# PRINCIPAL READINESS AND PROFESSIONAL DEVELOPMENT TO CONDUCT EFFECTIVE TEACHER EVALUATIONS THAT LEAD TO IMPROVED STUDENT ACHIEVEMENT

A Doctoral Research Project
Presented to
Associate Professor Ann Myers, Ed. D.
Doctoral Research Committee Chair
School of Education
The Sage Colleges

In Partial Fulfillment of the
Requirements for the
Degree of Doctor of Education
In Educational Leadership

Shawn Hunziker

December 12, 2012

© Copyright by Shawn Hunziker, 2012
All Rights Reserved

PRINCIPAL READINESS AND PROFESSIONAL DEVELOPMENT TO CONDUCT EFFECTIVE TEACHER EVALUATIONS THAT LEAD TO IMPROVED STUDENT ACHIEVEMENT

We represent to Sage Graduate School that this thesis/dissertation and abstract are the original work of the author and do not infringe on the copyright or other rights of others.

Shawn Hunziker — Date of Signature

Dr. Ann Myers
Associate Professor
Dissertation Chair — Date of Signature

**Abstract**

Education reform is the focus of many of the political agendas today. The research is clear that the best way to increase student achievement is by having highly effective teachers in the classroom. As a result of prior research, both the state and federal governments have created mandates and legislation aimed at achieving that goal. One of the areas where transformational change is occurring is in the way teachers are evaluated. Evaluation, once a tool to document competency, is now a tool to promote professional growth and assist in personnel decision-making. The focus of the evaluation has turned from teaching to student learning. This change in focus will require principals to acquire a new set of skills in order to be effective teacher evaluators. This exploratory study surveyed principals from Upstate New York to determine what they perceived to be the best forms of professional development to help them hone their skills as teacher evaluators. The results of the study revealed that there was a difference in perceived readiness between genders and years of experience. The findings of the study also discovered weak to moderate correlations between reported professional development and perceived readiness to conduct effective teacher evaluations. Principals have often been the coordinators of their own professional development. This study shows that principals desire a change and want to be able to collaborate and work with their colleagues. As a result of the findings of this study, it is recommended that system leaders create a transitional step by taking an activity that principals previously did in isolation and making it a collective process aimed at increasing principals' skills and readiness to be effective teacher evaluators.

Keywords: educational reform, evaluator, feedback, growth, professional development, principal, Race To The Top, student achievement, teacher evaluation

Dedicated to my grandmother, Eleanor Schmidt

**Acknowledgements**

A doctoral program is a journey that leads to contributing research for the field. Although a dissertation has a sole author, it is not without many passengers. The process of completing a doctoral program requires the encouragement, navigation, advice, and roadside assistance from others along the way. Without the help and support of others, I would have never reached my destination, let alone crossed the finish line.

First and foremost I have to give thanks to my partner and best friend, Seven, for his encouragement and support. It was not only his dedication to experience the victories, bumps, and disasters along the way, but also his mission to keep me focused throughout the journey. It was Seven who reminded me that there was no "perfect" time to apply to the program when I was contemplating it and that I needed to apply at that time. During the journey, when things got rough and blurry, he reminded me that I needed to enjoy it and that I was granted a privilege that not many others would experience. When I neared the completion of the program and others would ask me what was I going to do with my "free time," it was Seven who quickly reminded me that I would be doing the lawn work, painting the window sills, taking care of the dogs, and all those other household chores I was able to forget about for the past two and a half years. You excelled at your job as the co-pilot, and I commend you for your dedication and am forever grateful for your continued support.

Dr. Ann Myers, what a great experience it has been. I could have not asked for a better dissertation chair. Thank you for making me take ownership for my learning and for keeping me in the driver's seat even when I wanted to become the passenger or better yet ride in the backseat! It was because of the way you challenged my assumptions and made me analyze and test my theories that I have emerged a better person and professional. I have enjoyed each one of our conversations, which elevated not only my work, but me as a person, while enjoying a coffee and tapioca pudding.

Dr. Peter McManus, you were not only a great executive coach, but you became an inspiration and a friend to me. I am forever grateful for the first visit we had together, and I will never forget the advice you gave me. It was during our first meeting that you asked me what my goals were as both a doctoral student and as a building principal. When I began with the words "I try to" you quickly stopped me and gave me the best advice, which was "you don't try – you do. You either do and you succeed or you do and you fail but to say 'I try' is a cope out." I reflect upon that lesson often. I am thankful for all the times you emailed to see how I was doing, for the advice on how to navigate through the program, and of course, for the proofreading and feedback that you had given to me throughout the journey. Anyone would be lucky to have you as his or her mentor and coach.

Dr. Ray O'Connell, thank you for your patience and guidance. You are a true educator. The patience you showed while trying to get me to grasp some of the statistical analyses was amazing. I am going to miss our many Skype sessions and having you say "Shawn drag my image down to the left side of the screen so I can see the screen better." Thank you for your tutorials, flexibility, and dedication. I would have never made it to the end without your help and support.

I want to thank all of my cohort IV members. I have enjoyed getting to know each one of you. I could not have gone on this journey with a better group of people. I will always treasure our long weekends, weeks of residency, conference calls, emails, and discussions, both in and out of the classroom. We have shown solidarity in both good and trying times. I have learned something from each of you, which in turn has made me a better-rounded person.

The cohort model at Sage is truly inspirational in many ways, but the way it creates bonds and facilitates interconnectedness is priceless. The completion of this program would not have been possible without the guidance from prior cohort members: Dr. Mark Stratton, Dr. Steve Dana, Dr. Christopher Wojeski, and especially, Dr. Peter DeWitt. Peter, thank you for the many encouraging emails, the talks at Starbucks, your insight, but most of all for your friendship.

Liz Fisk, you are the best. You gave me invaluable advice from the very beginning stages of developing my survey to the use of SPSS.

Dr. Robin Young, you are an expert and have made my work look like a masterpiece through your edits and formatting. I truly appreciate your efforts, guidance, and help.

I would be remiss if I did not thank both Mrs. Barbara Newman and Ms. Colleen Yeager. Thank you so much for your support, encouragement, and check-ins to make sure I was on task and that my assignments where done on time. You guys are the BEST support staff anyone could ask for. Thank you for the many bags of Swedish Fish!!!

I also want to thank my friends for their understanding and support. Entering into a doctoral program requires a lot of sacrifices, one of which is socializing. You knew when to provide support and encouragement, but you also knew when to remind me that I needed to temporarily step away from my research and go enjoy a quick bite to eat with friends in order to become refreshed and regain my focus. Thank you to David, Pat, Randy, Jim, Brian, Ralph, and Steve, Karen, and Patti. Laughter is the best way to relieve stress.

Thank you to the dedicated teachers and staff at the Myers Center and to my colleagues at the WSWHE BOCES, especially Mr. Jim Dexter, District Superintendent, and Mrs. Nancy DeStefano, Director of Special and Alternative Education. Your support was insurmountable.

Thank you to the members of both my panel of experts and my defense panel. I appreciate the time you gave to me and treasure the feedback you provided.

Last, but not least, I want to thank my mother and sister for always being there, encouraging me to do my best, giving me a solid foundation to build upon, and sticking with me through the highs and the lows. They say experience shapes who we are, and the opportunities you have afforded me throughout life have enabled me to reach my goal of attaining my doctoral degree.

## Table of Contents

Abