, Shulman, & Sullivan, 2007). Hence, through the examination of data from multiple sources, supervisors are able to discern teachers' instructional skill levels and personal knowledge of students while concurrently facilitating teachers' utilization of these data to address student needs. Consequently, supervisory and evaluative activities are not teacher responsibilities added to the school day, but are rather well-integrated elements of the instructional life of the school (Glanz, Shulman, & Sullivan, 2007).

Tthe use of multiple sources of information in analyzing teacher performance may differ with the experience and skill level of the individual teacher. For teachers at all levels of development, it is imperative that these data are examined in conjunction with one-another as researchers caution against the overuse of a single method of data gathering or reliance on one type of data when judging educator performance. Doing so can produce a narrow or distorted view of classroom performance and student learning (Frase & Streshly, 1994; T. J. Sergiovanni & Starratt, 1993; Starratt, 1996).

Concerning teachers at differing levels of experience and instructional skill, Danielson and McGreal examine the topic of differentiated supervision and evaluation and the concept of utilizing multiple sources of data in great depth. For example, they outline a system which utilizes three separate "tracks" for teacher evaluation which are differentiated by teachers' experience levels and instructional expertise (Danielson & McGreal, 2000) as utilized in one Michigan School District (Danielson, 1996). While differentiated in terms of support levels and criteria, in this plan all teachers are held to the following six instruction standards: Classroom Environment; Preparation and Planning; Instruction; Assessment; Communications; and Professional; Responsibilities.   This is expanded from the four teaching "domains" as outlined

in Danielson's earlier work (1996). Absent from these four domains is the aforementioned

Communications standard as adopted by the state of Michigan. In essence, these standards

reflect the RBT's Principles of Effective Teaching examined in the previous chapter. The

Principles of Effective Teaching, however, outline two additional standards which provide

indicators for assessing teachers' curriculum content knowledge and knowledge and

implementation of instructional strategies that ensure equity of education within the classroom as

well as celebration of diversity in the classroom (Department of Education, 2001).

The track system as adopted by the state of Michigan in 2001 provides for three plans for

staff development, supervision, and evaluation: Plan I – Initial Staff Development; Plan II –

Professional Growth; and Plan III – Specific Staff Development (Danielson & McGreal, 2000).

Teachers are grouped according to their knowledge and implementation of the six

aforementioned standards. Under Plan I, for example, non-tenured teachers, teachers new to the

state, and struggling tenured teachers are supported to understand, accept, and demonstrate the

five standards. The purpose of this track is not only to help teachers internalize the standards,

but also to help teachers in implementing them in the classroom and to provide data with which

evaluators make decisions for ongoing employment (Danielson & McGreal, 2000). Veteran or

master teachers who exhibit the six Standards for Effective Teaching in Michigan are supported

in the second track or Plan II. This track allows teachers and their supervisors to focus on key

areas of instructional improvement with the ultimate goal of augmenting student learning and

reaching school-wide improvement goals. While the methodology in Plan I calls for more

formal evaluative activities, such as scheduled observations and feedback, review of a

professional portfolio, and professional mentoring, Plan II relies mostly upon informal

observations, conferencing, and peer collaboration as methods for reaching school-wide learning goals (Danielson & McGreal, 2000).

Plan III provides a framework for those tenured teachers who are having difficulty meeting specifics instructional standards. Here, teachers are supported through more direct administrative observation and formal written feedback akin to those outlined in Plan I. In the event that the tenured teacher continues to struggle in his practice, Plan III ensures that the procedural due process of evaluation is followed with an eye toward disciplinary action and possible dismissal should the teacher fail to improve within the specified areas (Danielson & McGreal, 2000).

Regardless of the differentiated teacher support and evaluation activities that will be examined, according to Danielson and McGreal all teachers are held to the five standards that the authors delineate (2000). Sparks (2004) and Glickman (2006) express concern in this regard. What the aforementioned researchers consider to be differentiated supervision and evaluation Sparks sees as "professional development apartheid" in which two disparate systems of professional development and teacher supervision/evaluation exist within a school (Sparks, 2004, p. 305). The first tier closely resembles "Plan II" of Danielson and McGreal's track system (2000). Sparks defines this tier of professional development in the following way:

> The first tier is an emerging system that advocates the development of professional community and the exercise of professional judgment … teachers in schools that have embraced this system of professional development are generally committed to collective school and team goals, use data and other forms of evidence to make decisions, engage in extended study and discussions of educational issues, and enjoy the benefits of supportive, collegial interactions (Sparks, 2004, p. 304).

Here, Sparks is aligned with Danielson and McGreal (2000) and Marshall (2006) in his belief that this first tier is built on a foundation of shared leadership, collegiality, and professional analysis, implementation, and reflection (2004). This has tremendous implications for supervision and evaluation, as teachers themselves become instructional leaders and as administrators and formal evaluators share supervisory tasks with these professionals (Marshall, 2003).

In contrast to this, Sparks paints a portrait of a professional development system that reflects and relies upon a more hierarchical or Taylorist educational model. He describes this tier of professional development as one reliant upon teacher adherence to imposed instructional and curricular mandates. He indicates that professionalism does not thrive in schools where curricula and instruction that are prescribed, directive, and micro-managed through systems of supervision and evaluation. Instead, teachers are treated as automatons in professional cultures that value compliance and conformity rather than collegiality, collective professional judgment, knowledge, and creativity (2004). Glickman (2006) supports this belief by stating, "externally approved, 'research-based' programs, teaching methods, and assessments that are officially promoted by state and federal governments" sap the willingness to fail and the consequent imaginative resolutions from professional cultures (p. 690). Further, Sparks indicates that within schools and districts which implement two-tiered systems such as that espoused by Danielson and McGreal (2000), the second, more prescriptive and hierarchical tier, invariably becomes the dominant model and where "if the benefits of professional community occur at all under this approach, they are, at best, by-products for a minority of teachers rather than an intended outcome for all" (Sparks, 2004, p. 304). To be sure, Sparks does qualify his statements by demonstrating a need for directive and highly-structured curricula and instructional practices in some instances. For

example, in schools and districts that employ a more novice staff with a less sophisticated

understanding of instructional strategies and curriculum development, Sparks sees prescribed

curricula as a very good starting point and as a means of entry into in-depth professional

understanding and discussion surrounding student learning as linked to high quality instructional

practice. Sparks believes that as ends unto itself, highly directive/structured curricula and

supervision and evaluation that support them will breed teacher reliance rather than collegiality

and professional creativity. Worse, they will foster mediocre and superficial student

achievement rather than true understanding, excellence, and the power of imagination

(Glickman, Gordon, & Ross-Gordon, 2001; Sparks, 2004).

Sergiovanni and Starratt (2007) reflect this school of thought and consequently argue

against a narrow range of standards that are used to script instruction and forge a broad spectrum

instructional performance standards that are utilized in the following ways: "(1) to provide an

overview of standards that can be used to construct a definition of effective teaching; (2) to show

how standards can be embedded in practice at different levels of quality and competence; (3) to

provide a bank of standards that can be used by teachers to examine their practice alone,

together, or with supervisors; (4) to provide a bank of standards that schools can use in crafting

teacher evaluation systems" (T. Sergiovanni & Starratt, 2007, p. 179). Banks of standards

should serve as frameworks, rather than templates or checklists, for practice (T. Sergiovanni &

Starratt, 2007). Frameworks are more fluid and encompassing than templates, and provide for

teacher growth at different levels of teacher competence, as evidenced by the aforementioned

work of Danielson (1996). These scholars contend that for models of supervision and evaluation

to be successful in aiding instructional improvement, these standards must exist as frameworks

that serve both as a common language for instructional dialogue and as a baseline for teacher

practice. Templates, on the other hand, serve as checklists for teacher performance indicators that are easily observed and quantified by supervisors. They do not, however, lead to deep and powerful conversations regarding the development and assessment of individual teacher goals based upon foundational frameworks for practice.

In more practical and essential terms, the use of frameworks for instructional practice also demands that teachers and supervisors revisit and revise said standards to account for differing levels of teacher development, student learning strengths and needs, and teachers' objectives both as stated and as taught (T. Sergiovanni & Starratt, 2007). As outlined earlier, Danielson's frameworks provide areas for instructional performance within four domains (Danielson, 1996). Sergiovanni and Starratt contend that teachers and supervisors could observe and review a lesson as planned and as executed through each of these domains noting areas of strength and areas for instructional improvement according to the indicators provided, thus providing for the professional reflection and flexibility needed to examine and improve instructional practice (T. Sergiovanni & Starratt, 2007).   Should elements within the lesson exist that do not coincide with indicators in the framework, then the framework needs to be modified. Mandated templates allow for no such reflection or flexibility, nor do they account for the unending complexities inherent in classroom instruction.

The ineffectiveness of top-down mandates and the need for teacher ownership for supervision and evaluation systems are constant themes in the contemporary study of teacher performance evaluation. This is also true of the of the inexorable link that must exist between supervision and evaluation systems and teacher driven professional development (Marshall, 2003, , 2006; Vandeweghe & Varney, 2006; Zimmerman & Deckert-Peton, 2003). While the above-mentioned authors perspectives have been examined at length, Vandeweghe and Varney

(2006) relate their findings regarding the development of a "School Based Study Group" in one

urban public school in Denver, Colorado. According to Vanderweghe and Varney, several major

factors served as catalysts for the formation of the study group. The first was teacher

dissatisfaction with an administrative directive that mandated teacher research with the

implication that said research would weigh heavily in their annual performance reviews.

"Teachers saw the mandate as busywork, no one took it seriously, and the next year it

disappeared" (Vandeweghe & Varney, 2006, p. 283). The second cause for teacher uneasiness

with the system of professional development and concurrent evaluation process was the use of

educational consultants from outside of the school to demonstrate their own expertise to the

faculty and staff of the school in question. This type of professional development was an

administrative mandate and teachers saw little or no alignment between this type of professional

development and the teaching and learning taking place in their classrooms. Consequently, not

only did teachers have no ownership for their own training, but they also resented the sessions as

a waste of their valuable time. Last, teachers wanted professional development that consisted of

in-depth collegial discussions with their co-teachers around specific instructional strategies and

student learning. Teachers believed that this type of development should be integrated into and

not detached from the school day (Vandeweghe & Varney, 2006).

These elements served as the original impetus for development of a teacher-led school-

based study group which evolved over six school years. The basic tenets which served as the

foundation for the group were similar to those espoused by authors who examined reformed

supervision, evaluation, and professional development systems of the early 1990s as examined

earlier in this chapter. These include school cultures that support shared and distributed

leadership, student-centered conversation, and peer support (Black, 1993; Colby, 2002;

Danielson & McGreal, 2000; Glanz, Shulman, & Sullivan, 2007; Glickman, Gordon, & Ross-

Gordon, 2001; Marshall, 2006; Nevo, 1994; Rettig, 1999; Rooney, 1993; T. J. Sergiovanni &

Starratt, 1993; Shen, 1998).   While rooted in these basic elements, the school-based study group

in question demonstrated a profoundly thorough, yet simple approach, to the examination and

analysis of instructional practices and student learning (Vandeweghe & Varney, 2006).   The

group was facilitated by professors and clinical teaching staff from a partnering university and

composed of various school based teaching staff and other staff members (titles not specified).

The group's goal was to utilize the internal experience and expertise of the staff to arrive at

sophisticated levels of understanding of teacher behavior and concomitant student performance.

Group facilitators were certain to commence work in one well-defined instructional area for

improvement, in this case, teacher talk (Vandeweghe & Varney, 2006). Given teachers' past

experience with top-down mandates accompanied by little or no follow-up, it was of utmost

importance that the study group work to create an culture and atmosphere of trust that valued

collective inquiry and improvement over judgment and blame. To this end, group members

utilized both empirical and anecdotal data from the teaching and learning within their school as

the foundation for their discussions on understanding instruction as a precursor to improving it

(Vandeweghe & Varney, 2006). This data included both formative and summative indicators,

such as quantitative and qualitative student test data. As with Danielson and McGreal's (2000)

multi-track system, the Denver study group utilized other more anecdotal, yet at least equally as

powerful and revealing sources of data, including: group viewing of video-recorded lessons; peer

observations; professional articles; and examination of state-mandated assessments

(Vandeweghe & Varney, 2006).

In subsequent years, the school-based study group still reached out to the professional knowledge base for guidance However, they looked to implementation of various instructional strategies, which they name "tiny experiments," in order to further their understanding of teacher talk and teacher and student questioning in particular (Vandeweghe & Varney, 2006).

To be certain, school administrators must provide visionary leadership that drives the entire school community toward common instructional goals (Lashway, 2002). However, Vandeweghe and Varney speak of a system that is nearly completely owned and run by teachers. This has tremendous implications for instructional leadership in terms of teacher supervision and evaluation as differentiated supervision models demand that administrators function not as the centers, but as the facilitators of the previously examined educational reform efforts and instructional improvement (Lashway, 2002; Peterson, 2004; Prybylo, 1998; Starratt, 1996). As Starratt (1996) indicates, administrators must aim to create a trustful professional environment that fosters teacher reflection on their practice and effective open and honest conferencing among teachers and administrators. Like Vandeweghe and Varney (2006) and Danielson and McGreal (2000), Starratt warns against the counter-productivity of instructional mandates and outlines the role of administrators in differentiating supervision and evaluation practices through the leadership tenet of commitment:

> In schools administered by commitment, there is no attempt to monitor each teacher's classrooms according to a predetermined, uniform set of expectations. Rather, there is the expectation that each teacher, carefully reading the talents and interests in her class, is responding to both the curricular objectives and the complex chemistry of the human beings in her class and designs a variety of learning activities to assist in the appropriation of the material and their creation of knowledge through those activities.

Administrators do not control what goes on the classroom; instead, in their visits to classes they listen to teachers explain how they are using their professional talents and skills to create exciting learning possibilities for the children. Administrators promote and reinforce the teachers' commitment to make learning exciting, interesting, and fulfilling (1996, p. 119).

From the 1840s to the present, teacher evaluation systems have evolved from subjective processes that measured characteristics of teacher's personalities and largely reported upon readily observable management practices. As stated above, contemporary models of evaluation allow for more differentiation and flexibility in the promotion of teacher reflection and improvement of their practice. The previously examined literature demonstrates that even these models are not without their own limitations. The absence of teacher ownership for the evaluation process, the lack of administrative time and supervisory expertise in conducting evaluations, the need for effective conferencing and the use of limited sources of data that reflect a teacher's practice all contribute to the ineffectiveness of current evaluation models. When properly differentiated and facilitated, though not always led, by administrators, however, current teacher evaluation systems can promote true instructional improvement by helping educators to sift through and analyze complex human behaviors, beliefs, and interactions. From the promotion of teacher reflection to the implementation of constructs such as peer observation and portfolio assessment, teachers and administrators get a more complete picture of classroom instruction from which they discern instructional strengths that are heralded as best practices from instructional deficiencies that demand review and reconsideration. The next section of this review examines the perspectives of professional educators that outline many of these facets of teacher evaluation systems and their implementation.

Perceptions of Teacher Supervision and Evaluation

Teacher evaluation is the predominant means of instructional quality control in education (Worthen, 1987). However, outlined in the current body of professional literature surrounding the topic are several mitigating factors that inhibit use of evaluation as a means of insuring the ongoing improvement of teaching and learning. The existence of these limitations is perpetuated by the perceptions of administrators (or other evaluators), the perceptions of teachers toward prevailing evaluation systems, and how these perceptions are intensified by collective bargaining agreements and state mandates. As outlined in the previous section, the presence of these perceptions results in a lack of ownership for all parties involved thus rendering supervision and evaluation activities futile and perfunctory, yet mandatory exercises.

Blumberg (1980) describes relationships between teachers and their supervisors as a "cold war" that is perpetuated by several factors. The first is teacher and administrator perceptions of their own roles and those of the other group within the evaluation process. For example, administrators might indicate that their largest instructional challenge stems from their interactions with teachers and the unwillingness of teachers to change as opposed to their limited knowledge of new instructional techniques and curricular materials. Likewise, many teachers see supervision and evaluation activities as perfunctory at best, and damaging at worst, due in large part to supervisors' lack of instructional and curricular information (Blumberg, 1980). Blumberg (1980) contends that the interpersonal relationships and concomitant behaviors within a school are influenced greatly by the organizational structure of the supervisory system itself. The system of selecting teacher evaluators, for example, assumes that administrative certification is tantamount to competency in teacher evaluation which is not always the case. The hierarchical nature of evaluation models also serves as an impediment to collegial behavior and serves to

preserve the professional chasm between teacher and supervisor. While changing one's place within an organizational structure might change one's professional perspective, it need not change one's attitudes. According to Blumberg, however, the latter is almost always the case. He cites that as individuals are promoted from union to managerial positions, their attitudes become more favorable toward management when as union representatives, the same individuals possessed attitudes more favorable toward the union membership. This perpetuates and strengthens the dichotomous relationship that exists between teachers and supervisors, consisting of two sides of a struggle intent upon protecting their attendant territories. He also finds that evaluation systems, by both design and implementation perpetuate destructive norms for professional behavior. He argues that the absence of clear behavioral indicators and expectations for both teachers and supervisors, professionals will believe that all is well with the organization barring overt crises or complaints to the contrary. The same rings true, according to Blumberg (1980), of systems of evaluation and the role of supervisors within them. Absent calamity or outrage, the system will continue to function without review or revision.

Blumberg (1980) asserts that the basis for change within an evaluation system lies within the understanding and propagation of positive interpersonal relationships. Here, the idea of collegiality replaces that of oversight and supervisors are trained to work with teachers as part of the same mechanism that supports instructional practice. Supervisors will set goals and, understanding the interpersonal dynamics of the staff, will work to move each person toward self-motivation through attending to hygienic needs, i.e. the need for basic resources. In addition, teachers need to be acknowledged for positive and effective instructional performance as a means of leading to self-actualization and motivation to improve performance (Blumberg, 1985). These observations and assertions are reflected as well in the literature that follows.

Administrators' Perceptions. Effective teacher evaluation is not only the duty, but also the ethical responsibility of administrators (Fullan, 2003; Pratt, 1996; Starratt, 1996), however the body of literature indicates that administrators are faced with several limitations when attempting to pursue such an end. A serious lack of time is highlighted as a major obstacle toward the use of teacher evaluation as a means of improving teaching and learning. Although administrators value their roles as instructional leaders, they believe that the complexities of these roles prevent them from making instructional improvement a primary focus within their schools and districts on a daily, weekly, and yearly basis. In the Commonwealth of Massachusetts, for example, "despite the fact that principals are expected to evaluate … teachers, only 6% felt strongly that they are able to direct the major focus of their time to the improvement of student learning, and 52% said that they are not able to spend time on this" (MESPA, 1998). The Massachusetts Partnership for Schools also found that nearly 50% of the teachers it surveyed did not believe that "administrators spend enough time in classrooms to carry out district's expectations for evaluation" (MassPartners, 2000). Participant-administrators of this study cite a serious lack of time as the major obstacle to their presence in the classroom and many were willing to intimate this on an individual basis. Excerpts from focus groups and surveys revealed principals' attitudes and observations regarding their roles as evaluators and the limited availability of time for the task: "[Research for Better Teaching] says that I can evaluate seven teachers adequately, then the others lose. I am evaluating over twenty teachers" (MassPartners, 2000).

More recent literature describes exactly how the limitations of time perpetuate a belief in the superficial and perfunctory nature of teacher supervision and evaluation among education professionals. As one former Boston principal reflects:

"1) my evaluations were based on grossly inadequate information; 2) they were therefore superficial and often missed the target; 3) this was why teachers paid so little attention to them and rarely made changes based on what I wrote; and therefore, 4) spending hours and hours on this process was not a good use of my time" (Marshall, 2006, p. 3).

Marshall (2006) cites the shortage of time observing teaching and interacting with this particular teacher (less that 1% of the teacher's overall time) as the defining factor in his self-assessment. Thus, effectiveness of evaluation processes, from administrative classroom observations to the collection of artifacts and other instructional data, is not only severely hemmed in by this limitation (Meloche, 2000), but also it also supports educator beliefs that the process is largely perfunctory and one of compliance (Marshall, 2003, 2006; MassPartners, 2000).

As Marshall indicates, this trend is not confined to the Commonwealth of Massachusetts, nor is it solely a complaint of building principals. For example, Bredeson (1996) surveyed over three hundred superintendents. In this survey, he asked public school superintendents to rank nine administrative tasks first by the order of importance they placed on the tasks, and then by the amount of time they spent executing the tasks. Bredeson found that while superintendents ranked curriculum and instructional leadership (including behavior that supports teacher evaluation) as fourth in importance from a list of nine administrative responsibilities, the same duty was ranked as the seventh administrative task by the amount of they time spent on these same duties (Bredeson, 1996). Hence, while administrators acknowledge the importance and

necessity of carrying out regular and effective teacher evaluation in order to improve instruction,

they also acknowledge what they perceive to be a serious lack of time that plagues their efforts at

executing evaluative tasks (Bredeson, 1996; Frase & Streshly, 1994; MassPartners, 2000;

MESPA, 1998).

As participants in these studies indicate, the managerial tasks of school administration

dominate administrators' calendars; thus administrators are faced with little or no time for daily

supervision or evaluation of teachers. The concentration of these tasks undermines the

effectiveness of evaluation as a means of improving instruction, especially in terms of dismissing

ineffective or incompetent teachers (DeMitchell, 1995). Without regular and effective

evaluation, administrators are unable to gather evidence regarding a teacher's effectiveness in the

classroom, hence, they tend to err on the side of caution in written summative evaluations by

being vague and thus, avoiding conflict:

> Administrators often believe that they are unable to spend enough time in a teacher's
>
> classroom because of the daily press of running a school. They hesitate to be critical in
>
> the face of incomplete information (DeMitchell, 1995, p. 89).

While the lack of time is cited in these studies as a major limitation to effective and

ongoing evaluation, (Bredeson, 1996; Frase & Streshly, 1994; Goldrick, 2002; MassPartners,

2000; MESPA, 1998), this perception is listed by some administrators as an excuse, not a reason,

for not carrying out teacher evaluations (Frase & Streshly, 1994).

The findings of the aforementioned works indicate that the lack of administrative time

spent on evaluation is not a phenomenon caused solely by the enormity of managerial tasks that

confront administrators (DeMitchell, 1995; Frase & Streshly, 1994; Lashway, 2002; Peterson,

2004). These authors indicate that the failure of administrators to focus on evaluation as a means

of improving instruction is also a choice that is consciously made due to their lack of expertise in the area of evaluation. Administrators may avoid use of teacher evaluation as a means of improving teaching and learning because they lack knowledge and skills necessary to do so, especially in the case of marginal and incompetent teachers. This may include a lack of the skills necessary to document legally the progress, or lack thereof, of both marginal and outstanding practitioners (DeMitchell, 1995). Also, some administrators are reluctant to constructively criticize or to begin a dialogue on the improvement of teaching and learning because they don't believe they have the ability to assess the performance of practitioners who teach different subjects and at different grade levels than they themselves taught. One study found that twenty-five percent of principals agreed, "due to a lack of a sophisticated understanding of classroom instruction evaluation, administrators are reluctant to dismiss a teacher with professional status whose performance has declined and has failed to improve" (MassPartners, 2000).

These administrative problems hinder school administrators from creating a school culture with norms that support teachers in pursuing the goal of ongoing instructional improvement. Further, the unwillingness of administrators to carry out effective, regular evaluation stems from more psychological, personality-driven issues that dictate the administrators desire to avoid conflict (Frase & Streshly, 1994).

The aforementioned elements result in administrative practices that compound the obstacles to effective evaluation. For example, "principals often use minimal teaching competencies (associated with direct instruction) as criteria to judge teachers' performance" (Weiss, 1998). Such competencies include readily observable aspects of teacher instruction that emphasize compliance with initiatives and mandates over actual student learning and achievement. The prominent posting of daily lesson objectives, the creation of colorful

classroom displays, and impeccable record keeping all support classroom learning. When observed in isolation of student performance indicators, however, they tell evaluators nothing about the effectiveness of instructional techniques.   As Weiss (1998) indicates, this practice is futile, for outstanding teachers have already mastered the basic competencies of teaching. These behaviors result in a plethora of effects that are counterproductive to the end of improving teaching and learning, including the following consequences:

1. Evaluation ratings are inflated beyond reality;

2. Teachers … receive little substantive feedback for improvement from evaluations; 3. Professional growth plans are not aligned with personnel evaluation findings;

4. Evaluators fail to assume responsibility for teacher evaluations (Frase, 1994, p. 47).

 This is problematic, specifically because teachers and administrators alike acknowledge the value of effective professional development, conferencing, and alignment with professional standards all as paramount to the improvement of instruction. However, while large majorities of teacher and administrators in one study agree that "conferences between teachers and administrators are an important component of teacher evaluation," only thirty four percent of teachers and twelve percent of principals disagree that "conferences between teachers and administrators as part of evaluation are done well" (MassPartners, 2000). Frase (1994) contends that this is, in part, a result of the misalignment between teacher-generated professional growth plans and the conferences themselves. Consequently, the evaluation process becomes a cursory and meaningless one in which implementation of "the traditional model of pre-conference, observation, and post conference" becomes a perfunctory exercise that is owned neither by administrators nor by teachers and that fails to improve practice (Rooney, 1993).

Teachers' Perceptions & Beliefs. Teacher roles and beliefs regarding their own evaluations can likewise foster the idea that evaluation is a superficial, perfunctory act that has no bearing on instructional improvement. As with administrators, the negative effect of these beliefs causes a significant lack of ownership for teacher evaluation as a means of supporting teaching and learning. From the lack of positive and frequent feedback from evaluators, to the limited choice of evaluation options, to the tendency of marginal teachers to deny their own need for instructional improvement, the aim of using teacher evaluation as a tool for improving teaching and learning is grossly inhibited (Fuhr, 1993; Mertler, 1997; Ballard, 1998; Colby, 2002).

The examples of professional research that follow suggest that the construct of evaluation instruments themselves severely limits the usefulness of teacher evaluation in improving instruction. For this reason, "evaluation is usually perceived as a means to control teachers, to motivate them, to hold them accountable for their services, or to get rid of them when their performance is poor," rather than as a means of improving their performance (Nevo, 1994). Furthermore, evaluation instruments in some districts tend to be summative in nature and of the stereotypical "checklist" or positivist variety, and therefore encompass observation of elements that are not relevant to instructional improvement. As late as the 1990s, for example, often on a single form, "a principal is required to comment on such things as the teacher's appearance, professional development, rapport with colleagues, and instructional ability" (Rettig, 1999).

While these systems may allow evaluators to report on teacher and possibly student behavior, they do not require evaluators to make specific recommendations to teachers for improvement of their practice. Furthermore, they lack the flexibility that allows administrators to take into account several key individual instructional indicators, including student behavior

and performance, individual class personality, student needs and learning, teacher experience, goal setting, or professional development needs. In essence, sole use of summative evaluation systems does not support teaching and learning for the practice inherently denies the quintessential belief that teaching is a profession. Consequently, the basic elements necessary for the instructional improvement of teachers tend to be absent in buildings and districts that utilize summative evaluation processes alone. These include professional development that is tied directly to teacher evaluation, individual practitioner goal setting that flows from building and district goals and that is linked to teacher evaluation, an ongoing dialogue between teachers and evaluators that is consistently focused upon the improvement of teaching and learning, and most important, teacher input and judgment regarding their own supervision and evaluation (Ballard, 1998; Black, 1993; Howard, 1996; Prybylo, 1998). In the MassPartners study, some teachers affirm these perspectives, eighty percent of teachers-participants agree, for example, "Ongoing professional development in techniques of teacher evaluation is needed for evaluation to improve teaching and learning." Further, seventy percent of teacher-participants agreed that "the goal setting process required of teachers by evaluation systems at the beginning of each year should be linked to teacher evaluation." Finally, concerning teacher input into the evaluation process, forty-two percent of teachers disagreed that in their respective districts, "there is regular discussion about what [professional standards for teaching] look like in the classroom" (MassPartners, 2000).

The effect of the absence of this ongoing conversation is two-fold. First, the use of professional standards (e.g., RBT's Principles of Effective Teaching as utilized by the participating school districts in the pilot study) is ineffective if administrators do not extend to teachers conversations about the specific practical applications of these standards. Second, the

absence of this conversation demonstrates to teachers that administrators are not interested in the improvement of instruction in any practical or specific sense, nor are they willing to empower teachers to own or interpret professional standards in a way that materially and substantially improves teachers' practice. As Shen (1998) indicates, administrators may believe that they are empowering teachers to effect changes in practice by changing the organizational structures that exist in schools. School administrators must be clear regarding the place for teacher performance standards in the school and how they will be used both to evaluate teachers and to improve instruction. Without this direction, teacher-led activities such as common planning sessions and observation of peer practice are amount to little in terms of instructional improvement because interpretation of professional standards becomes arbitrary at best; thus, the improvement of instruction becomes a chance occurrence. Consequently, while administrators may believe that teachers have more leadership capacity within the organizational structure of schools, teachers do not, in fact, feel empowered to improve their practice within the context of formal evaluation structures (Shen, 1998).

Hence, researchers stress the importance of an evaluation system which utilizes both formative and summative evaluation and supervision models, that is flexible enough to address specific student learning and teacher professional improvement needs, and that allows teachers a substantial voice in their own goal setting and professional development agendas (Marshall, 2005; MassPartners, 2000; Shen, 1998). In the presence of these and the aforementioned factors, the process of evaluation becomes one that is teacher centered and owned, rather than one that is perfunctory and bereft of professional and personal meaning for teachers and administrators. Educators indicate that ongoing professional development in teacher evaluation, peer observation, and effective conferencing techniques is needed in order to maintain an evaluation

system that supports the ongoing improvement of teaching and learning (Mass Partners, 2000).

Marshall (2005) adds:

> The engine that drives high student achievement is teacher teams working collaboratively
> toward common curriculum expectations and using interim assessments to continuously
> improve teaching and attend to students who are not successful (p. 731)

Consequently, while systems such as those aforementioned do support and augment student learning and professional instructional techniques, they are extremely complex to initiate, and are as limited in their effectiveness as the resources (especially time) granted to their implementation and maintenance by administrators and teachers alike (Ballard, 1998; Dawson, 1998). This again demands that administrators maintain a strong focus on instruction in their buildings and districts in order that teachers are expected and empowered to improve their practice.

Teacher attitudes concerning their own professional remediation can also prove a formidable barrier to effective evaluation. For example, Waintroob (1995) contends that the teachers who deny their own need for remediation and professional development are often those who need it the most. Further, as Hoerr (1998) intimates, while outstanding teachers tend to be self-evaluating regardless of what system of evaluation is in place, marginal or incompetent teachers tend to criticize others. Waintroob (1995) agrees, "inevitably, the non-remediable teacher's denial that he or she has a problem is accompanied by an attack on the credibility, the competence, or integrity of the administrator." In those cases previously outlined, where an administrator either lacks the skills to determine instructional effectiveness of a teacher, is uncomfortable with confrontation, or is not willing to invest time in the complaints, remediation, or proceedings for dismissal of the teacher in question, the teacher is retained, teaching and

learning are not improved, and the potential benefits of evaluation are invariably eliminated.

Participants in the MassPartners (2000) study affirm this conclusion as teachers indicated that

principals are somewhat reticent to evaluate and to dismiss teachers with many years of service,

citing personal relationships or familial concerns as the root of their reticence. This is contrary

to the stated needs of teachers who believe that effective supervision and evaluation depends

deeply on the trust, sincerity, consistency, and evaluative expertise of the administrator (Nikolic,

2001).

Perceived Obstacles & the Disconnect Between Teachers and Evaluators

In the absence of obstacles presented by administrative and teaching practices and

perspectives, factors extraneous of the school setting often have a negative effect upon even the

best laid efforts toward improving teaching and learning through the evaluation process. These

factors tend to be rooted within state mandates, district policies, and/or collective bargaining

agreements. Due to their different roles, teachers and their evaluators necessarily perceive the

impact of these outside factors differently. This can result in a strong disconnect between

teachers and evaluators regarding both the role and efficacy of standards-based teacher

evaluation systems (Kelly, 1999; MassPartners, 2000; MESPA, 1998).

Some state mandates for teacher evaluation often have the direct opposite of their

intended effect upon teaching and learning. This is due, in part, to the way teachers and

administrators view the mandates: "When mandates are perceived to be unreasonable,

impossible, ineffective, or otherwise inappropriate, they are felt to be coercive" (Kelly, 1999, p.

543). As with state mandates for high curriculum standards and teacher and student testing,

mandates for evaluation have not improved teaching and learning because of the aforementioned

lack of ownership for them. While Kelly recognizes the benefits of state mandated evaluation

systems, he readily admits their "awkward and complex" nature, and that positive results are

derived there from only when teachers and evaluators follow the mandates (1999). And as

demonstrated earlier, educator creativity and imagination can ge stifled by this: "When problems

occurred, teachers were operating … outside of the [mandated] paradigm" (Dawson, 1998).   As

Kelly intimates, "If you want to improve the product, what you must constantly do is ask – not

tell – the workers how to do it" (1999, p.544).

Collective bargaining agreements and the practice of teacher unions and associations also

can obstruct the improvement of instruction through use of an evaluation instrument. Gerwin

(1999) reports, for example, on the growing number of school administrators in the

Commonwealth who feel frustrated by limitations imposed upon them by union contracts. This

is especially true in cases of teacher evaluation relative to the professional development, support,

and retention of non-tenured teachers. When one district attempted to implement a

comprehensive professional development and evaluation program for its non-tenured teachers in

order both to improve their instruction and to determine their candidacy for tenure, the teachers'

union grieved the "working conditions" imposed on these teachers. Consequently, the program

was altered to meet union demands, and by the district superintendent's admission, "'it wasn't

something that by [administrative] design would have been the best way to do it'" (Gerwin,

1999, p.60). This directly involves the theory that high rates of resignation of non-tenured

teachers are directly linked to poor administrative efforts at supporting improvement of their

instructional practice (Armstrong, 1999).

Study findings affirm the assertions of both Gerwin and Kelly as over half of surveys

agreed that collective bargaining agreements acutely prohibit teacher evaluation as a means of

improving teaching and learning (MassPartners, 2000). Furthermore, a large majority (seventy

percent) of principals surveyed in the same study intimated that collective bargaining agreements inhibit them for dismissing incompetent educators (MassPartners, 2000).

Research also indicates that the concept of seniority inherent in the vast majority of teacher contracts essentially negates any efficacy for improving instruction that is derived from teacher evaluation. In some districts, for example, principals may not be allowed to hire new teachers for vacant positions within their schools if a teacher with tenure and seniority in the district elects to fill the position, regardless of a principal's objections due to prior unsatisfactory performance evaluations. Such practices all but negate the significance of teacher evaluation, especially for teachers with professional status. This is especially true, reports Thomas Hoerr, in terms of the rewards afforded to teachers for superior performance. "The star teacher with 13 years of experience and a master's probably gets the same raise as the mediocre teacher with 13 years of experience and a master's degree" (Hoerr, 1998). This leads to a lack of investment in the evaluation process by the more effective practitioner who is more likely to be a self-evaluator, and less likely to own a process that rewards her and the marginal teacher similarly (Hoerr, 1998; Kelly, 1999). Some states and districts, however, have negotiated with teachers' unions to either mitigate or wholly eliminate the use of teacher seniority as a factor in teacher placement.   In cities such as Seattle, teachers are selected for employment in a given school by a team of school-based professionals (teachers and administrators) regardless of the number of their number of years of service within the system (Merrow, 2000). Merrow attributes this to forward thinking union and district officials working to include teachers in the hiring of their prospective colleagues.

When analyzed into their component parts as done above, it is clear that each of the aforementioned factors affects teacher evaluation in complex and ongoing relation to the other.

For example, administrators may not readily uphold strict standards for evaluation in cases where marginal and outstanding practitioners are offered the same rewards by contract. Likewise, marginal teachers with professional status may not comply with state or district mandates because past practice dictates that administrators will pursue an avenue of little or no recourse in terms of their remediation or dismissal.

However, when viewed in the broadest terms, two powerful themes emerge from this analysis. The first is the strong disconnect in both perception and communication that exists between administrators and teachers regarding teacher evaluation processes. The second theme is directly derived from the first, that is the lack of shared understanding among teachers and evaluators of what effective teaching looks like, in specific terms, in the classroom.

For example, most principals surveyed in the MassPartners study agreed that union contracts inhibit the use of teacher evaluation as a means of improving teaching and learning, and that the contracts inhibit administrators from dismissing incompetent teachers. However nearly half of teachers surveyed disagreed with these same assertions (Mass Partners, 2000). Anecdotal evidence generally reflects the opinions of principals and of the public, while union officials believe administrators' perceptions to be misguided: "The perception is that it's really hard to get rid of an under performing teacher and it baffles me as to why that's true. But I think that over the years it's been very difficult to have a person who's under performing [dismissed]" (MassPartners, 2000, p. 56). Still principals perceive union action and contracts to be a major obstacle to any real effort improving teacher performance through evaluation: "The union makes life intolerable for a principal whose observation and evaluation reports are not 'glowing'" (MassPartners, 2000, p. 56).

This disconnect between teachers and administrators is also evident in the findings of Shen (1998). As she points out, despite efforts at restructuring schools, while principals believe that teachers now have more opportunity to be involved in areas such as school governance and curriculum development, teachers surveyed in her study indicated that they felt no more empowered than they had before the "new" structure had been implemented. Thus, these data underscore the very real presence of the differences in perception of obstacles to effective teacher evaluation within public schools.

The lack of a common vocabulary and a shared understanding surrounding teacher evaluation is also a major theme shared within the body of literature. Danielson and McGreal (2000), for example, believe that a common vocabulary for and understanding of teaching and learning must exist in order for any meaningful instructional dialogue to take place among colleagues. They recommend that the mentors and evaluators of first year teachers spend the first year ingraining the teaching standards and their indicators as understood within the school within new professionals. The necessity for this type of indoctrination is readily apparent. For example an overwhelming majority of teaching and administrative participants in one study agreed that their evaluation systems were based upon a set of predetermined professional standards, and that they were familiar with those standards as used in the evaluation process (MassPartners, 2000). However, when pushed to reflect more deeply, the lack of a specific understanding of how these principles are made manifest in the classroom is evident. Very low percentages of participants surveyed agree that a common understanding and vocabulary regarding effective teaching existed within in their districts (Mass Partners, 2000). This directly coincides with the lack of regular discussion around teacher evaluation as reported by teacher-participants in these same districts. Likewise, Black (1993) and Rooney (1993) respectively

comment on the lack of any meaningful discussion around the improvement of teaching and learning in the schools in which they had served as administrators. In the search for new, more effective systems of evaluation, they each discovered that peer observation of teachers was an integral part of improving instruction. Also, these administrators found it necessary for teachers to hold meaningful and ongoing discussions surrounding their practice and its improvement. Black (1993) and Rooney (1993) also found it imperative that they, themselves, participate in these conversations. In doing so, they were not only able to clarify criteria for performance ratings on summative evaluations, but more importantly, teachers *and* administrators became more adept at identifying both effective teaching practices and at expanding the instructional repertoires of all of their colleagues. As these theoretical and empirical studies indicate, without ongoing dialogue, the presence of a common vocabulary alone surrounding teacher evaluation will not improve teaching and learning in the classroom. Teachers and administrators must open a dialogue in which specific behaviors that model the professional standards are examined and defined.

Conclusion

From the 1840s, the development of teacher evaluation systems in the United States has shown marked shifts in philosophy and in purpose. Earlier systems provided for little teacher growth and were based on the arbitrary criteria as selected by the evaluator, namely a state or county official. To be sure, some of these criteria reflected the instructional beliefs of the time, i.e., the teacher's ability to make material interesting and relevant to student lives and the ability of a teacher to motivate students. Though such indicators did prosper in certain states, they indicators were far from universal . In these states (New York for example), evaluation ratings were inexorably linked to teacher licensure at the time.

This idea was reinforced and somewhat largely accepted in the 1920s and 1930s as schools adopted the Taylorist model of hierarchical organization as adopted from American industry. Schools were viewed as assembly line-based factories where teachers served as the workers and students were seen as their as finished products. Consequently, teacher were treated as workers who needed constant oversight if they were to be effective. Supervision and evaluation reflected these beliefs as systems for evaluating teachers were highly efficient in terms of this model. Ironically, teacher, or worker performance ratings were not based upon the product (i.e. students) they were producing as was the case with their industrial counterparts. Rather, they were rooted in superficial and readily observable factors such as teacher dress and behavior or physical classroom appearance. Feedback to teachers was quantitative in nature and student learning was not a priority in these evaluation models as that was seen largely as the responsibility of the student.

This model remained largely unchanged until the 1970s and 1980s first with the arrival of the clinical supervision model as developed by Andersen, Cogan, and Goldhammer and subsequently with the advent of the Personnel Evaluation Standards. The standards were developed not as indicators for individual teacher performance, but as guidelines for state and local agencies to develop effective teacher evaluation systems. As a result, more evaluation systems became standardized, and states and districts sought consistent teacher performance indicators upon which to evaluate these systems. Resultant systems promoted teacher accountability by their very nature and as in the past, relied upon administrators and evaluators to observe teaching and lead individual teachers in discussions surrounding their performance. Ultimately, the duty of forming a judgment about teacher performance fell to the administrator/evaluator. In essence, the evaluation process was still administratively led;

however, the necessity for differentiated systems within the same evaluative construct was evident based upon the needs of professionals at different places in their careers became evident.

The 1990s and 2000s saw a paradigm shift regarding supervision and evaluation with state accountability, high-stakes assessments, and curriculum standards serving as the catalyst for these changes.. While administrators still provided the instructional focus and vision for schools, ideally they were no longer at the center of the instructional improvement conversation. Instead of administratively led dialogue, teachers facilitated conversations about specific aspects of student learning and teaching techniques with an eye toward whole-school improvement. Teacher evaluation standards became the common language for instructional improvement in schools and districts and ideally teachers collaborated to define specific behavior that reflected these standards. In many instances, all administrative supervision became "informal" and as shared leadership became a behavioral norm, teachers supervised themselves through peer observation and the in-depth dialogue that followed. As a result, much of the research from this period demonstrates that the line between professional development and supervision and evaluation is almost nonexistent as teachers work collegially to improve instruction based upon their own students' learning and their own practice.

As evidenced by the literature examined, creating a culture of reflection-based instructional improvement as outlined above results in dramatic increases in student performance over time. However, as demonstrated literature on teacher and administrator perceptions of evaluation systems, many obstacles still prevent the ideal collegial framework for instructional improvement from becoming a reality in many schools and districts. These perceived obstacles include sufficient time, evaluative expertise, state and local educational mandates, collective bargaining agreements, and a lack of common understanding between teachers and

administrators. The literature clearly demonstrates, however, that these obstacles must and can

be overcome in order for a healthy culture of instructional improvement to grow. This includes

the implementation of an evaluation system that is truly effective as a means to improving

teaching and learning.

Chapter 3: Research Design

Research for the present study was conducted using both qualitative and quantitative methods. In this chapter the researcher offers a detailed overview of the research design. The chapter begins with a restatement of the research question for this study. This is followed by a presentation of the research methods that were employed to gather data. Methods included an initial focus group of urban secondary teachers for the purposes of refining the survey instrument and distribution of said instrument to the population of urban secondary teachers. The theoretical rationale for selection of the aforementioned methods as well as for participant selection is also outlined.

The present study was modeled after The Study of Systems for Evaluating Massachusetts Teachers (MassPartners, 2000). In particular, the survey instrument from the aforementioned study served as the prototype for that used in the present study. Next, selected methods methods of data gathering and analysis are described. Last, the researcher cites the statistical and practical limitations of the present study.

Questions for Research

The primary research question for consideration in this dissertation is, What are teachers' perceptions of evaluation in an urban high school that uses a standards-based evaluation system? The researcher sought to answer this question in terms of teacher familiarity with the stated standards and their indicators and the extent to which the standards are utilized as the common language for evaluation activities and instructional improvement within a school.

There were three resultant statistical hypotheses that the researcher tested in this study. The first of these (H_1: $\mu_1 > \mu_2$) was: Teachers who are familiar with the language and associated

activities of a standards-based evaluation system perceive significantly higher levels of instructional improvement from the process than those who are not familiar with the language and function. The corresponding null hypotheses (H$_0$: $\mu_1 = \mu_2$) then, was stated as: There is no significant difference in the perceived instructional improvement of teachers who are familiar with the language and associated activities of a standards-based evaluation system and those who are not familiar with the language and function of a standards-based evaluation system. The second statistical hypothesis (H$_1$: $\mu_1 > \mu_2$) was: Teachers with ten or fewer service in the district perceive that a standards-based evaluation system has more of an impact on instructional improvement than do veteran teachers. The consequent null hypothesis (H$_0$: $\mu_1 = \mu_2$) was: There is no significant difference in perceptions of the impact of a standards-based evaluation system on instructional improvement between teachers with ten or fewer years of service and teachers with more than ten years of service.

The researcher's third and final hypothesis (H$_1$: $\mu_1 > \mu_2$) was: Teachers without graduate degrees more strongly agree that a standards-based evaluation system has a positive impact on instructional improvement than do teachers with graduate and post-graduate degrees. Accordingly, the null hypothesis H$_0$: $\mu_1 = \mu_2$) was stated: There is no significant difference in perceptions of the impact of a standards-based evaluation system on instructional improvement between teachers without graduate degrees and teachers with graduate and post-graduate degrees.

Research Methodology

The researcher utilized mixed methods in gathering data for the present study. Data collection took place in two phases: 1) an initial focus group to refine the survey instrument; 2) distribution of a survey instrument that utilized a Likert attitude scale and open-ended response

questions. The Likert-based research instrument served as the primary means of data collection in the present study. As a result, qualitative data collected from open response survey items were utilized to confirm, refute, and/or clarify the findings derived from the attitude scale.

Survey Instrument Development. The researcher had not discovered an existing survey instrument that quantitatively measured teacher beliefs about the effectiveness of instructional performance standards. As aforementioned the researcher therefore created a survey instrument modeled after that utilized in the MassPartners (2000) study (Appendix C). Thus, the Standards-Based Evaluation System Attitude Scale*,* was developed to address the present study's research questions. Areas for modification included: 1) participant demographics; 2) participant familiarity with the participating district's system of formal evaluation; 3) participant familiarity with instructional performance standards within the aforementioned system; 4) participant perceptions about the effectiveness of the language of the standards as a means of improving instruction; and 5) perceptions of the effectiveness of evaluation and evaluation activities as a means to improving teacher practice. For the purposes of maintaining participant attention and interest, the entire instrument was limited to twenty-one Likert-based items and four open response items.

Letters of invitation from both the researcher and the school principal describing the study, its purpose, and its significance served as cover sheets for the survey. These letters thanked the teachers in advance for their participation in the study.

While demographic items are necessary parts of any instrument, they were limited in number, and supplied information essential to the answering of the research questions. Survey participants were asked to identify their gender. For the purposes of subsequent analysis, teachers were asked to signify their years of service within the participating district by selecting

from a list of year-range choices. They were also asked to enumerate the years of service within

the building where they presently serve, as well as their teaching assignment within the current

building. Last, participants were asked for their level of education or degree level, and their

primary area(s) of instruction. This information was used to compare levels of significance in

participant responses to survey items based upon each teacher's primary area of instruction,

years of service within the district, years of service within the current school building, , and

education level.

The survey instrument itself (see Appendix D) was divided into five sections, each aimed

at gaining information germane to the problem statement. Section A, entitled "The Seven

Standard Areas for Teacher Performance," gauged teacher familiarity with the instructional

performance standards as part of the evaluation process. Items 1 through 6 were offered in this

section. The second section, Section B, of the instrument, "Evaluation Activities and

Communication," sought to measure the extent to which the standards were utilized as a

common language for evaluation. In this section items 7 through 14 assessed the regularity with

which the standards and their descriptors were used as a common language in all written and oral

evaluation and instructional activities.

Section C of the survey, "Evaluation and Instructional Improvement," consisted of seven

survey items, three of which were adapted with permission from the *Professional Evaluation

Systems Survey* (Zimmerman, 2003). Items 15 to 21 measured the extent to which the seven

standard areas for performance have influenced instructional improvement within the building.

This was gauged by ascertaining the specific impact teachers believed that use of the standards

had upon their individual teaching practice. Section D of the instrument, contained four open

response items. Here, respondents were asked to provide information on the strengths and

weaknesses of their district's evaluation system. Participants were also required to comment upon other aspects of the system or experiences that they believed were relevant to the present study.   Data obtained from this section were used to clarify points of the survey data analysis. The fifth and final section of the instrument contained five items that asked for participant demographic information. This section was situated at the end of the survey in order to diminish influence of the chosen demographic items on individual participant responses.

Qualitative Methods. The very nature of the research question and the data it was intended to uncover demanded a mixed methods research design. Focus group data were integral to this design. Therefore, prior to survey distribution, a focus group of select participants was conducted in lieu of individual interviews. The purposes for this decision were threefold: 1) to gain insight into participant perceptions of the study instrument; 2) to probe focus group perceptions of the language of the instrument and of the teacher evaluation process; and 3) to gather evidence of teachers' experiences and perceptions of the evaluation process itself. Focus group data provided anecdotal information from which biases inherent in the survey instrument could be eliminated. The four open-ended responses on the survey instrument also served as a participant "check-in" by allowing individual participants to address concerns that were not measured or addressed by the Likert items.

The primary research question, "What are teacher perceptions of evaluation in an urban high school that uses a standards based evaluation system," required that a phenomenological approach be taken toward gathering qualitative data. Gay and Airasian (2000) believe the purpose of qualitative research is to "capture the human meaning of social life as it is lived, experienced, and understood by its participants" (p. 201). Consequently, they define a phenomenological approach to qualitative research as one that helps to answer the question,

"What is the experience of an activity or concept from these particular participants' perspective[s]" (Gay and Airasian, 2000).

Fontana and Frey's argument that the efficacy of the "group interview" as a means of data collection over that of the individual interview coincides with Gay and Airasian's aforementioned belief. "The use of the group interview is not meant to replace individual interviewing, but it is an option that deserves consideration because it can provide another level of data gathering or a perspective on the research problem not available through individual interviews" (Fontana and Frey, 1994, p. 361). In support of that belief, Fontana and Frey argue, "the group interview has the advantages of being inexpensive, data rich, flexible, stimulating to respondents, recall aiding, and cumulative and elaborative, over and above individual responses" (Fontana and Frey, 1994, p. 365). Though not completely synonymous with a group interview, the interaction between and among focus group participants yields rich and multidimensional data that is not readily collected from individual interviews. As a result, analysis of the data yielded from this type of research was invaluable to the researcher in seeking illuminating qualitative feedback on development of the survey instrument.

Participants & Sampling. The first phase of research took place in School A (see Table 1) within the selected urban school district. With the permission of the school principal, an invitational flyer (Appendix F) was distributed to all full time teaching professionals within School A via their school mailboxes. The researcher sought a maximum of eight professionals to participate in the focus group and focus group members were selected on a first-response, first-to-participate basis. Four full-time teaching professionals expressed interest and participated in the focus group. Prior to the one-hour focus group, the researcher provided each participant with a copy of the draft survey and the research questions. The group protocol (Appendix G) focused

on participants' perceptions of survey item clarity, their pertinence to the research questions, the

scope of the items, and consistency of item structure. Further, members of the group served as a

"panel of qualified experts in the content domain" who reviewed the content of the attitude scale

items to ensure validity of content (Crocker & Algina, 1986). Even though participants were

drawn from a convenience sample as stated above, group members did represent a cross section

of the teaching staff in School A. Table 1 provides some descriptive and demographic data of

the focus group participants that illustrate this cross-section.

Table 1

Focus Group Participant Demographic and Descriptive Information

Group Member	Race/Ethnicity	Gender	Content Area	Years in School A	Education Level
Participant 1	Caucasian	F	English/Language Arts	2	M.A.
Participant 2	Haitian	M	Mathematics	15	M.A.
Participant 3	Latino	M	Spanish	2	M.A.
Participant 4	Caucasian	M	English/Language Arts	32	M.Ed.

The focus group made several observations and recommendations regarding the both the

overall survey and individual survey items. The group first reacted to the clarity of the survey

items. Group members believed that the survey items were clearly stated and that they knew

what was being asked of them. The group had some concerns regarding the relationship between

certain survey items and the research question. For example, the initial draft of the survey

instrument contained a series of items regarding obstacles to effective evaluation posed by

collective bargaining agreements. To a member, the group did not believe that this section

served to answer the research question: What are teachers' perceptions of evaluation in an urban

high school that uses a standards-based evaluation system? They believed that said section asked

teachers their perceptions of collective bargaining agreements, not of standards-based teacher

evaluation. Consequently, the researcher removed this section from the final draft of the survey

instrument. Group members then commented on the demographic section of the instrument.

Participant 1 (see Table 1) believed that the demographic section should be limited so as not to

include items on race and gender. This participant intimated that she took offense to the items

and that their presence may affect the responses of similarly sensitive participants. The three

other focus group members did not agree with Participant 1; they did, however, suggest that the

demographic section be relocated to the end of the survey in order to limit the concerns of

Participant 1. As a result, all demographic items remained, however the section was relocated to

the end of the instrument.

The sample for the second phase of research was drawn from the teaching staff of a pre-

selected large high school within the same urban school district. The school has been previously

identified as School B. All participants were full time teachers under contract within said

district. As labor contracts and structures for professional evaluation differ among bargaining

units, to avoid complication, staff assistants, teacher aides, and other part time or

paraprofessional employees were not considered for participation in this study.

To determine an appropriate sample size, the table of "Sample Sizes (S) Required for

Given Population Sizes (N)" was consulted (Gay and Airasian, 2000, 135). The population of

full time teachers within the selected school is 80. Here, Gay and Airasian stipulate, "for smaller

populations, say N=100 or fewer, there is little point in sampling; survey the entire population"

(Gay, 2000, 132). Consequently, no sampling was done and the entire full-time teaching staff of

School B was invited to participate in the quantitative segment of the present study. Of that

population, 30 of the 80 full-time teaching staff agreed to and did participate in the survey portion of the present study.

   Research Site Selection and Entry. As aforementioned, this study was conducted in two secondary schools within a large urban district. School B in the urban district was chosen as the principal site from which to gather survey data for the present study. School and district demographic information are presented below in Tables 2 and 3.   Both tables provide descriptive data that underscore the urban characterization of the participating district and schools. Table 2 outlines and compares student demographic data among the district, Schools and B, and the Commonwealth. These include: percentages of minority students; percentages of students from low socioeconomic backgrounds; the percentage of special education students; and the percentage of students who have limited English proficiency. Table 3 provides professional teacher information once again compared at the state, district, and school levels. This includes: the number of teachers; licensure information; highly-qualified status; and student/teacher ratio.

Table 2

District and School Student Demographic & Enrollment Indicators

Descriptor	District	School A	School B
Total No. of Students	56,163	370	1,305
No. of Full-Time Teachers	4,275	24	80
Student Race by Percentage			
African American	39.3	36.4	40.2
Asian	8.5	23.2	8.5
Hispanic	36.7	21.1	49.4
Native American	0.4	0.3	0.7
White	5.0	18.6	13.4
Native Hawaiian, Pacific Islander	0.1	0.0	0.1
Multi-Race, Non-Hispanic	1.5	0.5	0.7
% of Students with Low Income Status	72.7	65.1	82.4
% of Special Education Students	19.7	16.9	17.7
% of Students with Limited English Proficiency	18.3	15	18.3

Note. Massachusetts Department of Education School and District Profiles Enrollment/Indicators Report. Massachusetts Department of Education, retrieved February 24, 2008 from http://profiles.doe.mass.edu/.

Table 3

District and School Teacher Data

Descriptor	District	School A	School B
Total No. of Teachers	4,275	24	80
% of Teachers Licensed in Teaching Assignment	94.2	79.2	96.3
Total # of Teachers in Core Academic Areas	3,297	22	61
% of Core Academic Teachers Identified as Highly Qualified	92.1	68.9	98.1
Student/Teacher Ratio	13.2:1	15.6:1	16:1

Note. Massachusetts Department of Education School and District Profiles Teacher Data Report. Massachusetts Department of Education, retrieved February 24, 2008 from http://profiles.doe.mass.edu/.

The data provided in Tables 2 and 3 are significant as the primary intent of the 1993 MERA was to provide equity of access to quality educational resources within the Commonwealth's school districts. As held by the Supreme Court of Massachusetts in 1993 in *McDuffy v. Secretary of the Office of Education*, the system of educational funding that existed in the Commonwealth prior to the passage of MERA had proven grossly inadequate and inequitable, especially for the representative districts with lower socioeconomic demographics and diverse racial-ethnic student populations as those outlined above. The Court further held that this inequity of funding translated directly into the lack of quality classroom instruction in these typically urban districts:

"In their 1991 report of the Committee on Distressed School Systems and School Reform, the defendant members of the board speak of a 'state of emergency due to the grossly inadequate financial support,' and admit that 'certain classrooms simply warehouse children at this time, with no effective education being provided'" (North Eastern Reporter, 2nd Series, 1993, 176, *McDuffy v. Secretary of the Office of Education)*.

In addition to all of these indicators, the foremost factor in selection of this district as a research site was its committed use of a standards-based teacher evaluation instrument and investment in substantial professional development for teachers and administrators on standards-based instruction and evaluation. In essence, these local standards reflect the Principles of Effective Teaching as disseminated to Commonwealth public school districts. However the seven standards and the respective evaluation tool used in the schools being studied were bargained at the local level to accommodate the specific needs of the district.

As part of the massive MERA initiative, the Principles of Effective Teaching were established as one vehicle through which the quality of classroom instruction could be monitored

and improved in an ongoing way, especially in districts with diverse student populations. Again, School B resides within a district that is characteristic of those addressed by the Commonwealth High Court's 1993 ruling. This is the primary reason the researcher chose School B's population of full time teaching professionals served as the pool of study participants.

The schools' headmasters granted first verbal and then written permission for research to be conducted within their schools. A copy of the letter from each headmaster urging honest and open participation on the part of participants accompanied each invitation to participate in the study. Upon completion and acceptance of the dissertation proposal, implementation of the design as outlined above began in earnest.

Participant Confidentiality.   For the purposes of all original research methods described herein, the researcher abided by the "Ethical Standards of the American Educational Research Association":

Informants and participants have a right to remain anonymous. This right should be respected when no clear understanding to the contrary has been reached. Researchers are responsible for taking appropriate precautions to protect the confidentiality of both participants and data. Those being studied should be made aware of the capacities of the various data-gathering technologies to be used in the investigation so that they can make an informed decision about their participation. It should also be made clear to informants and participants that despite every effort made to preserve it, anonymity may be compromised. Secondary researchers should respect and maintain the anonymity established by primary researchers (American Educational Research Association. 2003. Ethical Standards of the AERA [On-line]. Available: http://www.aera.net/about/policy/ethics.htm.).

In accordance with these standards, participant confidentiality was guaranteed through measures outlined below and in the "Data Gathering Procedures" section of this chapter.   Confidentiality was offered to participants as a mode of facilitating the accuracy and veracity of their responses. The sample from this study is drawn from the population of full time teachers within School B. The researcher offered the headmaster and school professionals, including teachers, a full report of data and findings upon completion of the present study. Consequently, complete participant confidentiality was offered to all participants in order to allay concerns of professional consequences that may have arisen from potentially controversial observations. This maximized the level of trust between the researcher and the participants, and aided in the gathering of the most accurate data available regarding teacher perceptions of their evaluation system.

Protecting the personal and professional privacy and mental health of all participants was of paramount importance to the researcher.   While the protection of participant confidentiality remains the ethical responsibility of all researchers, in this case it promoted the integrity of data collection and analysis. As such, all who participated in the study were assured that the researcher took sufficient measures to protect their privacy and confidentiality. These measures were reported in the letter of informed consent that all participants were required to sign (see Appendix E).

Participant confidentiality was maintained throughout the collection of and reporting upon both qualitative and quantitative data. For those professionals who completed the survey instrument, confidentiality was protected in several ways. First, participants were not required to provide any specific vital information (e.g. name, date of birth, employee identification or social security number, race, etc.) in the demographic section of the survey. Confidentiality was an essential factor in survey distribution as well. The principal researcher addressed the teaching

staff of School B at the end of a school-wide professional development day. The headmaster was not in attendance during this part of the meeting. The principal investigator explained the study objectives and procedures, including the informed consent document. After the meeting, the researcher provided a sealed participant packet to each staff member present. Those who chose to participate in the study returned the completed packet to the investigator via inter-office mail. Each packet also contained a pre-addressed and postage paid post card. Each participant was asked to write his or her name and email address on the postcard and to mail it to the researcher separate from the survey packet. By process of elimination, the researcher was able to send follow-up invitations to those faculty members who had not yet completed the survey packet. The post cards were destroyed immediately following completion of the study.

The confidentiality of focus group participants was also protected during data collection and reporting. Each teacher in School A (approximately 30 in total) received a copy of a flyer in his or her school mailbox accompanied by a letter of endorsement from the school headmaster as invitation to the focus group. Interested teachers responded to the researcher via email. The researcher communicated focus group meeting time and location information via email only to those who replied to the invitation.   The headmaster of School A was excluded from any part of the recruitment process, and participant identities were not shared with the headmaster. These measures ensured that the rest of the school does not readily know participant identities.

Last, the researcher took all measures necessary to limit or eliminate identifying the specific schools and district in which data were collected. For example, the use of identifiers such as "School A" and "School B" has already been employed, and the district has heretofore been referred to as simply, "a large urban district" or "District A". The researcher also limited

the reporting of potential identifying characteristics to those that are absolutely essential to the integrity of the present study.

Pilot Study. In the summer of 2000, the Massachusetts Partnership for Schools (MassPartners) commissioned the Study of Systems for Evaluating Massachusetts Teachers. The Mass *Partners, or Massachusetts Partnership for Public Schools, is self-described in the following way:*

> [MassPartners is a] collaborative venture of the: Massachusetts Association of School Committees; Massachusetts Association of School Superintendents; Massachusetts Elementary School Principals' Association; Massachusetts Federation of Teachers; Massachusetts Parent-Teacher Association; Massachusetts Secondary School Administrators Association; and the Massachusetts Teachers Association. (MassPartners, 2004) http://www.masc.org/PDF%27S/child.pdf.

The principal researcher of the present study served as the primary researcher for the aforementioned study as well. Therefore, with permission from MassPartners, the *Study of Systems for Evaluating Massachusetts Teachers* served as the pilot to the present study. In essence, the purpose and design of the pilot study were replicated in the present study and the findings of the pilot study influenced both the scope and methodology of the present study.

The MassPartners believed that systems for evaluating teachers could be a valuable tool for the improvement of teaching and learning. The purpose of their study was to uncover data that either supported or refuted the hypotheses that teacher evaluation is not used to its fullest potential, is sometimes not used at all, and/or is viewed as a perfunctory obligation by both the evaluator and the person being evaluated.

Through focus groups and interviews, and through the distribution of a survey instrument (see Appendix) in three Commonwealth districts, the principal researcher for MassPartners sought to provide answers to the following research questions:

- Is the evaluation of teachers being carried out in an effective manner, i.e. in a way that substantially supports the improvement of teaching and learning;

- If the evaluation of teachers is not being performed in an effective manner, then according to teachers, administrators, and other evaluators, what are the specific factors that inhibit the use of teacher evaluation systems to improve instruction?

As summarized below and recorded in the unabridged *Study of Systems for Evaluating Massachusetts Teachers,* the MassPartners reported findings in three key areas regarding teacher evaluation in the Commonwealth.

During the summer of 2000, focus groups and interviews were conducted in participating Commonwealth districts which were selected by the MassPartners by MassPartners based upon their diverse demographic and socioeconomic characteristics.   These districts employed the Principles of Effective Teaching and the RBT model for teacher evaluation as the basis of their teacher evaluation systems.

The survey instrument employed in the pilot study was developed from consistent themes that emerged from the aforementioned focus groups and interviews conducted during the pilot study.   The survey was distributed to a random selection of twenty percent of teachers, other evaluators (i.e. any professional conducting formal teacher evaluations other than central office administrators and school principals), and to all administrators in the participating districts. A total of 184, or thirty one percent of the surveys were returned for analysis.

Survey response data revealed distinct yet common patterns relating to: 1) familiarity with the Principles of Effective Teaching and the district evaluation process; 2) the evaluation process and its relationship to improving teaching and learning; 3) conferencing and dialogue as part of the evaluation process and; 4) obstacles to dismissing ineffective and/or incompetent teachers. In addition the data also revealed a significant readily observable perceptual disconnect between teachers and their evaluators in virtually all aspects of teacher evaluation. Regarding the evaluation process and its relationship to improving teaching and learning, large majorities of teachers and their evaluators agreed that they were familiar with the Principles of Effective Teaching and that the evaluation system used in their district was based upon these principles. Significant differences in perception became evident, however, concerning a common understanding of behavior that reflected the principles. Sixty-one percent of principals, for example, agreed that regular discussion took place among teachers and principals about what the Principles of Effective Teaching look like in the classroom. Conversely, only thirty-six percent of teachers agreed with this belief.

The MassPartners study revealed a similar disconnect concerning the evaluation process and the improvement of teaching and learning. As quantitative data illustrated, most teachers and administrators agreed that the annual goal setting process required of teachers by their district evaluation system should be linked to teacher evaluation. While teachers and administrators spent significant time developing these goals, administrators rarely or never revisited them as part of teacher evaluation.

This pattern is replicated by respondents regarding the purpose and function of teacher evaluation and other evaluation activities. For example substantial numbers of teachers indicated that the formal evaluation process aids them in improving their practice. These same teachers

claimed to find fault, though, with the implementation of teacher evaluation. They reported that administrators did not spend enough time in their classrooms to carry out their district's expectations for evaluation, nor could they articulate specific criteria for performance ratings, i.e. "Meets/Does Not Meet Expectations."   Further, a large majority of teachers surveyed (90%) believed that conferences between them and their administrators were a key component of teacher evaluation. Fewer than half of the respondents (41%) thought that evaluation conferences with their administrators were conducted effectively. Thus the MassPartners reported that administrators must engage in ongoing professional development around teacher supervision and evaluation.

The findings of the pilot study supported the MassPartners belief that in practice, most teachers surveyed perceived formal evaluation as a superficial, perfunctory exercise that is carried out in large part to satisfy bureaucratic mandates for compliance. Response data revealed other trends, however, that have far reaching implications for the present study and for teacher evaluation in general. First, the study revealed that most teacher participants possessed the fundamental belief that teacher evaluation and all of its corollary activities, can be a very effective means of improving teacher practice and consequently, student learning. This is especially true of annual goal-setting processes, effective conferencing models among teachers and administrators, and the attainment of a common language around good teaching as well as a school wide understanding of behavior that reflects that language.

In practical terms, however, teachers believed that their experiences with formal evaluation did not live up to their own expectations. In particular, the causes of this perceived lack of effectiveness ranged from limited administrative follow up on annual teacher performance goals, inadequate time allocated for classroom observation and ongoing conversation around individual

teacher practice, and continuing professional development for administrators in effective evaluation of instruction.

The majority of participating administrators articulated their belief in the efficacy of formal evaluation as a key means of improving instruction. Their perceptions on all three factors mentioned above are almost diametrically opposed to those of teachers. This perceptual rift served to erode professional confidence in standards-based evaluation as a means of improving instruction. The principal investigator of the present study sought to replicate in part the MassPartners study in order to ascertain teacher perceptions of standards-based evaluation within one large school. By limiting the present investigation primarily to one urban high school, the researcher aimed to uncover more detailed perspectives from teachers regarding their evaluation system and its implementation.

Methods of Data Analysis

Both the quantitative and qualitative (open response item answers) data gathered in Phase II of the study were examined through use of Windows-based data analysis applications. The data gathered through the Likert-based *Teacher Evaluation System Attitude Scale* are descriptive, categorical data. Participant responses gathered from survey items 1-21 were entered into a custom database developed using the SPSS 10.0 statistical software package. The principal investigator utilized the SPSS package to perform pertinent statistical analysis of participant responses. In terms of determining reliability of the instrument, the researcher ascertained its internal consistency by using *Cronbach's alpha.* This allowed the investigator, "to estimate the reliability of a composite when we know the composite score variance and the covariances among all its components" (Crocker & Algina, 1986, p. 117). Members of the focus group in

Phase I of the present study served as the panel of experts who determined the construct validity of the survey instrument.

Specific statistical operations were performed on the data gathered from all Likert-based items. Once the data were input into the SPSS database, the principal investigator produced a summary of the data including the count and percentage of response frequency for each item by generating a Table of Frequency Results. Next, a chi-square analysis was conducted to indicate differences in response to specific Likert items among groups of respondents. These groups included all demographic indicators outlined in Section E of the survey instrument, including years of service, race, gender, instructional field, and education level. The chi-square analysis was preferred to that of a *t-test* because the data yielded from the present instrument were discreet and categorical, and in order to be scientifically relevant, *t-tests* are normally conducted using continuous data. The researcher had to account for an inflated overall alpha level resulting from multiple chi-square tests by sustaining statistical meaning at an alpha level of .002 (i.e. dividing the desired .05 alpha by the number of Likert items, in this case, 27). In addition, the researcher planned to generate statistics for individual and groups of respondents on the subscales, or groups of similar questions (in this case, items from Sections A, B, & C respectively, of the instrument). However, a factor analysis revealed only one factor present in the study. As a result, statistics included the mean, standard deviation, and *t-tests* for comparing mean responses among participants for groups of questions are not reported in the study findings.

The qualitative data uncovered from Section D, the open response section of the survey, were analyzed through use of *Hyper Research* data analysis and encoding software. The written responses to survey items 22-25 from Section D of the survey were transcribed, verbatim, into Microsoft Word 2003. These data were then imported into the *Hyper Research* program where

they were coded according to categories that emerged from data analysis. The categories were determined by the, but reflected the exact, or *in vivo,* language of the participants. Once coded, the data were analyzed through two lenses. First, the primary researcher determined the relationship, if any, the qualitative research bore to the findings of survey items 1-21 (i.e., data that either support or refute the survey findings). Second, the data were analyzed to reveal teacher perceptions or attitudes about evaluation that were not found through analysis of the qualitative data.

Formats for Reporting the Data

The data gathered from each phase of the present study were presented according to the type of data collected. The data collected in the first and third phases of the study are qualitative in nature, and were coded as previously described. These data are subsequently presented in narrative summary format according to those codes that emerged from the focus groups.

The data gathered from the attitude scale survey instrument were categorical and discreet in nature, Therefore the researcher presents most of the research data in table form. In addition, a scree plot was utilized to display the results of the factor analyses presented. Bar graphs were also utilized as a means of presenting data, for example, when illustrating response frequency for a given item, or mean responses for groups of survey items.

Study Limitations

There are several limitations that are inherent in the present study. First, the researcher recognizes those limitations that are intrinsic to survey research. As Crocker and Algina state: "Because psychological constructs are abstractions which can only be assessed indirectly, the design of instruments to measure such variables presents several challenging problems" (1986, p.5). First, the scope of the survey instrument employed was limited in terms of the attitudes that

it was designed to measure. In other words, the items presented on the survey instrument could not account for all possible teacher attitudes concerning standards-based evaluation. For example, the instrument measures the attitudes of teachers regarding a common understanding of teacher performance rating criteria; however, the instrument does not measure teachers' understanding of or agreement with the criteria.

The second limitation stems from the design of the data collection process itself. As Crocker and Algina state, "Most psychological measurements are based on a limited sample of observations and usually are taken at one point in time" (1986, p.6). As surveys in the present study were administered after the end of a regular school day, participant fatigue and preoccupation with the day's events, for example, likely affect participant attention, investment, and response. For example, a teacher receiving a poor performance rating on the day of survey administration might respond to items differently in the morning than in the afternoon after the rating had been delivered.

The study was further restricted not only by the number of survey respondents (30), but also by the response rate (30/80).   This narrowed the ability of the researcher to examine item response patterns among demographic groups, i.e. participants of different education levels, races, genders, and years of service.   Last, a factor analysis revealed only one factor present among survey items. As a result, the researcher did not perform statistical operations which compared responses among survey sections A, B, and C.

Conclusion

The investigator sought to collect data from participants regarding their perceptions of the standards-based evaluation system used to assess their performance as professional teachers. Qualitative data would provide rich descriptive data that while powerful, could not be used to

draw general conclusions to an entire school setting. Likewise, discreet and categorical data

upon which statistical operations could be performed provide a broader array of responses, but

limit areas for participant response to those on the survey instrument. It is for these precise

reasons that the researcher chose to utilize both written quantitative and qualitative methods of

investigation in the present study. Through use of this design, the researcher provided for

respondents' written clarification of the statistical data presented. The researcher presents these

findings, their analyses, and significance, in the chapter that follows.

Chapter Four: Research Findings

Introduction

In this chapter and the next, the researcher presents the statistical findings of the present study and then situates these findings within the body of literature examined above. Chapter 4 is comprised of several sections. In the first, the researcher reviews characteristics of the data collection site, including descriptors of the participating school and its home district. Next, as this study is quantitative in nature, the researcher provides scale information, i.e. statistics and figures concerning the reliability of the survey instrument, as well as a principal component analysis. Cronbach's alpha was used to determine scale reliability. The principal component analysis was conducted to determine the number of components being measured by the scale. Next, the researcher examines descriptive statistics derived from data analysis as presented in table format. Inherent in this section is a discussion of the data in terms of the research question and the researcher's stated hypotheses. This discussion is supplemented by qualitative data gathered from survey instrument written response items. Limitations of the present study are examined at the close of this chapter. Following is Chapter Five within which the study findings are summarized and are then situated within the body of literature as reviewed in Chapter Two. Chapter Five concludes with a discussion of the theoretical and practical implications of the study at hand as well as recommendations for further study.

Site & Participant Characteristics

The present study was conducted in a large urban high school in Massachusetts. Thirty full-time teachers participated in the survey. Participants were asked to provide information on several demographic items, including race, gender, education level, primary program area, and

years of service within the participating school and district. Table 4 outlines specific

demographic information for the participants.

Table 4

Survey Participant Demographics

Demographic	# Out of 30 Participants	Percentage of Sample
Participant Race by Number		
African American	2	7%
Caucasian	21	70%
Latino	1	3%
Not indicated	6	20%
Years of Service in District		
0 – 10 Years	21	70%
10 + Years	9	30%
Gender		
Female	21	70%
Male	8	26%
Education Level		
BA/BS	5	17%
MA/MS	25	83%
Primary Program Area		
Core Subjects	17	57%
Special Education	5	17%
Electives	8	26%

As noted in the preceding chapter and as indicated above, a total of 30 full-time teaching

professionals from School B chose to participate in the survey portion of the present study. This

has significant implications concerning data analysis and the descriptive statistics from which it

was generated. These implications are discussed in further detail in the section of this chapter

outlining the study limitations.

Scale Analysis

The researcher conducted several statistical tests in order to determine the reliability of

the survey instrument utilized. A principal component analysis (PCA) was conducted in order to

assess the number of components being tested. Initially, it was hypothesized that the survey

would yield three separate factors. After the initial scale examination (outlined below), however,

it was determined that the survey measured one prominent component. A second PCA was

conducted after the researcher determined that data from two test items should not be included in

any of the principal component or reliability analyses or statistics. The results of the first PCA

are represented by the scree plot shown in Figure 1 below. As in most cases of this nature, this

plot indicates that the first PCA extracted six components; however, the plot shows that, relative

to the first component, the eigenvalues of the next five components extracted are substantially

smaller than the first. This is displayed in more detail in Table 5 below which represents the

initial eigenvalues for the first PCA. The table illustrates that the first component extracted is

dominant as it accounts for over 38 percent of the total test variance with the next closest

component accounting for only 13 percent of the total variance. At first glance, these data

provided evidence that the survey is primarily measuring one component; however, there also

appears to be at least two smaller components being tested by the survey instrument, as

components two and three have eigenvalues of 2.752 and 2.181 respectively. Together, these

first three components account for about 62% of the total test variance.
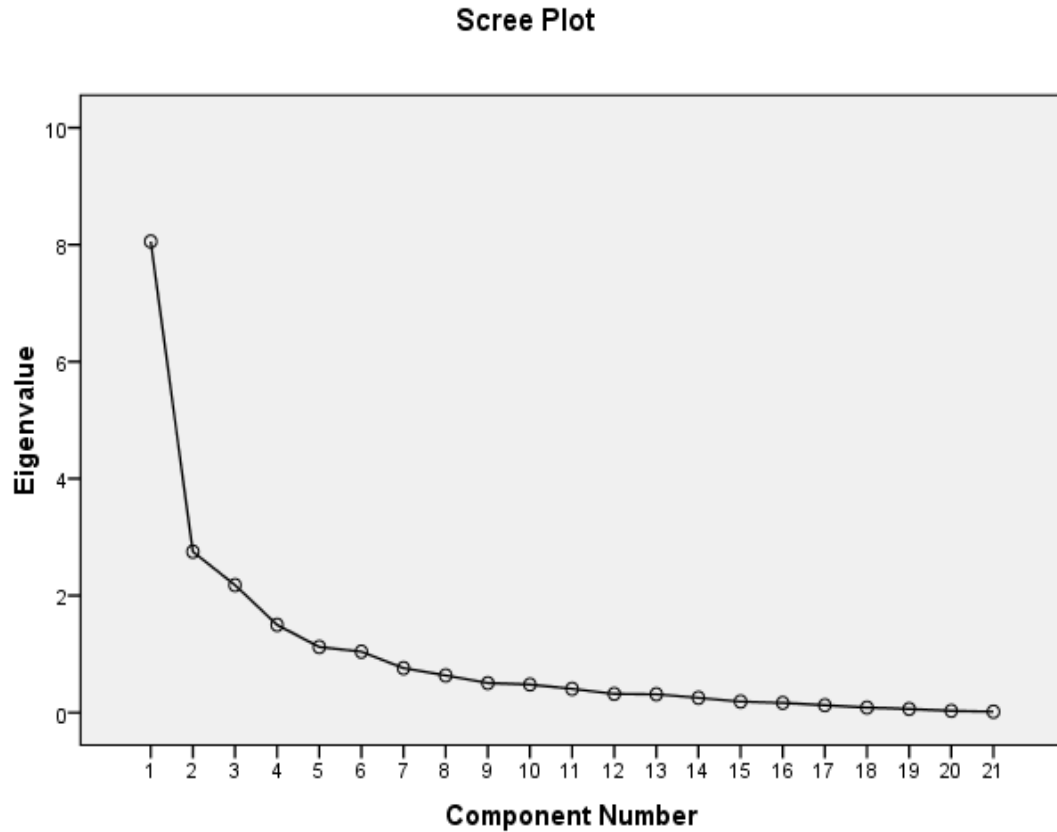
Scree Plot

Figure 1. *Scree plot of principal component analysis #1*

Table 5

Initial Eigenvalues for Principal Component Analysis #1

Component	Total	% of Variance
1	8.05	38.35
2	2.75	13.10
3	2.18	10.38
4	1.50	7.14
5	1.12	5.34
6	1.04	4.96

To understand which items make up each of the three components, the PCA solution's

oblique factor loading matrix (when only three factors were extracted) was examined. The

oblique factor loading matrix was used here, as opposed to the orthogonal, because the researcher hypothesized that answers from the three sections of the survey should be correlated. Table 6, the oblique component matrix from the first PCA, shows the loadings of each test item to each of the three components extracted. These loadings represent the correlation between each survey item and the three extracted components being tested.

As evidenced by Table 6 below, eleven out of twenty items loaded most highly on the first, most dominant component (i.e. items 2-6, 12, 14-17, and 20). Six items load most highly on the second component (i.e. items 8-11, 19, and 21). Last, four items load most highly on the third component (i.e. items 1, 7, 13, and 18). The items that load on each of these components are not related in terms of the content they were constructed to measure. For example, the eleven items that loaded most highly on the first component are representative of all three sections of the survey instrument. This is true of the items that loaded most highly on the second and third components. As each survey section was developed to measure perceptions on different subjects (i.e. evaluation language, evaluation activities, evaluation and instructional improvement), the items in each component are not related in any obvious way.   The researcher speculates that this is due in large part to the low survey response rate. The low rate indicates that the participant group, by definition, was not representative of the larger faculty within the school.

This first PCA failed to reveal any obvious, qualitatively interesting components. Nevertheless, to supplement the PCA and to help determine the reliability of the survey instrument, the researcher utilized coefficient alpha, also known as Cronbach's alpha. This method was chosen because it effectively estimates "the internal consistency of items … which have a wide range of scoring weights, such as those on some attitude inventories or attitude scales" (Crocker & Algina, 1986, p. 138). As illustrated in Table 7 Cronbach's alpha was

calculated first using all twenty-one standardized items from the original survey. As illustrated

later in this chapter, the reliability statistics were again formulated (for reasons discussed below)

after eliminating items seven and twenty-one from the calculation. While these reliability tests

revealed that the original scale was, in fact, highly reliable, removal of the two aforementioned

items resulted in an even higher reliability determination. From this, the researcher concluded

that responses to items seven and twenty-one did not correlate with the single component being

tested.

Table 6

Oblique Component Matrix from First PCA

	Component		
	1	2	3
Item 1	.450		-.639
Item 2	.657	.451	
Item 3	.816		-.305
Item 4	.783		
Item 5	.831		
Item 6	.856		-.391
Item 7			.726
Item 8	.551	.627	-.349
Item 9	.398	.833	
Item 10	.349	.843	
Item 11		.732	
Item 12	.598		-.582
Item 13	.494		-.684
Item 14	.575	.357	
Item 15	.829	.400	
Item 16	.712		
Item 17	.721		
Item 18	.331	.554	-.763
Item 19		.700	-.375
Item 20	.653	.383	
Item 21	-.312	.431	.393

Table 7

Reliability Statistics for Original Scale

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.902	.901	21

This conclusion was further supported by the item-total statistics represented in Table 8 below. Here, the corrected item total correlation for items seven and twenty-one was near zero. This indicates that these two items did not correlate with the corrected item total for the other scale items. Further, Table 8 reveals that the reliability statistic used, i.e. Cronbach's alpha, increases substantially for the overall scale when either item seven or twenty one was deleted.

Consequently, the researcher reexamined both scale items to determine qualitative reasons for this statistical outcome. Item number seven concerned the relationship between annual goal-setting and the teacher evaluation process. The item asked participants to indicate their level of agreement with a condition that did not exist: "In my school, an annual goal setting process for teachers should be linked to teacher evaluation." Annual goals for instructional improvement are teacher-generated and are one means of providing ownership for the evaluation process when linked to teacher evaluation (Glanz & Neville, 1997). The researcher postulated, then, that this item (with which a large majority of respondents showed agreement) was leading in that it essentially asked teachers whether or not they preferred to have more input and ownership regarding evaluation of their performance.

Table 8

Item-Total Statistics for Original Scale

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	50.00	102.138	.469	.899
2	50.40	95.007	.681	.893
3	50.00	99.310	.651	.894
4	50.53	103.637	.601	.896
5	50.37	100.861	.672	.894
6	50.30	101.734	.691	.894
7	49.40	110.938	.004	.913
8	49.57	98.116	.687	.893
9	49.03	101.895	.604	.896
10	49.00	102.345	.563	.897
11	49.63	102.585	.460	.899
12	50.13	101.361	.607	.895
13	50.27	102.892	.484	.898
14	49.77	102.116	.568	.896
15	49.70	98.493	.757	.891
16	49.80	102.579	.541	.897
17	49.47	99.499	.599	.895
18	49.03	101.895	.539	.897
19	49.07	105.857	.384	.901
20	49.90	100.369	.597	.895
21	48.63	113.344	-.133	.908

Similarly, item twenty-one concerns teacher ownership for improvement of their

instructional practice: "I rely upon reflection on my own teaching to improve student

performance." The researcher again postulated from the high levels of participant agreement

with this item that it too was a leading statement. As individuals generally want to see

themselves in the best light, the disproportionate levels of agreement here are understandable.

Upon reexamination, the researcher likens the item to the statement of a side-show fortune teller,

"You are a very independent person," to which a customer is likely to respond, "Why yes, I am!"

After all, regardless of one's independent nature, who would not agree that they were so? In more statistical terms, it is also the lack of variability in responses to items twenty one and seven that contributes to the lowering of the alpha.

In light of these qualitative concerns unearthed by the reliability analysis, the researcher decided to re-run all previous statistical operations deleting items seven and twenty-one from the scale. The differences in statistical results are readily observable in Figure 2 and Table 9 below. Figure 2 represents the scree plot generated from the second PCA.

Scree Plot

Figure 2. Scree plot of principal component analysis #2

Like that in Figure 1 above, the plot in Figure 2 shows that one primary component is being tested. A more detailed analysis of this exists in Table 9, where the initial eigenvalues for these five major components are represented. In contrast, six components were extracted from the original scale as represented in Table 5 above. As aforementioned and demonstrated here, once items seven and twenty-one were deleted from the scale, the first component extracted remains by far the dominant component being tested. Rather than accounting for 38 percent of the total variance as when these two items were included in the scale, this component now accounts for 42 percent of the variance while the next nearest component accounts for only 13 percent of the total test variance. After examination of Table 9, the researcher determined that only the first two components accounted for enough variance to warrant extracting and examining their corresponding factor loadings. Admittedly, a subjective decision was made here to cut the three smallest components out of this analysis. This decision was based in part because of the closeness of components 3-5 eigenvalue to 1, as well as the graphical examination of the leveling off of the scree plot after the second component.

Table 9

Initial Eigenvalues for Principal Component Analysis #2

Component	Total	% of Variance
1	8.022	42.219
2	2.500	13.160
3	1.584	8.339
4	1.369	7.206
5	1.011	5.320

Deletion of these items from the scale also resulted in augmented reliability statistics as demonstrated in Table 10.

Table 10

Reliability Statistics for Scale with Items 7 & 21 Deleted

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.920	.921	19

Table 11 below represents the oblique factor loadings generated from the second principal component analysis, i.e. the analysis conducted once items seven and twenty-one were deleted from the scale (and when only two factors were extracted). Here, a majority of items load moderately to strongly (at least at .3) on both components. Only items 11 and 19 do not load even moderately (at least at .3) on the first component. Only items 3-6 and 16-17 do not load even moderately on the second component. In a qualitative sense, items 3-6 are related as they seek participants' perceptions of the language of teacher performance standards. Likewise, items 16-17 are related to the role of supervision and evaluation relative to instructional improvement. Collectively, however, these items with low loadings are not related. The researcher also concluded that the two components do not form distinct qualitatively interesting factors.

Analysis of the attitude scale utilized in the present study revealed several key factors for which the researcher accounted when generating descriptive statistics regarding participant responses. First, a principal component analysis extracted three components being measured by the scale; the first component extracted accounted for 38 percent of the total scale variance. Though three components were extracted, the items comprising the components did not coincide with the researcher's original postulation that each of the first three sections of the scale would

be extracted as an individual component. Second, the original scale was found to be highly

reliable with an overall coefficient alpha of .901.

Two scale items, however, were deleted from the scale analyses when it was revealed that

they did not correlate with statistics for the other nineteen items. Consequently, the researcher

chose to regenerate statistics for both the component analysis and the reliability test.

Table 11

Oblique Component Matrix
from Second PCA

	Component	
	1	2
Item 1	.506	.427
Item 2	.655	.509
Item 3	.808	
Item 4	.763	
Item 5	.809	
Item 6	.863	
Item 8	.611	.654
Item 9	.376	.794
Item 10	.344	.781
Item 11		.777
Item 12	.668	.414
Item 13	.584	.336
Item 14	.598	.405
Item 15	.813	.415
Item 16	.705	
Item 17	.735	
Item 18	.436	.740
Item 19		.790
Item 20	.636	.330

Absent the aforementioned scale items, the reliability analysis indicated that the overall scale was highly reliable. However, the PCA results again failed to indicate that three factors (one corresponding to each section) comprised the full measure. Instead, it appears that one dominant factor (measured by all but two items) is primarily being measured by the instrument. While a small second factor is also measured by a majority of the items. The researcher was unable to assign distinct qualitative descriptions to either of the two components. The PCA results are not entirely unexpected, however, because the sample of subjects consisted of only 30 people. Obtaining stable factor analysis results for 19-21 items from a sample of 30 people is not likely, for as Crocker and Algina state, "A common rule of thumb for the minimum sample size in factor analysis is to use the larger of the following: 100 examinees or 10 times the number of variables" (1986, p. 296).

Due to the inconclusiveness of the PCA, but the high reliability of the overall scale, the researcher decided to simply treat the overall measure as one highly reliable scale, as opposed to three smaller scales as was originally hypothesized. Also, due to the tenable nature of the understanding of the overarching component(s) being measured by the overall instrument, the researcher decided to also examine individual item responses to attempt to understand the specific aspects of the teachers' attitudes.

Descriptive Statistics: Organization & Interpretation

In generating descriptive statistics, the researcher sought to perform statistical tests that would best answer the question for research: "What are teachers' perceptions of evaluation in an urban high school that uses a standards-based teacher evaluation system?" As aforementioned, the principal component analysis determined one dominant component being tested. This component measured the extent to which teachers believe that their evaluation process serves to

improve instruction. The attitude scale utilized consisted of four levels of agreement for each

item listed, ranging from "1", indicating strong disagreement with the item, to "4", indicating

strong agreement with the item. The mid-point for the scale, then, is 2.5. Consequently,

participants' mean scores that fell below this mid-point for a given item indicated general

disagreement with the item, while mean scores above 2.5 indicated general agreement with the

item. Quantitative findings are either supported or refuted by qualitative data offered in

participants' open responses to survey items 22 through 25. These qualitative data are grouped

into four categories, one for each of the four open response survey items. Survey item 22, for

example, asked participants to identify a major strength of their evaluation system. The

hypotheses are then restated along with a summary of the study's findings.

The small number of survey respondents significantly limited the use and efficacy of

statistical operations. While the researcher planned to derive data from chi-square analyses,

analyses of variance (ANOVA), *t-tests,* and cross-tabulations, the data generated from these

operations were neither statistically significant nor scientifically sound as they lacked statistical

power due to the small sample size. Therefore, descriptive data in the form of aggregate item

statistics tables and total mean score tables by demographic descriptors serve as the primary

sources of response data for the present study.

The first data examined were the aggregate summary item statistics as presented in Table

12 below. With a mean response score falling below the midpoint by only .088, data indicate

that in general, participants were neutral in their perception of the usefulness of teacher

evaluation processes as a means of improving instructional practice.

Table 12

Aggregate Summary Item Statistics

Mean	Minimum	Maximum	Range
2.412	1.667	3.200	1.533

The data presented in Table 12 are, by their nature, overarching and demand a closer look at individual item responses. For Section A of the survey, i.e. items one through six, these data are outlined in Table 13 which illustrates item statistics in the form of overall mean response scores to these items. Data in this table provide a more detailed view of participant responses from which several conclusions were drawn. In addition, these findings prompted further disaggregation of the data by demographic groups.

Survey Section A: The Seven Standards for Teacher Performance

Section A of the scale was designed to gauge: teachers' familiarity with the seven standards for teacher evaluation; teachers' understanding of the standards and their use as the language for effective instruction; and teachers' perceptions of the role of the seven standards in the evaluation process. As Table 13 illustrates, participants' mean response scores represent general disagreement with these items. Participants demonstrated more agreement with items one and three, items which asked participants their level of familiarity with the seven standards for evaluation and the role of these standards in the evaluation process. Participants showed very low levels of agreement with item two, which asked them if their evaluation process was based wholly on the seven standards for teacher performance. Items four, five, and six concerned the understanding and use of the standards as the language for teaching, learning, and instructional improvement within their school. Participants generally disagreed that regular discussion of

behavior that models the descriptors existed in the school. They also did not agree that the seven standards and their descriptors were used as the common language for instructional improvement within the school. This is especially true of item five which dealt with the presence of a common understanding of specific instructional practices used within the school.

Thus, in aggregate, teachers disagree slightly that they are familiar with the seven standards upon which their evaluation system is based. They indicate a similar level of disagreement regarding the role of these standards in the evaluation process. However, teachers disagreed more strongly when asked if the standards were used as the basis for teacher evaluation. Likewise, participants did not agree that the seven standards were used as the common language for instructional improvement within the school. The data in Table 13 also demonstrate the lack of a common understanding of instructional practices that model the seven standards among teachers and administrators.

Table 13

Survey Section A Aggregate Scale Item Statistics

Item	N	Scale 1= Strongly Disagree 4 = Strongly Agree	Mean	SD
1. I am very familiar with the seven standards for teacher performance as used in my school district.	30		2.20	.961
2. Our evaluation process is wholly based upon the seven standards for teacher performance.	30		1.80	1.186
3. I understand how the seven standards for teacher performance are used in our teacher evaluation process.	30		2.20	.925
4. In my school, there is regular discussion about what the seven standards for teacher performance look like in the classroom.	30		1.67	.661
5. In my school, there is a common understanding among teachers and administrators regarding specific behaviors that model the seven standards for teacher performance and their descriptors.	30		1.83	.791
6. The seven standards for teacher performance and their descriptors serve as the common language for effective teaching in my school.	30		1.90	.712

A closer examination of item response statistics by participants' years of service in the

district, however, yielded notable differences in responses between the two identified groups. As

illustrated in Table 14, participants with fewer than ten years of service in the district showed

slightly more agreement on all items in Section A than did their counterparts with ten or more

years of service.   This indicates that in general, teachers with ten or more years of service in the

district, i.e. more veteran teachers, were less likely to be familiar with the seven standards for

teacher performance. The data also indicate that these veteran teachers were less familiar with

the language of effective instruction and with the role of the seven standards relative to the

evaluation system.

Table 14

Survey Section A Item Statistics by Years of Service in District

	0-10 Years of Service in District			10 + years of Service in District		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Item 1	2.33	.966	21	1.89	.928	9
Item 2	1.81	1.123	21	1.78	1.394	9
Item 3	2.29	1.007	21	2.00	.707	9
Item 4	1.71	.717	21	1.56	.527	9
Item 5	1.90	.889	21	1.67	.500	9
Item 6	2.00	.775	21	1.67	.500	9

A qualitative reason for these differences in perception may stem from state and district

policies and their implementation. As stated in Chapter One, the Massachusetts Education

Reform Act of 1993 mandated that each school district within the state adopt a standards-based

teacher evaluation system by 1996. At the time of data collection, the standards within the

participating district had been in place for just over ten years. Thus, those participants with ten

or more years of service within the district had been evaluated under a non-standards-based

evaluation system from the beginning of their careers in the district until 1996. Participants with

fewer than ten years of service in the district would only have been evaluated under the

standards-based instrument.

Further, the district teacher evaluation policy and local collective bargaining agreement

provide mandates that teachers without professional status undergo at least two formal

summative evaluations per year for their first three years of service. In contrast, teachers with

professional status are evaluated every two years. It is postulated, then, this latter group's

stronger levels of disagreement with the items in survey Section A stem from less exposure to

the standards through the evaluation process.

Table 15 illustrates item statistics for survey Section A disaggregated by education level.

Statistics for participants having earned only undergraduate degrees and for those having earned

graduate degrees indicate general disagreement with items one through six. While all

participants are more likely to agree that they are familiar with the seven standards for teacher

performance, in aggregate, they are less likely to believe that the evaluation system is wholly

based upon these standards. Further, on each of the six items, participants with bachelor's

degrees showed slightly less agreement, on average, than did their colleagues who possessed

graduate degrees. Again, this indicates slightly less familiarity with the standards and how they

are used as the basis for evaluation and the common language for instructional improvement. In

particular, the undergraduate group showed a modicum of disagreement with survey items two

and five in particular. Hence, participants with bachelor's degrees are less likely to agree that the

evaluation system is based wholly on the seven standards. Further, they are less likely to agree

that a common understanding of instructional practices that models the standards exists within their school. Three out of the five respondents who held bachelor's degrees indicated that they had six or more years of service within the school and within the district; the other two members of this group indicated one to five years of service within the school district.

Table 15

Survey Section A Item Statistics by Education Level

	BA/BS			MA/MS/Med		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Item 1	2.20	.837	5	2.20	1.000	25
Item 2	1.20	1.304	5	1.92	1.152	25
Item 3	2.00	1.000	5	2.24	.926	25
Item 4	1.40	.548	5	1.72	.678	25
Item 5	1.40	.548	5	1.92	.812	25
Item 6	1.80	.837	5	1.92	.702	25

Table 16 outlines survey participants' written responses related to teacher familiarity with the role and language of instructional performance standards. Responses were sorted into one of four categories: major strength of the evaluation system; substantial change to the evaluation system; major drawback to the evaluation system; and other comments. In all, there were seven responses related to Survey Section A.

Five out of nine study participants with ten or more years of service in the district offered comments related to the language of the seven standards, of the evaluation process, and of effective instruction. These comments as outlined in Table 16, explain the aforementioned

difference in perception between this group and the group of teachers with fewer than ten years

of service.

Table 16

Written Responses Related to Survey Section A: The Seven Standards for Teacher Performance

Response Number	Education Level	Years in District	Comment	Response Category
1.	M	1-5	"I was handed a copy of the [standards] for evaluation that were used on me."	Major strength of evaluation system
2.	M	6-10	"First explain [the standards]. Review [them] with the faculty. Offer examples of best teaching practices so teachers can model these practices."	Substantial change needed in evaluation system
3.	M	10-20	"More time is needed…for formal and informal discussion so that communication is open and teachers have a very clear view of what [the] expectations are."	Substantial change needed in evaluation system
4.	M	20+	"Make people aware of [the standards]. I know nothing about it."	Substantial change needed in evaluation system
5.	M	20+	"I know nothing about [the standards]."	Major drawback of evaluation system
6.	M	10-20	"More information about what an administrator wants to see in the classroom."	Substantial change needed in evaluation system
7.	M	20 +	"School-wide discussions with faculty, administration, and support staff concerning standards [and] interpretation of same might increase teachers' comfort with the process."	Substantial change needed in evaluation system

Also indicated in this table, only 2 out of 21 participants with 1 - 10 years of service

provided comments related to the items in Section A of the survey. These comments are not

representative of an overall theme expressed by this demographic group. In fact, as detailed later

in this chapter, members of this demographic group provide qualitative responses that reflect

very different concerns regarding the evaluation process. However, the absence of such

comments from this demographic upholds quantitative findings that teachers with fewer years of

service indicate more familiarity with the seven standards and how they are used in the

evaluation system. This group of teachers is less likely, then, to perceive the language of

evaluation as a major drawback to the system or its implementation.

In stark contrast, four out of nine participants with ten or more years of service provide

written comments related to the seven standards and their use as the common language for the

overall evaluation process. All of these comments fell into the major drawback or substantial

change categories. The common theme among these comments was the need and desire for

presentation and clarification of the seven standards. This shows that respondents see the value

of instructional standards and their utilization as a common language for instructional

improvement. The comments also indicate, however, that participants in this demographic desire

clear definitions of the standards and examples of how they can be effectively implemented in

the classroom.   They believe that both teachers and administrators can reach consensus

regarding clear and common expectations for teacher performance. This finding coincides with

the quantitative data for Section A of the survey, which suggests that teachers with more than ten

years of service are less likely to agree that they have been provided with a clear explanation of

the standards used to evaluate their performance.

Section B: Evaluation Activities and Communication

Section B of the survey consists of items 8 through 14. These seven Likert items were

utilized to measure participants' perceptions of evaluation-related activities and communication

between teachers and their supervisors. In particular, these items address teachers' perceptions

Table 17

Survey Section B Aggregate Item Statistics

Item	N	Scale 1 = Strongly Disagree 4 = Strongly Agree	Mean	SD
8. Administrators in my school communicate a sophisticated understanding of effective teaching.	30		2.63	.964
9. Conferences between teachers and administrators are an important component of teacher evaluation.	30		3.17	.791
10. When done well, conferences between teachers and administrators as part of evaluation are helpful in improving teaching and learning.	30		3.20	.805
11. In my school, conferences between teachers and administrators as part of evaluation are done well.	30		2.57	.935
12. Administrators spend sufficient time in classrooms in order to carry out the district's expectations for evaluation.	30		2.07	.828
13. Administrators spend sufficient time in classrooms in order to carry out teachers' expectations for evaluation.	30		1.93	.868
14. Teachers and administrators in my school share a common understanding of specific criteria for performance ratings ("*meets/does not meet expectations*") in the evaluation process.	30		2.43	.817

of administrator knowledge and competence regarding teacher evaluation, the role of

conferencing in the evaluation process, and the criteria for teacher performance ratings. Table 17

shows the aggregate scale item statistics for each of these items.

Items 9, 10, and 11 in particular deal with evaluation based conversations. According to

the data above, participants were more likely to agree that conferences linked to evaluation are

an important part of the evaluation process. Further, aggregate mean responses illustrate that

participants believe that supervision/evaluation based conferences, when done well, are helpful

as a means of improving teaching and learning. Last, mean aggregate statistics showed

participants' agreement that conferences linked to evaluation were done well in their school.

This mean response score here was lower, however, than the above-mentioned items, indicating

participants' belief that conferences could be conducted in more effective ways within their

school.

Three survey items in Survey Section B measure teachers' perceptions of administrators'

roles in evaluation. According to the aggregate mean score for item 8, for example, participants

agreed that administrators in their building possess a sophisticated understanding of effective

instructional practices and communicate effectively this knowledge and understanding.

According to response data for items 12 and 13, administrators did not spend as much time in

teacher classrooms as expected by teachers and by the school district. Once again, this indicates

that teachers do believe in the power of effective conferencing as a means of improving both the

evaluation experience and ultimately as a means of improving instruction. Teachers are further

inclined to agree that teachers within their building possess a sophisticated knowledge of

instructional practices and are able to communicate this understanding effectively. Participants

also agreed that evaluation conferences in their school were done well. Teachers did not, in

aggregate, believe that administrators met their or the district's expectations for the amount of time they spend observing classroom instruction. According to the mean response score for item 14, teachers also did not agree that they and their supervisors shared a common understanding of specific criteria for performance ratings. These data are significant because they suggest a fundamental belief that relationships between administrators and teachers and conversations around teaching and learning are a key component in fostering a culture of instructional improvement.

Next, the mean item response data were disaggregated to show whether the aforementioned findings held true for demographic groups. Table 18 contains item statistics for Section B of the survey by participants' years of service within the district. As indicated above, these relate to evaluation activities and evaluation–related communication between and among teachers and administrators.

Table 18

Survey Section B Item Statistics by Years of Service in District

	0-10 Years of Service in District			10 + years of Service in District		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Item 8	2.71	.956	21	2.44	1.014	9
Item 9	3.24	.768	21	3.00	.866	9
Item 10	3.29	.784	21	3.00	.866	9
Item 11	2.52	.981	21	2.67	.866	9
Item 12	2.14	.793	21	1.89	.928	9
Item 13	2.00	.775	21	1.78	1.093	9
Item 14	2.43	.811	21	2.44	.882	9

Participants showed notable levels of agreement with items 8 through 11. According to the item statistics in Table 18, the two groups agreed that administrators in their building possessed a sophisticated understanding of teaching and learning. They also agreed that pre and post-evaluation conferences were helped to improve teaching and learning when done well; participants also shared the perspective that conferences within their building were done well, i.e., they served to improve teaching and learning. With the exception of item 11, mean response scores for these items were higher for participants with fewer than ten years of service than those of participants with more than ten years of service. While still scoring above the 2.5 midpoint score on item 11, the 0-10 year demographic shows less agreement than their more veteran counterparts with the statement that conferences between teachers and administrators within their school are done well.

As suggested by the data in Table 17 regarding items 12 and 13, participants believed that administrators did not meet their expectations for time spent observing classrooms. Data in Table 18 show a substantial disparity between those with more than ten and fewer than ten years of service within the district. The demographic with fewer years of service scored noticeably higher than their counterparts, indicating more agreement with these items. A qualitative explanation stems from the policies and contractual obligations of the participating district. Teachers new to this district must be formally evaluated two times per year for each of their first three years of service. Teachers are to be evaluated every two years thereafter. Given the overwhelming number of participants within the zero-to-ten year demographic, it is reasonable to suggest that administrators would spend more time in the classrooms of newer teachers. Even if this is the case, again, overall participant scores indicate that the time spent is insufficient. The mean response score for item 14 is nearly identical for both groups in this demographic, both

scoring slightly below the midpoint score of the scale. As there is no noticeable difference

between either of these groups and the aggregate response score for this item in Table 17, the

findings here reflect those following the aggregate item statistics.

Table 19 provides statistics for items in survey Section B as disaggregated by education

level. Overall, mean responses for participants in the BA/BS demographic were higher for each

item in this section with the exception of one item. Participants in this group, for example, agree

more strongly than their graduate-level counterparts that administrators in their building convey

a sophisticated understanding of teaching and learning.   This is also true concerning participant

Table 19

Survey Section B Item Statistics by Education Level

	BA/BS			MA/MS/Med		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Item 8	3.20	.837	5	2.52	.963	25
Item 9	3.40	.548	5	3.12	.833	25
Item 10	3.00	.707	5	3.24	.831	25
Item 11	3.00	.000	5	2.48	1.005	25
Item 12	2.40	.894	5	2.00	.816	25
Item 13	2.20	.837	5	1.88	.881	25
Item 14	2.20	.837	5	2.48	.823	25

beliefs about the role of effective conferencing in improving teaching and learning. Item ten

represents a departure from this trend as the BA/BS demographic did not believe as strongly as

their MA/MS colleagues that conferencing within their schools as part of the evaluation process

was done well. In addition, participants in the undergraduate group scored at higher levels on

items 12 and 13 than did the graduate demographic. The data here reflect, however, those

presented in Tables 17 and 18 which show that overall, participants do not agree that

administrators spend enough time observing classroom instruction.

The qualitative evidence in Table 20 below provides some context for these findings.

Again, these data were collected from the four open response items provided in survey Section

D.

Table 20

Written Responses Related to Survey Section B: Evaluation Activities and Communication

Response Number	Education Level	Years in District	Comment	Response Category
1.	MA/MS	10-20	"There is a conference between teacher and administrator before the evaluation. This will help the teacher prepare better for evaluation. There is a conference after the evaluation. Allows teacher and administrator to get into more details."	Major strength of evaluation system.
2.	MA/MS	1-5	"As far as I can tell, there is very little evaluation that is conducted. No administrator has been in to observe my classroom this year. I have only been evaluated once in 3 years."	Major drawback to the evaluation system.
3.	MA/MS	1-5	"I am considered a strong teacher, but because I am not evaluated, this reputation seems tenuous. In the absence of evaluation I do not feel like I am receiving adequate support or feedback."	Substantial change needed in evaluation system.
4.	BA/BS	1-5	"Observing one class may not give an adequate representation of teachers' habits, abilities, strengths/weaknesses."	Major drawback to evaluation system.
5.	BA/BS	6-10	"[I have been evaluated] only once in 6 years. This is my review year."	Additional feedback regarding evaluation experiences.

Table 20

Written Responses Related to Survey Section B: Evaluation Activities and Communication

Response Number	Education Level	Years in District	Comment	Response Category
6.	BA/BS	10-20	"Administrators are required to make several visits to the classroom. I would suggest that they don't just pop in for 5 minutes but they come and take part as often as possible."	Substantial change needed in evaluation system.
7.	MA/MS	1-5	"Feedback from the headmaster on teaching style with a good sit down conference [is needed]."	Major drawback to evaluation system.
8.	MA/MS	6-10	"[Evaluation] is seen as a punishment and as a fake show because it is done so infrequently."	Major drawback to evaluation system.
9.	MA/MS	1-5	"I would like everything to be more consistent. I haven't been observed all year. Last year I was observed once a month."	Substantial change needed in evaluation system.
10.	MA/MS	1-5	"[A major strength] does not exist. I have taught in my school for three years and have not been observed once by an administrator."	Major strength of evaluation system.
11.	MA/MS	1-5	"More time to pre/post conference."	Substantial change needed in evaluation system.
12.	MA/MS	1-5	"My administrators have been very open to helping improve my teaching and are very knowledgeable as to how to do so."	Additional feedback regarding evaluation experiences.
13.	MA/MS	1-5	"More time with administrators before and after formal evaluation."	Substantial change needed in evaluation system.
14.	MA/MS	1-5	"I have not been evaluated as of yet."	Additional feedback regarding evaluation experiences.
15.	MA/MS	6-10	"Conference with me [about] student exemplars so administrator understands my field."	Substantial change needed in evaluation system.

Table 20

Written Responses Related to Survey Section B: Evaluation Activities and Communication

Response Number	Education Level	Years in District	Comment	Response Category
16.	MA/MS	6-10	"More administrators involved in evaluation. I have not been observed by an administrator in <u>4</u> years!"	Substantial change needed in evaluation system.
17.	MA/MS	1-5	"If not done well, with an effective and communicative administrator, the teacher is set up to fail."	Major drawback to evaluation system
18.	MA/MS	1-5	"Not enough communication 1:1 with supervisors unless there is a problem."	Major drawback to evaluation system
19.	MA/MS	20+	"Many teachers feel uncomfortable with what they perceive as unprepared, incompetent people conducting evaluations."	Major drawback to evaluation system
20.	MA/MS	1-5	"Observe more often, have specific goals in mind for each evaluation, discuss post-evaluation."	Substantial change needed in evaluation system.

The topic of evaluation activities and communication was a very important one for study participants as evidenced by the number of written responses. The majority of these represented major changes needed within the evaluation system or major drawbacks to the system. The most frequently recurring theme aligns with quantitative data regarding survey items 12 and 13.   As illustrated in Tables 17-19, participants believed that administrators did not spend enough time observing classroom instruction to meet their and their district's expectations. Fourteen of the twenty written responses in Table 20 speak directly to this factor, the common thread being that respondents wish for consistent administrator presence in their classrooms during instructional time. This sentiment crosses the demographics of years of service and education level. In addition, response 12 in Table 20 directly supports both the aggregated and disaggregated mean

response statistics for survey item 8. These data revealed that on average, survey participants agreed that their evaluators possess and communicate a sophisticated understanding of effective instructional practices.

Six written responses also speak to participant perceptions of conferencing as linked to evaluation. Within these responses, teachers indicate the importance of conferencing with administrators in a regular basis, both before and after classroom observations. These responses support quantitative data for survey items 9, 10, and 11 across demographics. These data showed that teachers perceived that conferences as a part of evaluation, when done well, are effective in improving instruction. While quantitative data indicate that participants are neutral, at best, regarding survey item 14, there were no written responses that directly related to the concept of criteria for performance ratings.

Thus, several findings flow from the data gathered through survey Section B. Qualitative and quantitative data for this section show that teachers value the efficacy of effective conferencing related to teacher evaluation. In addition, teachers, on average, believed that their administrators possessed and communicated a sophisticated understanding of effective instructional practices. Time emerged as an important strong theme within this section, as teachers believed that frequent visits by administrators for observing and/or engaging in instruction within the classroom was essential to the improvement of teaching and learning. However, teachers also indicated that administrators did not spend sufficient time in classrooms to meet teachers' or the district's expectations for observations. On average, teachers disagreed slightly that a common understanding of specific criteria for performance ratings existed within their school.

Survey Section C: Evaluation and Instructional Improvement

       Survey Section C consisted of six items. These items measured the attitudes of participants concerning the effectiveness of their teacher evaluation system as a means of improving classroom instruction. More explicitly, the items ask teachers to discern the role of teacher evaluation, peer observation, and administrator feedback in improving the quality of teaching and ultimately student performance. In addition, teachers are asked to ascertain the extent to which their own evaluation is linked to the school's professional development initiatives and whether or not multiple sources of student performance data are part of the evaluation process. Last, participants shared their perceptions regarding the need for ongoing training and professional development of administrators in teacher evaluation techniques as a means of aiding teachers in improving their practice.

       Table 21 shows the aggregate item statistics for Survey Section C. Several important findings are revealed within these data. First, teachers overall mean score for item 15 is the scale midpoint, or median score of 2.5. This indicates overall participant indecision or neutrality regarding the effectiveness of teacher evaluation in increasing student achievement. Further, participants showed slight disagreement, on average, with survey item 16 thus indicating that they do not believe teacher evaluation to be linked with the professional development program within their school. The mean response score for item 17 shows that participants agree, though not strongly, that the feedback that they get from evaluation improves their instruction. Likewise, mean response data for item 18 indicate that feedback teachers receive from peer observations helps improve the quality of their teaching. According to response data for item 19, participants solidly affirm the belief that administrators need ongoing professional development around teacher evaluation, if evaluation is to be used as a tool to improve teacher practice. Last,

Table 21

Survey Section C Aggregate Item Statistics

Item	N	Scale 1 = Strongly Disagree 4 = Strongly Agree	Mean	SD
15. The process of teacher evaluation leads to increased student achievement in my school.	30		2.50	.861
16. Teacher evaluation is closely linked to my school's professional development initiatives.	30		2.40	.814
17. The feedback I get from evaluation helps me to improve the quality of my teaching.	30		2.73	.980
18. Feedback from peer observation helps me to improve the quality of my teaching.	30		3.17	.874
19. Administrators need ongoing professional development around teacher evaluation for evaluation to improve my practice.	30		3.13	.730
20. Multiple sources of student performance data are used as part of my evaluation.	30		2.30	.915

mean response data for item 20 show that teachers do not believe that multiple sources of information are used as part of the teacher evaluation system.

Though there are some variations among demographic groups for individual items, in general, these findings hold true for data disaggregated by years of service within the participating district and by education level. Table 22 outlines item statistics for survey Section C as disaggregated by years of service within the district. Accordingly, data disaggregated by participant education level are presented in Table 23. According to data in Table 22, there are several notable differences between the two groups for items 15, 16, 18, 19, and 20.

Table 22

Survey Section C Item Statistics by Years of Service in District

	0-10 Years of Service in District			10 + years of Service in District		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Item 15	2.57	.811	21	2.33	1.000	9
Item 16	2.52	.750	21	2.11	.928	9
Item 17	2.76	1.044	21	2.67	.866	9
Item 18	3.29	.784	21	2.89	1.054	9
Item 19	3.24	.700	21	2.89	.782	9
Item 20	2.56	.981	21	2.19	.726	9

Statistics here reveal that those with ten or more years of service on average do not agree that the process of teacher evaluation leads to increased student achievement. Their less-veteran counterparts show slight agreement with this item. Members of this latter group also believe, on average, that teacher evaluation is closely linked to their school's professional development plan, while those in the ten-plus years demographic were much less likely to agree with this assertion.

Both groups represented in Table 22 agree that feedback from evaluation and from peer observation helps to improve their instructional practice and that administrators need ongoing instructional improvement regarding teacher evaluation. Respondents with ten or fewer years of service, however, showed markedly higher levels of agreement with these beliefs than did those with ten or more years of service. This pattern is disrupted, however, when the latter group, on average, shows slight agreement with Item 20, that administrators use multiple sources of student performance data as part of their evaluation. Participants with ten or more years of service indicate stronger disagreement with this statement. Explanations for these responses will be presented in light of participants' written responses presented later in this section.

Table 23

Survey Section C Item Statistics by Education Level

	BA/BS			MA/MS/MEd		
	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Item 15	2.40	.548	5	2.52	.918	25
Item 16	2.40	.548	5	2.40	.866	25
Item 17	3.20	.447	5	2.64	1.036	25
Item 18	3.80	.447	5	3.04	.889	25
Item 19	3.20	.447	5	3.12	.781	25
Item 20	1.60	1.140	5	2.44	.821	25

As presented in Table 23, data disaggregated by participants' education levels reflect some similar response patterns represented by aggregate item statistics and those disaggregated by years of service. For example, those participants with undergraduate degrees on average showed less agreement with Item 15, that teacher evaluation leads to improved student

performance. Given that four of the five participants with undergraduate degrees have five or fewer years of service this statistic represents a departure from the data presented in Table 22. A qualitative cause for this could include the overall *pro forma* and inconsistent implementation of the evaluation process within the participants' school. Also according to Table 23, both teachers with and without graduate degrees showed less agreement with Item 16, that their evaluation is closely linked to their school's professional development plans. This, again, is indicative of the underutilization of the district's standards for teacher evaluation as a common language and basis for instructional practice and improvement. Likewise, one can speculate that the school's professional development plan is not based upon a common language. This would explain response data for this item. Participants with undergraduate degrees show clear agreement with Items 17 and 18, that feedback from administrator and from peer observation aids them in improving the quality of their teaching. They also are more likely to agree that administrators need ongoing professional development around supervision and evaluation if the process is to improve instruction. Item statistics for participants with graduate degrees also indicate agreement with these items; however on average the level of agreement is lower than that of participants with undergraduate degrees. Participants with graduate degrees did show stronger agreement with the statement that multiple sources of student performance data were used as part of their evaluation. Those with undergraduate degrees showed marked disagreement with this statement.

The findings here coincide with item statistics disaggregated by years of service. Statistics for participants with ten or fewer years of service mirror those of participants with undergraduate degrees and responses for those with ten or more years of service reflect the

Table 24

Written Responses Related to Survey Section C: Evaluation & Instructional Improvement

Response Number	Education Level	Years in District	Comment	Response Category
1.	BA/BS	1-5	"I think it would be very helpful to get feedback on my teaching but I have never had any formal feedback from administrators."	Additional feedback regarding evaluation experiences.
2.	BA/BS	1-5	"[The] evaluation process has an informal tone at my school. The informality alleviates tension and promotes best practices. Peer observation is extremely productive. The evaluation process is not used enough. Bad teaching continues because of a lack of incentive and a lack of prescription."	Major strength of evaluation system.
3.	BA/BS	1-5	"Current evaluation system is fine, but administrators need to evaluate routinely those who are obviously under performing."	Major drawback to evaluation system.
4.	BA/BS	1-5	"In my personal experience, the evaluation process has been fair and productive because it was used as a tool to better my teaching not as a tool to threaten my job."	Additional feedback regarding evaluation experiences.
5.	BA/BS	10-20	"I believe feedback from administrators is crucial in the evaluation process. It should be timely and informative.	Major strength to evaluation system
6.	BA/BS	10-20	The administrators are required to make several visits to the classroom. I would suggest that they don't just pop in for 5 minutes but they come and take part in the class."	Major drawback to evaluation system.
7.	BA/BS	10-20	"Regular evaluations would assist with teacher performance, including accountability.	Additional feedback regarding evaluation experiences.

Table 24

Written Responses Related to Survey Section C: Evaluation & Instructional Improvement

Response Number	Education Level	Years in District	Comment	Response Category
8.	MA/MS	1-5	"I have only been evaluated once in three years."	Additional feedback regarding evaluation experiences.
9.	MA/MS	1-5	"[Evaluation] is not used to improve teaching and learning in my school. This is a major drawback."	Major drawback to evaluation system.
10.	MA/MS	10-20	"Some administrators are not suitable to evaluate a teacher."	Additional feedback regarding evaluation experiences.
11.	MA/MS	20+	"Many people feel uncomfortable with what they perceive as unprepared, incompetent people conducting evaluations."	Major drawback to evaluation system.
12.	MA/MS	20+	"[Evaluation] should be a part of professional development."	Major drawback to evaluation system.
13.	MA/MS	20+	"More frequent short/brief visits to my classroom versus one longer, prepared lesson plan."	Substantial change you would make to your evaluation system.
14.	MA/MS	1-5	"Administrators focus on student success and whether or not teachers are facilitating this well."	Major strength of evaluation system.
15.	MA/MS	6-10	"More administrators involved in evaluation. I haven't been observed in 4 years!"	Major drawback to evaluation process.
16.	MA/MS	6-10	"[I] feel the more formal evaluations do not occur [more] frequently."	Additional feedback regarding evaluation experiences.

responses of participants who have earned graduate degrees. These comparisons stand to reason

as four of five respondents with undergraduate degrees have five or fewer years of service within

the district. The researcher speculates, however, that these findings may be confounded as a

result of the low survey response rate. In other words, as the participating population is not

sufficient to represent the faculty at large and the findings reflect this underrepresentation.

Table 24 offers participant comments related to survey Section C. There are several

consistent themes that have emerged from these responses. The first is related to the amount of

time spent on evaluation and the amount of time spent by administrators in the classroom.

Respondents who commented on this topic shared a common belief that administrators did not

spend enough time in the classrooms. This sentiment took several forms, including a lack of

multiple administrator visits, both formal and informal, to classrooms; the infrequency or lack of

formal evaluation of some teachers over multiple years; the desire by teachers for more

administrator engagement with students and actual classroom instruction during classroom visits.

Participants also intimated preferences regarding the type of observation to take place, some

preferring more frequent, informal visits and others preferring longer, more formal observations.

In contrast to this last finding, several respondents comment on the productiveness of the

evaluation system due precisely to the informal tone set by administrators. Teachers credit the

non-threatening manner in which administrators implement the evaluation process as a major

supporting factor in the use of teacher evaluation as a means of improving teaching and learning.

In addition, participants questioned the competency of the administrators conducting teacher

evaluations as well as the lack of evaluations of teachers perceived by their colleagues to be

underperforming. These responses crossed demographic lines and readily support the

quantitative data for Section C as presented in Tables 21 through 23 above.

Restated Hypotheses

As outlined earlier in this work, the primary research question for consideration in this

dissertation is: What are teachers' perceptions of evaluation in an urban high school that uses a

standards-based evaluation system? The researcher sought to answer this question in terms of

teacher familiarity with the stated standards and their indicators and the extent to which the

standards are utilized as the common language for evaluation activities and instructional

improvement within a school. There were three resultant statistical hypotheses that the

researcher tested in this study. Each hypothesis is provided below and restated in light of the

scale and data analysis conducted.

The first of these (H_1: $\mu_1 > \mu_2$) was: Teachers who are familiar with the language and

associated activities of a standards-based evaluation system perceive significantly higher levels

of instructional improvement from the process than those who are not familiar with the language

and function. The corresponding null hypotheses (H_0: $\mu_1 = \mu_2$) then, was stated as: There is no

significant difference in the perceived instructional improvement of teachers who are familiar

with the language and associated activities of a standards-based evaluation system and those who

are not familiar with the language and function of a standards-based evaluation system. Within

these hypotheses, μ_1 represents the degree of participant familiarity with the language and

associated activities of a standards-based teacher evaluation. Likewise, μ_2 represents levels of

instructional improvement. This hypothesis was dependent upon the researchers belief that more

than one component was being tested by the survey instrument. In this case, the researcher

postulated that survey Section A would test as a single component. As demonstrated earlier in

this chapter using scale analysis data, however, the component analysis extracted only one

dominant component from this survey. Given this, it is reasonable to postulate that the

alternative hypothesis (H_1: $\mu_1 > \mu_2$) is not tenable because the survey section in question does not

lend itself to direct examination. Therefore, the researcher cannot discern whether teachers who

are familiar with the language and associated activities of a standards-based evaluation system

perceive different levels of instructional improvement from the evaluation process than those who are not familiar with the language and function.

The second statistical hypothesis (H_1: $\mu_1 > \mu_2$) was: Teachers with ten or fewer years of service in the district perceive that a standards-based evaluation system has more of an impact on instructional improvement than do veteran teachers. The consequent null hypothesis (H_0: $\mu_1 = \mu_2$) was: There is no significant difference in perceptions of the impact of a standards-based evaluation system on instructional improvement between teachers with ten or fewer years of service and teachers with more than ten years of service. Here again, μ_1 represents the perception of teachers regarding the degree of participant familiarity with the language and associated activities of a standards-based teacher evaluation. Likewise, μ_2 represents levels of instructional improvement. Based upon the item statistics presented above in which participants with ten or fewer years of service in the district outscored the ten-plus year demographic on 18 out of 20 scale items, the researcher finds the null hypothesis to be untenable and the alternate hypothesis (H_1: $\mu_1 > \mu_2$) to be tenable. This is further supported by the group statistics presented in Table 25 below which indicate that those participants with fewer years of service in the district have a higher total mean score than their counterparts with ten or more years of service. This confirms that in contrast with their veteran counterparts, these respondents perceive that a standards-based evaluation system has more of an impact on instructional improvement

Table 25

Group Statistics for Years of Service in District

	Years in District	N	Mean
Total Score	0-10 years	21	46.9524
	10+ years	9	43.2222

The third and final hypothesis (H_1: $\mu_1 > \mu_2$) tested within this study was: Teachers without graduate degrees agree more strongly than their colleagues that hold graduate degrees that a standards-based evaluation system has an impact on instructional practice. Accordingly, the null hypothesis (H_0: $\mu_1 = \mu_2$) was stated: There is no significant difference in perceptions of the impact of a standards-based evaluation system on instructional improvement between teachers without graduate degrees and teachers with graduate and post-graduate degrees. Within these hypotheses, μ_1 represents the perceived impact of a standards-based evaluation system upon instructional practice. Therefore, μ_2 represents the levels of instructional improvement. Item statistics data interpreted above do support this alternate hypothesis.

While participants with undergraduate degrees outscored their colleagues with graduate degrees on only 8 out of 20 items, the data in Table 26 below indicate that the former group held a slightly higher total mean scale score. These findings prove inconclusive as that those participants holding undergraduate degrees only show slightly more agreement that a standards-

based evaluation system has an impact on instructional improvement than do study participants with graduate and post-graduate degrees.

Table 26

Group Statistics for Education Level

	Education Level	N	Mean
Total Score	BA/BS	5	46.0000
	MA/MS	25	45.8000

Summary of Findings

There are several limitations inherent within the present study which will be outlined in the chapter that follows. Analysis of the quantitative and qualitative data as presented above has yielded several findings regarding the attitudes of urban secondary teachers toward standards-based teacher evaluation. The researcher has identified ten major findings that are presented according to the three sections of the survey instrument used in the present study.

Within Survey Section A: The Seven Standards for Teacher Performance, the researcher found that participants, in aggregate, were more likely to **DISAGREE** that:

- participants are familiar with the Seven Standards for Teacher Performance;
- participants' evaluations are wholly based upon the Seven Standards for Teacher Performance;
- within their school there are common instructional practices that reflect the Seven Standards for Teacher Performance.

Aggregate data analysis from Survey Section B: Evaluation Activities and Communication showed that participants were more likely to **AGREE** that:

- conferences between teachers and administrators are an important component of teacher evaluation;

- when done well, these conferences are an effective means of improving teaching and learning;

- conferences conducted as part of teacher supervision and evaluation at the participating school are done well;

- administrators in the participating school communicate a sophisticated understanding of effective instruction.

Findings from this section also indicated that participants were more likely to **DISAGREE** that:

- administrators spend enough time in classrooms to meet teacher and district expectations;

- teachers and administrators in the participating school share a common understanding of specific criteria for performance ratings ("*meets/does not meet expectations*") in the evaluation process.

Data from Survey Section C: Evaluation and Instructional Improvement revealed that participants were more likely to **AGREE** that:

- the feedback they receive from supervision and evaluation helps to improve their practice;

- feedback from peer observation helps teachers to improve their instructional practice;

- administrators need ongoing professional development around teacher evaluation for evaluation to improve my practice.

Data from this section also showed that teachers were more likely to **<u>DISAGREE</u>** with the statements that:

- evaluation in the participating school is linked to the school's professional development plan;

- multiple sources of student performance data are used as part of teacher evaluation.

In addition, response data proved inconclusive regarding participants' perspectives on whether teacher supervision and evaluation is an effective means of improving student performance.

The aforementioned statements constitute the major findings of the present study. In the next chapter, the significance of the study is examined. This examination includes a summary of the study findings and the situation of these findings within the body of research as delineated and examined in Chapter Two. The researcher then discusses the implications of these findings for educational policy and practice. The limitations of the study design and findings are then discussed.

Chapter Five: Summary, Discussion, & Implications

Introduction

Chapter Five is organized in several main sections. First, the researcher presents a concise review of the study findings. This is followed by a discussion of these findings. Here, the researcher frames the study findings in terms of the earlier studies and the theoretical literature as presented in Chapter Two of this study. Next, the study limitations are outlined, including an examination of the limitations of survey research in general, as well as those limitations inherent in this particular study. An examination of study implications for educational policy and practice as well as those for further research follows. Last, the researcher outlines conclusions regarding the study. Within this section, the researcher guides the reader through the personal, scholarly, and professional journey which he had undertaken. This includes a bevy of learning opportunities afforded to the researcher while conducting the present study, as well as the myriad opportunities the research provided for reflection upon and refinement of the researcher's own base of knowledge and practice within the realm of standards-based teacher supervision and evaluation.

Summary of Findings

The present study revealed a series of findings related to each of the three sections of the survey instrument. As examined in Chapter Four, these concerned: participants' familiarity with the standards by which they were evaluated; the use of these standards as the common language for instructional practice and improvement within study participants' school and district; the use and efficacy of evaluation related activities (i.e. classroom observations and teacher-supervisor evaluation conferences); and the use of a standards-based system of teacher supervision and

evaluation not only as a means of quality control and accountability, but also as a means of improving instructional practice and student achievement. The major findings of the present study include:

- participants showed lower levels of familiarity with the Seven Standards for Teacher Performance;

- participants are less likely to agree that their evaluations are wholly based upon the Seven Standards for Teacher Performance;

- participants show less agreement that within their school there are not common instructional practices that reflect the Seven Standards for Teacher Performance;

- conferences between teachers and administrators are an important component of teacher evaluation;

- when done well, these conferences are an effective means of improving teaching and learning;

- teacher perceptions that conferences conducted as part of teacher supervision and evaluation at the participating school are done well;

- teacher perceptions that administrators in the participating school communicate a sophisticated understanding of effective instruction;

- teacher perceptions that administrators do not spend enough time in classrooms for supervisory and evaluative purposes to meet teacher and district expectations;

- teachers and administrators in the participating school do not share a common understanding of specific criteria for performance ratings ("*meets/does not meet expectations*") in the evaluation process;

- the feedback participants receive from supervision and evaluation helps to improve their instructional practice;

- feedback from peer observation helps teachers to improve their instructional practice;

- administrators need ongoing professional development around teacher evaluation for evaluation to serve as a tool for improving instructional practice;

- evaluation in the participating school is not linked to the school's professional development plan;

- multiple sources of student performance data are not used as part of teacher evaluation.

- participant responses were inconclusive regarding their perspectives on whether teacher supervision and evaluation is an effective means of improving student performance.

These findings were drawn from the descriptive, quantitative and the qualitative data examined in Chapter Four and represent, in aggregate, the attitudes of the thirty study participants. The quantitative data in particular were used to support the researcher's three hypotheses which, in turn, set a sound foundation for a discussion of the study findings.

Discussion of the Findings

In the second chapter of the present study, the researcher reviewed the body of literature related to standards-based evaluation systems and practices. This review was centered upon the primary research question for consideration, i.e. what are teachers' perceptions of evaluation in an urban high school that uses a standards-based teacher evaluation system? The researcher sought to answer this question in terms of teacher familiarity with the stated standards and their indicators and the extent to which the standards are utilized as the common language for evaluation activities and instructional improvement within a school.

The literature review was organized into two main parts, including the history and development of teacher evaluation standards and teacher ownership for supervision and evaluation processes. Included in the latter section were teacher and administrator perceptions of standards-based teacher supervision and evaluation. The discussion that follows situates key findings from the present study with salient points and themes found in the literature reviewed in both of these sections, with an emphasis on the latter section.

In accordance with the theoretical and research literature reviewed, the researcher asserts that the mere presence of a standards-based evaluation system does not guarantee that the system's use will affect teacher practice and student learning. Study findings, when compared with the literature reviewed, indicate that many of the deficiencies inherent in historically dysfunctional teacher evaluation practices and systems were present in that of the participating

school. The study of effective practices in a standards-based teacher evaluation system also led to the discovery of discrepancies within the present study regarding the improvement of instructional practice versus the improvement of student learning.

Participant perspectives of standards-based supervision and evaluation were in some ways aligned with descriptions of supervision of the late nineteenth and early twentieth centuries. A major aspect of these historical models includes the subjectivity of standards for performance (Blumberg, 1985). More precisely, the research indicates that from the mid-nineteenth through the late twentieth centuries, administrators assigned to oversee teachers served as the source - often the sole source - of professional authority and knowledge for teachers (Abbot, 1969; Beck, 1993; Beck and Murphy, 1993; Blumberg, 1985; Postel, 1955; Schott, 1956; Travers, 1981). Consequently, benchmarks for teacher performance were determined by each administrator, rendering evaluation criteria arbitrary. The development of standards-based teacher evaluation was meant to mitigate this subjectivity in favor of well-defined standards for and indicators of quality instruction. This work began as early as the 1970s with the Joint Committee on Standards for Educational Evaluation (Stufflebeam & Saunders, 1990). The intent of these standards was, as the term suggests, to eliminate the subjectivity and arbitrary nature of teacher evaluation so as to create a model that supports teaching, learning, and the professional growth of teachers. Efficacy of the standards-based system presupposes that teachers and administrators have an operational (if not expert) knowledge and understanding of the standards and instructional practices that model these standards. In fact, in keeping with Bloom's Taxonomy, Danielson (2006) believes that mastery of the content of teacher performance standards should be the first level of development in any differentiated teacher evaluation system.

However, even with the advent and promulgation of standards for teacher performance at the state and local levels in the 1980s and 1990s, the findings of the present study reveal that in aggregate, teacher participants were not familiar with teacher evaluation standards used in their district. In fact, many participants believed that their evaluations were not wholly based upon criteria related to the standards. While participants did not offer many examples of subjective criteria, written comments suggested that length of service in the building played a decided role in the evaluation process. For example, several participants with ten or more years of experience cited that they were unfamiliar with the standards for their performance. These participants indicated either that they had received satisfactory evaluations or had not been formally evaluated in years. Hence, as aligned with the aforementioned researchers, Rooney (1993) indicates that the mere existence of teacher performance standards does not serve to improve instructional practice. It is the common understanding of the standards among teachers and administrators and ownership for the associated supervision and evaluation process that drives actual instructional improvement.

Findings from the present study also align with some contentions of Black (1993), Danielson and McGreal (2000), and Rettig (1999). These all report that teacher evaluation systems, in practice, are punitive, arbitrary, and intimidating and work against the promotion of open and honest communication and collegiality; consequently, they do not support the ongoing improvement of teaching and learning. It is not surprising then, that participants in the present study were undecided regarding the link between their evaluations and the evaluation processes; furthermore, they underscored the arbitrary nature of their evaluation system even in the presence of standards for teacher practice.

If, as findings from the present study suggest, evaluation standards are not utilized as the common foundation for teacher evaluation, it is curious that study participants found that conferencing with their supervisors helped them to improve their practice. Absent a common language, and a common understanding of the standards for evaluation, it is difficult to ascertain what constitutes an effective teacher-supervisor conference. What further complicates the discussion is the perception among present study participants that evaluation conferences within their school are done well. Here, the researcher must indulge in speculation. For example, one could postulate that evaluation processes in the participating school revolve around superficial classroom observations concerning the implementation of administrative edicts, the compliance with previously agreed-upon and readily observable instructional tools (e.g. "word walls", technology, print-rich classroom environment), and the behavior of the teacher rather than the behavior and learning of the students. One could also speculate that teachers in this study find the conferences useful because they have respectful and trusting relationships with their supervisors. This aligns, in part, with the work of Glanz, Shulman, and Sullivan (2007) in which the researchers found that more organic evaluation processes led to open, honest, trusting, and collegial relationships among teachers and supervisors. The alignment between this research and the findings of the present study stops with this aspect, however. The Glanz study did not indicate that these relationships were developed in the absence of standards; on the contrary, the researchers found that the presence, understanding, and implementation of these standards led to positive and productive professional relationships in the school district studied. Unlike the present study, the evaluation processes in the Glanz study were flexible in nature, they were based upon common performance standards and a common understanding of the standards. Glanz, et. al. would be hesitant to judge how participants in the present study found their

evaluation conferences to be useful without being based upon performance standards and a common understanding of them.

Likewise, Mertler (1997), Rettig (1999), Shen (1998), and Weiss and Weiss (1998) all stress the importance of conferencing within the supervision and evaluation process. In particular, they stress that effective systems of supervision and evaluation are built upon respectful and trusting relationships among teachers and supervisors. As found in the Glanz, et. al. (2007) study, these relationships are a key to promoting a culture of collective responsibility for the achievement of all students within a school. Effective evaluation systems, according to the aforementioned researchers, are characterized by the presence of collegial, not hierarchical, relationships between teachers and supervisors and consist mostly of informal dialogue between and among staff members. These characteristics lie in direct contrast to those gleaned from the data of the present study. The assertions of Sparks (2004) and Glickman (2006) coincide with this, especially in terms of differentiated supervision and evaluation. Like Glanz, et. al. (2007), Sparks and Glickman advocate for a more organic supervision and evaluation process rather than one which is prescribed. They see hierarchical teacher supervision systems as impeding ongoing instructional improvement for they reinforce adherence to instructional and curricular mandates and do little to foster creativity within the classroom. This latter model resembles that perceived by participants in the present study.

Judging from present study findings, then, the researcher postulates that participants view their supervisors as the main - if not sole - repository of knowledge for effective instructional practices. This can be evidenced by their belief that administrators possess a sophisticated understanding of teaching and learning, that multiple sources of data are not used within the evaluation process, and that feedback from administrators helps them to improve their practice.

These conclusions, coupled with the data that indicate that teachers are not familiar with teaching standards and do not see standards for teacher practice as the common language for instructional improvement, reinforce the hierarchical notion that supervisors themselves do not employ the district performance standards as a foundation for teacher evaluation. Thus the elements of subjectivity, compliance, and superficial observation and discussion can continue to plague teacher evaluation practices even within a standards-based environment.

In yet another instance, findings from the present study illustrate that teachers perceive their evaluation system to be lacking many of the effective practices espoused by current research. This is certainly true concerning the existence of differentiated supervision and evaluation structures and practices. One such practice is the differentiation of evaluation activities according to individual teachers' specific needs. Nearly without fail, the research reviewed shows standards for instructional practice must serve as the basis for teacher supervision and evaluation and all related activities; however support around attaining these standards will necessarily differ based upon the level of individual teacher practice. Danielson and McGreal (2000), Glanz, et. al. (2007), Glickman (2006), Sparks (2004), Sergiovanni and Starratt (2002), all make this contention with differing levels of emphasis. A prominent feature of supervision and evaluation for new teachers must be the internalization and understanding of the district's standards for effective instruction as the language of ongoing instructional improvement. Regarding the present study, these researchers would certainly indicate that this foundation was not present in the participating school. This factor alone, Danielson and McGreal assert, can be seen as a strong predictor of the ambivalent perspective of study participants toward their evaluation system's relationship to instructional improvement.

This is likewise true of other dimensions of differentiated instruction, the use of multiple sources of data in the teacher evaluation process, for example.   The practices of observing, analyzing, and improving instructional practice through observations and conferencing alone cannot and will not support the ongoing improvement of student learning. In fact, Tucker and Stronge (2005) assert that tying specific student performance data from multiple sources to systems of teacher supervision and evaluation is key if the practices are to improve instructional practices that lead to increased student learning. According to these authors, tying evaluation to student learning refocuses professional development and creates a more collegial environment in which professionals use student performance data, along with observations of teacher practice, as the basis for conversations around the improvement of teaching and learning. The findings of the present study align with the research of Danielson and McGreal (2000), Sergiovanni and Starratt (2002), and Tucker and Stronge (2005). The present study found that multiple sources of data were not utilized in the evaluation of teachers at the participating school given the lack of familiarity study participants had with their districts performance standards; the researchers above conclude that, by design, many of the standards necessitate the examination of student behavior and learning when discerning a teacher's effectiveness. To be sure, this is the case with the standards utilized in the participating district. This both predicts and underscores the contradictory findings of the present study in which teachers believe that evaluation activities – conferencing in particular – are useful to them in improving their practice, but that they are undecided on the effectiveness of teacher evaluation as a means of improving student performance.   It is essential to ask, then, how professionals can know if their practice has improved absent discussions on formative and summative student performance data. This dichotomy has implications for practice which are stated later in this chapter.

The Massachusetts Education Reform Act of 1993 provided landmark legislation that mandated access to equitable educational resources throughout the Commonwealth of Massachusetts. These resources included increased instructional time, a guaranteed student entitlement, and state-wide standards for curriculum and for evaluating and improving instructional practice. In particular, the legislation provided that locally bargained instructional standards based upon the state standards - with the specific indicators bargained at the local level - must be in place in each Commonwealth district. It is clear from the results of the present study that these standards are, in fact, in place within the participating district. However, this framework does not guarantee the use of evaluation as a tool to promote effective practice; the data from the present study in large part underscore this contention. Therefore, as legislation which provides a structure for standards-based teacher evaluation remains in place, there are several key implications for research-based practices that are discussed further in this chapter.

Study Limitations

There are four overarching limitations that are inherent in the present study. First, there are those limitations that are intrinsic to survey research. As Crocker and Algina state: "Because psychological constructs are abstractions which can only be assessed indirectly, the design of instruments to measure such variables presents several challenging problems" (1986, p.5). First, the scope of the survey instrument employed was limited in terms of the attitudes that it was designed to measure. In other words, the items presented on the survey instrument could not account for all possible teacher attitudes concerning standards-based evaluation. For example, the instrument measured the attitudes of teachers regarding a common understanding of teacher performance rating criteria; however, the instrument does not measure teachers' understanding of or agreement with the criteria.

The second limitation stems from the design of the data collection process itself. As Crocker and Algina state, "Most psychological measurements are based on a limited sample of observations and usually are taken at one point in time" (1986, p.6). As surveys in the present study were administered after the end of a regular school day, participant fatigue and preoccupation with the day's events, for example, were likely to affect participant attention, investment, and response. A teacher receiving a poor performance rating on the day of survey administration might, for example, respond to items differently in the morning than in the afternoon after the rating had been delivered.

The study was further restricted by the number of survey respondents, in this case, thirty. By making several visits to the school site, explaining the study and its benefits to the entire teaching staff in the participating school, and providing incentives for teacher participation, the researcher sought to maximize participation rates. However, the low number of participants narrowed the ability of the researcher to examine item response patterns among demographic groups, i.e. participants of different education levels, races, genders, and years of service and response groups had to be collapsed. Further, the small number of survey respondents significantly limited the use and efficacy of statistical operations. While the researcher initially planned to derive data from chi square analyses, analyses of variance (ANOVA), *t-tests,* and cross-tabulations, the data generated from these operations were neither statistically significant nor scientifically sound as they were underpowered due to the small sample size. Therefore, descriptive data in the form of aggregate item statistics tables and total mean score tables by demographic descriptors served as the primary sources of response data for the present study.

Implications for Practice

The findings of the present study present significant implications for practice if, indeed, teacher evaluation is to serve as a primary means of improving teaching and learning. The findings have ramifications for the areas of: differentiated teacher evaluation practices; alignment of individual teacher evaluation goals with school and district professional development and instructional agendas; and the reallocation of time, funding, and personnel in order to implement said changes in practice.

The present study findings suggest the existence of several major areas of concern regarding the function and understanding of performance standards within teacher evaluation systems. This is evidenced by participants' lack of familiarity with the standards, their belief that the performance standards do not serve as the sole basis for teacher performance supervision and evaluation, and their perspective regarding the lack of a common language or understanding of instructional practices within their school.

These issues suggest that adoption of differentiated supervision and evaluation practices could improve the efficacy of teacher evaluation systems. While differentiation entails delineated support for teaching professionals in different stages of career and development, it does not entail differentiation of the performance standards themselves. As the standards serve as the foundation for all teacher evaluations, evaluative support and professional development would look different among teachers or groups of teachers. Within a system such as this, newer teachers, teachers whose performance is in question, and teachers who are self-actualized and who are self-motivated to improve their practice on an ongoing basis would be afforded differentiated supports based upon their place and station within their career and practice.

For example, to mitigate problems arising from teachers' lack of familiarity with performance standards, a differentiated evaluation model would frame support for first year/provisional teachers around understanding of and familiarization with the chosen standards for teacher performance. Beyond knowledge of the standards themselves, it is suggested that said teachers are trained in the actual implementation of evaluation instruments and associated activities, i.e. formal and informal classroom observations, conferencing, etc. As study participants believed that multiple sources of data were not used as part of the evaluation process, the acquisition and discussion of instructional artifacts could serve as an integral part of this stage. Examinations of items such as examples of teacher assessments, samples of student work, and lesson plans would paint a more complete picture of teacher practice and the alignment of these with school improvement plans and their resultant strategies. This stage of a differentiated system could also benefit more veteran teachers whose practice is floundering or in question.

Implementation of differentiated supervision and evaluation systems should also serve more skilled and self-actualizing practitioners as a means of expanding their instructional repertoire and skills. Veteran professionals who demonstrate strong reflective skills and self-motivation in their work would benefit from supervision and evaluation processes that allow for a more constructivist approach to instructional improvement. This could entail a provision for self-evaluation in an off-cycle year for said professionals. It could also allow for regular peer observation and feedback from colleagues in the same mode of evaluation because as indicated by the present study, teachers believe that the feedback they receive from peer observation helps them to improve their classroom practice. As professionals in this group would be, by definition, self-directing, the ownership they attain from directing their own professional development will

greatly enhance the collegial culture in their school and substantially benefit their instructional practice. To be certain, teachers on this improvement track would not operate in isolation of school administrative leadership. In fact, findings of the present study suggest that the presence of administrative support is essential for the improvement of instruction to take place through the formal supervision and evaluation process. For example, study data indicate that teachers believe administrators' feedback on classroom observations is a useful tool in aiding them to improve their instructional practice. In addition, present study findings suggest that during teacher and administrator scheduled conferencing time, administrators do not work with teachers to ensure that teacher performance goals are aligned with their school and district's overarching school professional development plans and goals. This is a significant implication for practice as the knowledge-base outlines this as an essential and effective practice.

As stated, supports within differentiated evaluation systems should address the needs of teachers at varying stages of their professional development. To be sure, however, there are common elements which must exist as foundational pieces within any such systems. For example, study participants indicated that their evaluations were not wholly based upon the standards for evaluation as written. Further, participants believed that a common understanding of behaviors that modeled the standards did not exist within their school. Last, participants indicated that multiple sources of data were not used as part of their performance evaluations. Pursuant to these findings and the elements of evaluation that they represent, it is suggested that practice be changed in several ways. First and foremost, teacher performance standards should serve as the centerpiece of a standards-based evaluation system. This cannot happen in a nominal way, as data suggest is the case within the participating school. Rather, study and review of the standards should happen on an ongoing basis for professionals at every level of

performance and tenure with an emphasis placed on internalizing the standards during the first year of practice. As suggested by the second finding mentioned above, once a mechanism for review of the standards is in place, then opportunities for identifying common behaviors that reflect these practices must be implemented. These might include frequent classroom observations by teachers and supervisors alike, followed by regularly scheduled "sacred" time for discussion of said observations using the language of the standards and their associated indicators as the common lexicon for instructional behavior.

Though the present study findings do not reflect the perceptions of administrators, according to participant responses, there are some direct implications for administrator practice. Data first suggest, for example, that teachers value the feedback of administrators concerning classroom practice. Participants also indicated that evaluative conferencing between administrators and teachers were done well at their school and were an important part of the evaluation process. These points demonstrate the essential nature of regular administrative direction and engagement in supervision and evaluation processes. Intermittent, compliant, and formulaic involvement in teacher evaluation reinforces historical beliefs in the perfunctory nature of supervision and evaluation in schools as examined earlier in this work. The implication here is that administrators should be actively, regularly engaged in teacher supervision and evaluation; instructional leadership asks no less. This demands that administrators organize their time and other resources to fulfill obligations to teacher improvement through formal supervision and evaluation systems.

For all the implications outlined in this section, it is recommended that professionals within schools practice reaching out to the professional knowledge base to research contemporary theory and best practices as a foundation for school-wide and individual

discussions on improvement of teacher practice. In addition, in order to implement a differentiated model of supervision and evaluation, funds must be earmarked for: administrator training in the use of supervision and evaluation as an essential tool for instructional improvement (to include instructional strategies and observation and conferencing skills); ongoing professional development for teachers in instructional improvement strategies aligned with the school and district's professional development priorities; and resources to provide for release time of teachers when conducting peer observations. Last, school leaders and teachers should schedule "sacred time" which is utilized for teacher-supervisor conferencing around classroom observations, alignment of practice with school professional development goals, the creation and periodic review of individual teacher goals for improved performance, and ongoing discussion of the district's performance evaluation standards and how their indicators define instructional practices within the school. An essential part of this is identifying and analyzing the student performance data that demonstrate the effectiveness or ineffectiveness of the instructional practices mentioned above. While demanding and complicated, this ongoing work is crucial if the improvement of instructional practice is centered upon increased student learning and achievement.

Implications for Further Research

As previously stated, the present study embodies several limitations which restrict the spectrum of conclusions that can be drawn from the data gathered. For this reason, the researcher makes several recommendations for further research. It is believed that these recommendations, if implemented, will expand the body of professional knowledge surrounding perceptions of standards-based teacher evaluation. These recommendations have implications for the expansion of both qualitative and quantitative research methodology that will, by default,

provide data that will serve as a basis for further improvements in policy and practice in this arena.

Consequently, the researcher offers recommendations in several areas including: the scope of research design in terms of target populations, survey distribution, and additional methodology; policy analysis; classroom implementation of standards and examination of the language of teacher evaluation.

It is recommended that the scope of further research be broadened in several key areas. First, the target population can be expanded to include elementary and middle school teachers. This would outline differences in perceptions of standards-based evaluation across grade levels. A subsequent sharing of best practices in evaluation and in teaching and learning could be generated by this research, and, depending on the research outcomes, increase the knowledge base surrounding the relationship between standards-based teacher evaluation and ongoing instructional improvement.

The voice of administrators and teacher evaluators should also be included in any subsequent research design. The reason for this is twofold. First, adding the perceptions of administrators would provide a rich base of data which can be compared and juxtaposed to teacher perceptions. By doing this, researchers can further examine teachers' perceived disconnects with their evaluators. Areas of possible discrepancy include: teacher and administrator familiarity with the standards for teacher evaluation; the use of teacher evaluation standards as the sole basis for formal and informal observations and evaluations; perceptions of the amount of time administrators spend in classrooms; and the existence and use of a common language and understanding of the standards. In practice, this connection or disconnect could then be analyzed and researchers could suggest implementation of behaviors and practices that

would foster reflection and alignment of evaluation language and practices. Subsequent researchers should also expand survey distribution to include several school sites at each level. By dint of having a larger sample from which to gather data, the research findings can speak to a more broad range of perceptions and therefore can be a more useful addition to the professional knowledge base. Researchers should be certain to conduct one or more follow-up focus groups through which quantitative data are vetted. While the present study utilized written feedback from participants through which they expanded upon their quantitative responses, the suggested focus groups would provide an additional lens through which researchers can view their findings.

Next, the researcher recommends that any subsequent review of the literature include a more contemporary examination of policy analysis. The literature review within the present study provided a historical perspective on the development and necessity for standards for teacher practice. This included an examination of policy development and implementation within the Commonwealth of Massachusetts. It is recommended that this development and implementation of standards be examined across several states. By doing this, researchers can ascertain both the commonalities and differences among standards and practices. This will better inform implications for local, state, and possibly federal policy development and revision in this area as a means of further improvement of teacher evaluation systems.

Last, the investigator in the present study recommends that further research incorporate the study of classroom observations and the associated pre and post conferences and the use of formative and summative student performance data in teacher evaluation. Implementation of these qualitative and quantitative measures would serve as a means of deepening research into the links and disconnects between evaluation standards, instructional practice, teacher and administrator perceptions, and ultimately to student achievement. Here, researchers could

ascertain the implementation of the standards, teacher and evaluator understanding of the standards in practice, and the extent to which the language serves as a common language for the evaluation process and the improvement of student learning. These data can then be utilized to support or refute those data gathered in the quantitative portion of the study. In comparing the two types of data, researchers can draw further attention to the need for ongoing dialogue between and among teachers and evaluators toward the end of using standards-based teacher evaluation as a means of improving student learning and achievement.

Conclusion

The findings and implications generated from the present study pose many questions regarding policy and practice in the realm of standards-based teacher supervision and evaluation. In doing this, this research has served to shed light on the state of standards-based evaluation in one school district over a decade after its adoption and implementation. Aggregate findings indicate that much work needs to be done to further the efficacy of standards-based evaluation as a consistent means of improving instructional practice and student achievement. To be sure, the present study does shed light on some effective practices that serve the above-mentioned end.

For example, findings demonstrated that conferences between teachers and administrators are a powerful part of supervision and evaluation systems. Participants also demonstrated that, in general, feedback from classroom observations serves as a tool for improving their own practice. Likewise, they indicated that feedback from peer observations is an effective practice that aides teachers in improving their instruction.

The present study also revealed that for standards-based teacher evaluation systems to serve as a primary vehicle for instructional improvement, several inconsistencies in practices and beliefs must be reconciled. For example, participants agreed on the one hand that conferencing

and feedback served to improve their instructional practice; however their responses were, in aggregate, inconclusive regarding whether teacher evaluation systems are an effective tool in increasing student achievement. This has implications for staff beliefs regarding the ability of all students to learn at rigorous levels given appropriate and rigorous instruction and support. Maintaining this belief necessitates a belief in the ability of the teaching profession to improve in an ongoing way in order to maximize the achievement of all students. A contrasting belief, in its essence, negates the legitimacy of the teaching profession as a whole. Other conflicting findings unearthed in this study include that while participants believe that their supervisors possess sophisticated understandings of teaching and learning, they do not perceive that multiple sources of data are used in the evaluation of their practice. And, while participants believe that feedback on their performance helps to improve their instructional practice, they do not agree that the Seven Standard Areas for Teacher Performance are used as a common language for teaching and learning or that their evaluations are wholly based on these standards. These findings are all underscored by the perception that teachers are not familiar with the evaluation standards used in their district. Again, these findings have implications for practice, policy, and further research that, if left unaddressed, will designate standards-based teacher evaluation as a perfunctory, compliance-based protocol that does nothing to improve instructional practice or student learning. If attended to on a deep level, the exact opposite will be true and standards-based supervision and evaluation will come to serve as an important tool for the ongoing improvement of teaching and learning and ultimately, the augmentation of student achievement.

REFERENCE LIST

Abbott, M. (1969). Hierarchical impediments to innovation in educational organizations. In F. D. Carver & T. Sergiovanni (Eds.), *Organizations and human behavior: Focus on schools*. 410. New York: McGraw Hill.

Ballard, L., & Gullat, D. (1998). Choosing the right process for teacher evaluation. *American Secondary Education, 26*(3), 13-17.

Beck, L. G., & Murphy, J. (1993). *Understanding the principalship: Metaphorical themes.* New York: Teachers College Press.

Black, S. (1993). How teachers are reshaping evaluation procedures. *Educational Leadership, 51*(2), 38.

Blumberg, A. (1980). *Supervisors and teachers: A private cold war* (Second ed.). Berkeley, CA: McCutchan Publishing Corporation.

Blumberg, A. (1985). Where we came from: Notes on supervison in the 1840's. *Journal of Curriculum and Supervision, 1*(1), 56-65.

Bolman, L. G., & Deal, T. E. (1997). *Reframing organizations: Artistry, choice, and leadership* (Second ed.). San Francisco, CA: Jossey-Bass, Inc.

Bredeson, P. (1996). Superintendents' roles in curriculum development and instructional leadership: Instructional visionaries, collaborators, supporters, and delegators. *Journal of School Leadership, 6*, 243-264.

Colby, S. A., Bradshaw, L. K., Joyner, R. L.. (2002). *Perceptions of teacher evaluation systems and their impact on school improvement, professional development, and student learning.*

Paper presented at the Annual Meeting of the American Educational Research

Association, April. New Orleans, LA.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York:

Holt, Rinehart and Winston.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria:

Association for Supervision and Curriculum Development.

Danielson, C., & McGreal, T. (2000). *Teacher evaluation: To enhance professional practice*.

Alexandria, VA: Asssociation for Supervision and Curriculum Development.

Daresh, J. C. (2001). *Supervision as proactive leadership* (Third ed.). Prospect Heights, Illinois:

Waveland Press, Inc.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the

organizational context. *Review of Educational Research*(53), 285-328.

Dawson, V., & Acker-Hocevar, M. (1998). Contradictions within a state-mandated teacher

evaluation system: A historical perspective. (pp. 28). Annual Meeting Mid-South

Education Association, March. New Orleans, LA.

DeMitchell, T. (1995). Competence, documentation, and dismissal: A legal template.

International Journal of Educational Reform, 4(1), 88-95.

Department of Education, M. (2001). Goals 2000: Five year master plan. From

http://www.doe.mass.edu/edreform.

Frase, L., & Streshly, W. (1994). Lack of accuracy, feedback, and commitment in teacher

evaluation. *Journal of Personnel Evaluation in Education, 1*, 47-57.

Fullan, M. (2003). *The moral imperative of school leadership*. Thousand Oaks, CA: Corwin

Press.

Glanz, J., Shulman, V., & Sullivan, S. (2007). *Impact of instructional supervision on student ahievement: Can we make the connection?* Paper presented at the Conference of the Council of Professors of Instructional Supervision, October, Staten Island, NY.

Glatthorn, A. (1984). *Differential supervision*. Alexandria, VA: Association for Supervision and Curriculum Development.

Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2001). *Supervision and instructional leadership: A developmental approach.* (5 ed.). Needham Heights: Allyn & Bacon.

Goldrick, L. (2002). *Improving teacher evaluation to improve teacher quality. Issue Brief.* (No. 141). Washington D.C.: National Governors' Association Center for Best Practices.

Howard, P. (1996). Interpreting the evaluation experience through embodiment, conversation, and anecdote. *International Journal of Qualitative Studies in Education, 9*(2), 167-180.

Kelly, T. (1999). Why state mandates don't work. *Phi Delta Kappan, 80*(7), 543-544,546.

Lashway, L. (2002). *Developing Instructional Leaders.* ERIC Clearinghouse on Educational Management.

Marshall, K. (2003). Recovering from HSPS (Hyperactive Superficial Principal Syndrome): A progress report. *Phi Delta Kappan, 84*(09), pp. 701-709.

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

Marshall, K. (2006). The why's and how's of teacher evaluation rubrics. *Edge, 2*(1), pp. 1-19.

MassPartners. (2000). *Unabridged study of systems for evaluating Massachusetts teachers*. Marlborough, MA: Massachusetts Partners for Public Schools.

McGregor, D. (1960). *The human side of enterprise*. New York: McGraw-Hill.

Meloche, J. N. (2000). Teacher evaluation: Educational process for pedagogical improvement or punitive, political tool? Retrieved October 22, 2003, from http://muse.widener.edu/~egr00001/EDControversy/Meloche.html

Merrow, J. (2000). Teacher seniority: Benefit or barrier. *Principal, 79*(3), 49-50.

Mertler, C., and Petersen, G. (1997). *A collaborative model of teacher evaluation: Roles and challenges faced by various constituent groups.* (Descriptive Report). Chicago, IL: Mid-western Educational Research Association.

MESPA. (1998). *Survey of the elementary and middle school principalship* (Survey). Marlborough, Massachusetts: Massachusetts Elementary School Principals' Association.

NCEE, N. C. o. E. i. E. (1983). *A nation at risk: The imperative of educational reform.* Retrieved. from www.ed.gov/pubs/NatAtRisk/index.html

Nevo, D. (1994). How can teachers benefit from teacher evaluation. *Journal of Personnel Evaluation in Education, 8*, 109-117.

Nikolic, V. (2001). *Investigation of effective supervisory practices in a multicultural context.* Toronto: University of Toronto.

Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin, 88*(639), 60-79.

Postel, H. H. (1955). Motivating reading through the school newspaper. In *Reading for today's children: Thirty-fourth yearbook of the Department of Elementary school principals.* Washington, D.C.: National Education Association.

Pratt, F. (1996). The ethical principal and teacher non-retention. *Journal of Personnel Evaluation in Education, 10*, 29-36.

Prybylo, D. (1998). Beyond a positivistic approach to teacher evaluation. *Journal of School Leadership, 8*(6), 558-583.

Rettig, P. (1999). Differentiated supervision: A new approach. *Principal, 78*(3), 36-39.

Rooney, J. (1993). Teacher evaluation: No more "super" vision. *Educational Leadership, 51*(2),

Schott, A. H. (1956). Adventure in arithmetic. In *Instructional materials for elementary schools: Thirty fifth yearbook of the department of elementary school principals.* (pp. 86-90). Washington, D.C.: National Education Association.

Sergiovanni, T. (1985). Landscapes, mindscapes, and reflective practice in supervision. *Journal of Curriculum and Supervision, 1*(1), 5-17.

Sergiovanni, T., & Starratt, R. (2007). *Supervision: A redefinition.* (Eighth Edition ed.). Boston, MA: McGraw Hill.

Sergiovanni, T. J., & Starratt, R. J. (1993). *Supervision: A redefinition* (Sixth ed.). Boston: McGraw -Hill.

Shen, J. (1998). Do teachers feel empowered? *Educational Leadership, 55*(7), 35.

Shinkfield, A., & Stufflebeam, D. (1995). *Teacher evaluation: Guide to effective practice.* Norwell, MA: Kluwer Academic Publichers.

Sparks, D. (2004). The Looming Danger of a Two-Tiered Professional Development System. *Phi Delta Kappan, 86*(4), pp. 304-306.

Starratt, R. J. (1996). *Transforming educational administration: Meaning, community, and excellence*. New York: McGraw-Hill.

Stufflebeam, D. L., & Sanders, J. R. (1990). Using the Personnel Evaluation Standards to Improve Teacher Evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation.* (pp. 416-428). Newbury Park, CA: SAGE Publications, Inc.

Travers, R. M. W. (1981). Criteria of Good Teachers. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 14-22). Beverly Hills, CA: SAGE Publication.

Vandeweghe, R., & Varney, K. (2006). The Evolution of a School-Based Study Group. *Phi Delta Kappan, 88*(04), pp. 282-286.

Wang, W., Cathy Day. (2002). *Issues and concerns about classroom observation: Teachers' perspectives.* Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (TESOL), April. St. Louis, MO.

Weiss, E., & Weiss, S. (1998). *New directions in teacher evaluation* (ERIC Publication. Eric Digest No. EDO-SP-97-9). Washington D.C.: ERIC Clearinghouse on Teaching and Teacher Education.

Worthen, B. R., and James Sanders. (1987). *Educational evaluation: Alternative approaches and practical guidelines.* New York: Longman Publication Group.

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

Appendix A: The Principles of Effective Teaching

 I. CURRENCY IN THE CURRICULUM

    A. The teacher is up to date regarding curriculum content

II. EFFECTIVE PLANNING AND ASSESSMENT OF CURRICULUM AND INSTRUCTION

    A. The teacher plans instruction effectively.

    B. The teacher plans assessment of student learning effectively.

    C. The teacher monitors students' understanding of the curriculum effectively and adjusts

      instruction, materials, or assessments when appropriate.

III. EFFECTIVE MANAGEMENT OF CLASSROOM ENVIRONMENT

    A. The teacher creates an environment that is positive for student learning and involvement.

    B. The teacher maintains appropriate standards of behavior, mutual respect and safety.

IV. EFFECTIVE INSTRUCTION

    A. The teacher makes learning goals clear to students.

    B.   The teacher uses appropriate instructional techniques.

    C. The teacher uses appropriate questioning techniques.

    D. The teacher evaluates, tries innovative approaches, and refines instructional strategies,

      including the effective use of technologies, to increase student learning and confidence to

      learn.

V. PROMOTION OF HIGH STANDARDS AND EXPECTATIONS FOR STUDENT

  ACHIEVEMENT

    A. The teacher communicates learning goals and high standards and expectations to

      students.

    B. The teacher promotes confidence and perseverance in the student that stimulate increased personal student responsibility for achieving the goals of the curriculum.

VI. PROMOTION OF EQUITY AND APPRECIATION OF DIVERSITY

    A. The teacher strives to ensure equitable opportunities for student learning.

    B. The teacher demonstrates appreciation for and sensitivity to the diversity among individuals.

VII. FULFILLMENT OF PROFESSIONAL RESPONSIBILITIES

    A. The teacher is constructive and cooperative in interactions with parents and receptive to their contributions.

    B. The teacher shares responsibility for accomplishing the goals and priorities of his/her grade/team/department, building and school district.

    C. The teacher is a reflective and continuous learner.

Appendix B: Participating District Teacher Performance Evaluation Standards

A. KNOWLEDGE OF SUBJECT MATTER, CURRENCY IN THE CURRICULUM AND

KNOWLEDGE OF CHILD DEVELOPMENT

1. The teacher is up to date regarding curriculum content.

2. The teacher shows a good understanding of the subject matter.

3. The teacher demonstrates an understanding of child growth and development.

B. SETTING THE STAGE FOR LEARNING

1. The teacher plans instruction effectively.

2. The teacher prepares materials that help all students learn the curriculum.

3. The teacher creates a classroom environment that is conducive to learning.

C. CLASSROOM MANAGEMENT

1. The teacher maintains clear standards of good behavior, mutual respect, and safety.

2. The teacher creates a positive environment for student learning and involvement.

3. The teacher functions in a controlled and effective manner under pressure.

D. EFFECTIVE TEACHING

1. The teacher communicates learning goals and high standards and expectations to students.

2. The teacher frames the learning for students.

3. The teacher uses appropriate instructional techniques that involve and motivate students.

4. The teacher develops students' independence and their higher-order thinking skills.

5. The teacher provides enough time and help for students to learn the subject matter.

6. The teacher evaluates, tries innovative approaches and refines instructional strategies

including the use of technologies to increase student learning and confidence to learn.

E. MONITORING, ASSESSMENT, AND FOLLOW-UP

1. The teacher continuously monitors students' learning to ensure student progress and inform instruction.

2. The teacher uses fair and effective assessments of student learning.

3. The teacher uses data from the classroom to adjust and improve teaching.

4. The teacher follows up with students who are not meeting expectations.

5. The teacher succeeds in realizing appropriate student learning gains.

F. PROMOTION OF HIGH STANDARDS AND EXPECTATIONS FOR STUDENT ACHIEVEMENT

1. The teacher organizes the instruction around the belief that sustained and directed effort can yield high achievement for all students.

2. The teacher communicates learning goals and high standards and expectations to students.

3. The teacher promotes confidence and perseverance in the students to stimulate increased personal student responsibility for achieving the goals of the curriculum.

G-1. COLLABORATION WITH PARENTS

1. The teacher keeps parents informed of what is expected and how students are doing.

2. The teacher reaches out to parents and makes them part of their children's education.

3. The teacher looks for other ways that parents can contribute to the class and school.

G-2. COLLABORATION WITH COLLEAGUES

1. The teacher is a collegial and contributing member of the staff.

2. The teacher joins in working toward grade/team/department, school, and [District A] goals.

3. The teacher works in a positive, ethical, and above-board manner.

G-3. SCHOOL RESPONSIBILITIES

1. The teacher sets an example to students by excellent attendance and punctuality.

2. The teacher performs routine duties in a professional manner.

3. The teacher includes all students in the full classroom and extra-curricular program.

4. The teacher cooperates in implementing the [school's improvement plan].

G-4. PROFESSIONAL GROWTH

1. The teacher tries innovative classroom strategies and evaluates their impact.

2. The teacher stays up-to-date on curriculum content.

3. The teacher is a reflective and continuous learner.

Appendix C: Pilot Study Survey Items

Please check the appropriate response to each of the following:

1. I am a(n)

   ☐ teacher.   ☐ principal.   ☐ other evaluator.

2. I work at the _____ school level.

   ☐ elementary   ☐ middle   ☐ secondary

3. I would characterize my district as

   ☐ a large urban district.   ☐ a small rural district.   ☐ a small wealthy suburban district.

Please refer to the following key when circling your response to each of the statements:

1 = Strongly Agree   **2** = Agree   **3** = Neutral   **4** = Disagree   **5** = Strongly Disagree

	SA	A	N	D	SD
1. Our evaluation process is based upon the Principles of Effective Teaching.	1	2	3	4	5
2. I am familiar with the Principles of Effective Teaching and how they are used in our evaluation process.	1	2	3	4	5
3. In my district, there is a common understanding between teachers and administrators of how to identify specific behaviors that model the descriptors of the Principles of Effective Teaching.	1	2	3	4	5
4. In my district, there is regular discussion about what the Principles of Effective Teaching look like in the classroom.	1	2	3	4	5
5. Teachers and administrators within my district have a common understanding regarding definitions of specific criteria and expectations for performance ratings in the evaluation process.	1	2	3	4	5
6. The goal setting process required of teachers by evaluation systems at the beginning of each year should be linked to teacher evaluation.	1	2	3	4	5
7. Teacher evaluation aids in the improvement of teaching and learning.	1	2	3	4	5
8. Ongoing professional development in techniques of teacher evaluation is needed for evaluation to improve teaching and learning.	1	2	3	4	5
9. Peer observation of performance is a more effective tool for improving teaching than formal observation.	1	2	3	4	5

	SA	A	N	D	SD

10. Administrators spend sufficient time in classrooms
     in order to carry out the district's expectations for evaluation. **1 2 3 4 5**

11. Conferences between teachers and administrators are an important
     component of teacher evaluation. **1 2 3 4 5**

12. Conferences between teachers and administrators
     as part of evaluation are done well. **1 2 3 4 5**

13. When done well, conferences are helpful in improving teaching and
     learning in my district. **1 2 3 4 5**

14. There is a shared understanding and a common language
     around effective teaching in my district. **1 2 3 4 5**

15. Teachers' union contracts do substantially inhibit use of teacher evaluation
     as a tool for improving teaching and learning. **1 2 3 4 5**

16. Due to obstacles imposed by the teacher contract, administrators
     are reluctant to dismiss a teacher with professional status whose
     performance has declined and has failed to improve. **1 2 3 4 5**

17. Due to personal reasons, i.e., teacher's possible loss of pension, teacher's
     personal financial obligations, etc., administrators are reluctant to dismiss
     a teacher with professional status whose performance has declined and
     has failed to improve. **1 2 3 4 5**

18. Due to a lack of a sophisticated understanding of effective classroom
     instruction evaluation, administrators are reluctant to dismiss a teacher
     with professional status whose performance has declined and has failed
     to improve. **1 2 3 4 5**

19. Due to a high turnover rate among staff without professional status,
     administrators are reluctant to hold teachers without professional status
     to high standards for evaluation. **1 2 3 4 5**

Please answer each of the following. Use additional sheets of paper as needed.

20. What is a major strength of the evaluation process in your district?

21. What is a major drawback in using evaluation as a means to improve teaching and learning in your district?

22. What major change would you make to improve evaluation in your district?

23. Please provide any additional information that you would like to share regarding your evaluation experiences
     and process or teacher evaluation in general.

Appendix D: Teacher Evaluation System Attitude Scale

Directions for Completing Sections A, B, & C

*Respond to each item in these sections by placing an **"X"** in the box that best reflects your level of agreement with the item. Please fill in only one box per item. Refer to the chart to the right when responding.*

SD	Strongly Disagree with the item
D	Disagree with the Item
A	Agree with the Item
SA	Strongly Agree with the Item

SECTION A: Items 1-6, The Seven Standards for Teacher Performance

SD = Strongly Disagree with the item; **D = Disagree** with the Item; **A = Agree** with the Item; **SA = Strongly Agree** with the Item

	SD	D	A	SA
1. I am very familiar with the Seven Standard Areas of Teacher Performance as used in the Boston Public Schools.	☐	☐	☐	☐
2. Our evaluation process is wholly based upon the Seven Standard Areas of Teacher Performance.	☐	☐	☐	☐
3. I understand how the Seven Standard Areas of Teacher Performance are used in our teacher evaluation process.	☐	☐	☐	☐
4. In my school, there is regular discussion about what the Seven Standard Areas of Teacher Performance look like in the classroom.	☐	☐	☐	☐
5. In my school, there is a common understanding among teachers and administrators regarding specific behaviors that model the Seven Standard Areas of Teacher Performance and their descriptors.	☐	☐	☐	☐
6. The Seven Standard Areas of Teacher Performance and their descriptors serve as the common language for effective teaching in my school.	☐	☐	☐	☐

SECTION B: Items 7-14, Evaluation Activities and Communication

SD = Strongly Disagree with the item; **D = Disagree** with the Item; **A = Agree** with the Item; **SA = Strongly Agree** with the Item

	SD	*D*	**A**	**SA**
7. In my school, an annual goal setting process for teachers should be linked to teacher evaluation.	☐	☐	☐	☐
8. Administrators in my school communicate a sophisticated understanding of effective teaching.	☐	☐	☐	☐
9. Conferences between teachers and administrators are an important component of teacher evaluation.	☐	☐	☐	☐
10. When done well, conferences between teachers and administrators as part of evaluation are helpful in improving teaching and learning.	☐	☐	☐	☐
11. In my school, conferences between teachers and administrators as part of evaluation are done well.	☐	☐	☐	☐
12. Administrators spend sufficient time in classrooms in order to carry out the district's expectations for evaluation.	☐	☐	☐	☐
13. Administrators spend sufficient time in classrooms in order to carry out teachers' expectations for evaluation.	☐	☐	☐	☐
14. Teachers and administrators in my school share a common understanding of specific criteria for performance ratings ("*meets/does not meet expectations*") in the evaluation process.	☐	☐	☐	☐

SECTION C: Items 15-21, Evaluation & Instructional Improvement

SD = Strongly Disagree with the item; **D = Disagree** with the Item; **A = Agree** with the Item; **SA = Strongly Agree** with the Item

	SD	D	A	SA
15. The process of teacher evaluation leads to increased student achievement in my school.	☐	☐	☐	☐
16. Teacher evaluation is closely linked to my school's professional development initiatives.	☐	☐	☐	☐
17. The feedback I get from evaluation helps me to improve the quality of my teaching.	☐	☐	☐	☐
18. Feedback from peer observation helps me to improve the quality of my teaching.	☐	☐	☐	☐
19. Administrators need ongoing professional development around teacher evaluation for evaluation to improve my practice.	☐	☐	☐	☐
20. Multiple sources of student performance data are used as part of my evaluation.	☐	☐	☐	☐
21. I rely upon reflection on my own teaching to improve student performance.	☐	☐	☐	☐

SECTION D: Items 22-25, Open Response Items
Provide a response to the following items. Please use only the space provided.

22. Describe a major strength of the evaluation process in your school.

23. Outline a substantial change you would make to the evaluation system in your school.

SECTION D: Open Response Items (Continued)
Provide a response to the following items. Please use only the space provided.

24. What is a major drawback to using evaluation as a means to improve teaching and learning in your school?

25. Share any additional feedback, warm and cool, regarding your experiences with the evaluation process in your school.

SECTION E: Items 26 - 31 Professional Demographic Information
Please complete all demographic information below by placing an "X" in the box that best applies. Fill in only one circle per item.

26. Gender	☐ **Female** ☐ **Male**
27. Race	☐ **African** ☐ **Caucasian** ☐ **Latino**
	☐**Asian/Pacific Islander** ☐ **Native American**
	☐**Other**_____
28. Years of Service in this district	☐ **1 – 5** ☐ **6 – 10** ☐ **10 – 20** ☐ **20+ years**
29. Years of Service in this building	☐ **1 – 5** ☐ **6 – 10** ☐ **10 – 20** ☐ **20+ years**
30. Last Level of Education completed	☐ BA/BS ☐ MA/MS ☐ Doctorate ☐ Post Doctorate

31. Primary Field of Instruction:

☐ ELA	☐ Math	☐ Special Education *(All Areas)*
☐ Science	☐ JROTC	☐ History/Social Studies
☐ Technology	☐ SEI/ESL	☐ Physical Education/Health
☐ Other Area **(Specify)**_____		

Appendix E: Informed Consent Document

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

Dear Professional Educator**:**

You are being invited to take part in a research study about teacher evaluation in the Boston Public Schools. The study is entitled "Urban Secondary Teachers' Perceptions of a Standards-Based Teacher Evaluation System." You have been chosen to participate because you are employed as a full time teaching professional in the City of Boston Public Schools. If you take part in this study, you will be one of over 100 people to do so.

Your participation is completely voluntary. Your decision whether or not to participate will have no effect on your employment status or professional standing within the Boston Public Schools or within your school. Please feel free to ask questions at any point if there is anything that you do not understand.

The person conducting this study is Jonathan D. Pizzi, a Ph.D. candidate in Educational Administration at Boston College. He is being guided in this research by Dr. Irwin Blumer, Chairman of the Educational Administration and Higher Education Department of the Lynch Graduate School of Education at Boston College. No funding has been received for this study, and neither Mr. Pizzi nor Dr. Blumer expects to receive any extra money from companies or organizations because of the results of this study.

Purpose:
By participating in this research, you will be providing valuable insight into teachers' perceptions of the standards-based teacher evaluation system that is utilized in your school and district.

Procedures:
Research will be conducted in three phases, the first at one high school within the Boston Public Schools (School A), and the second and third at a second Boston public high school (School B). The total amount of time you will be asked to volunteer for this study is expected to be one to two hours. You are only eligible to participate in this research study if you currently serve as a full time teaching employee at the secondary level within the Boston Public Schools.

If you participate in the first phase of the study, you will be part of a focus group of five to eight members from School A. You and the other participants will be asked to comment on the structure, content, and clarity of the survey that will be used in the second phase of research.

In the survey phase of the study, you will be asked for your responses in five areas. In the first three sections, you will indicate your level of agreement with statements about your evaluation system. The items relate to your familiarity with the evaluation system, your attitude concerning evaluation activities, and your beliefs about the overall effectiveness of teacher evaluation. The fourth section consists of four short open-response items concerning specific aspects of your

evaluation system. In the last section of the survey you are asked some questions about your background, such as where you have taught, your years of service, and your area of instruction.

Members who participate in the final phase of the survey will be part of a focus group of five to eight full-time teachers from School B. As a member of this group, you will be asked to respond to and clarify the data gathered from the survey.

Risks: To the best of the researchers' knowledge, the things you will be doing in this study have no more risk of harm to you than what you would experience in everyday life.

Benefits: You will not receive any direct benefit from being in this research study. However, the researchers hope to gather information that will help them to understand professionals' perceptions about their evaluation system.

Costs: You do not have to pay to participate in this research study.

Compensation: You will not receive any compensation for your participation in this study.

Withdrawal from the study:
You may choose to stop your participation in this study at any time. Your decision to stop your participation will have no effect on your employment status or professional standing within the Boston Public Schools or within your school.

Confidentiality:
Mr. Pizzi has designed the study to ensure your confidentiality, i.e. to protect your identity as a participant or non-participant in the study and to minimize risks to you. If you are a focus group participant, the focus groups will be recorded and the audiotapes will be transcribed using pseudonyms. The audiotapes will be destroyed immediately following their transcription.   For survey participants, each survey is numbered, however your name will not be on the survey, so your answers will be anonymous. This means that no one, not even the researcher, will know that the information you put on the survey came from you. You will be asked fill out a pre-addressed, stamped post card and return it to Mr. Pizzi separate from your survey. This will let Jonathan know you filled out a survey, but it will not tell him which survey is yours. Jonathan will use the postcards to select participants for a follow up focus group, and he will destroy all post cards immediately following completion of the study.

Once the study is complete, the only direct evidence of your involvement participation will be this signed document. This informed consent document with your name on it will be secured in a filing cabinet at the home office of Jonathan D. Pizzi. No one but Jonathan will have access to the cabinet. These forms will be destroyed three years after the results of the study are published. The original surveys and focus group transcripts will be kept for use in future research and might be shared with other researchers.

Although it happens very rarely, researchers may be required to show information that identifies you, like this informed consent document, to people who need to be sure they have done the

research correctly. These would be people from a group such as the Boston College Institutional Review Board that oversees research involving human participants.

The information you give will be entered into an electronic database and analyzed. In this process, your information will be combined with information from other people taking part in the study. When Mr. Pizzi writes the study to share with other researchers in his dissertation, at meetings, or in journals, he will write about this combined information. Individuals will not be identified in these written materials.

Questions:
You are encouraged to ask questions now, and at anytime during the study. You can reach Jonathan Pizzi at (781) 849-0383, or Dr. Irwin Blumer at (617) 552-1956. If you have any questions about your rights as a participant in a research study, please contact the Boston College Office for Human Research Participant Protection, (617) 552-4778.

Certification:
"I have read and I believe I understand this Informed Consent document. I believe I understand the purpose of the research project and what I will be asked to do. I have been given the opportunity to ask questions and they have been answered satisfactorily. I understand that I may stop my participation in this research study at anytime and that I can refuse to answer any question(s). I understand that my name will not appear on the survey I fill out or on focus group transcripts, and that I will not be identified in reports on this research. I have received a signed copy of this Informed Consent document for my personal reference."

"I hereby give my informed and free consent to be a participant in this study."

Signatures:

_____

Date   **Consent Signature of Participant**

Print Name of Participant

Signature of Person Providing Information and Witness to Consent

Print Name of Person Providing Information and Witness to Consent

Appendix F: Invitational Flyer for Focus Group

Call For Participants!

Jonathan Pizzi, a PhD candidate in Educational Administration at the Lynch Graduate School of Education, Boston College, is conducting a focus group of 5-8 teachers at all career and experience levels. Please see him or reply via the email address below if you would like to participate in a study that seeks to uncover teacher perceptions of standards-based teacher evaluations. You will be given some materials that will help you to prepare for the group.

DATE: Thursday, October 14, 2004

TIME: 2:00 – 3:30PM

LOCATION: Lower Library Conference Room

TOPIC: Urban Secondary Teachers' Perceptions of a Standards-Based Teacher Evaluation
                        System

RSVP to jpizzi@boston.k12.ma.us by Tuesday, October 12, 2004.

Refreshments will be served.

Appendix G: Focus Group Protocol

Jonathan D. Pizzi
Survey Development Focus Group Protocol
October 14, 2004
2:00 – 3:30 PM

1. Review & Explain Informed Consent. Sign & Witness.

2. Purpose of the Focus Group: to gather impressions and feedback regarding a proposed survey that measures teacher perceptions of a standards-based teacher evaluation instrument.

3. Participants take the survey.

4. Guiding Questions (if needed):

    a. What questions do you have regarding the survey questions

    b. What are your overall impressions of the survey and its structure?

    c. What do you think is being asked of you? Why?

    d. Which survey items need clarification? Why?

    e. Which survey items should be revised or reworded? Why?

    f. Should any items be added or removed? Why or why not?

5. Concerns or questions regarding the group

6. Adjournment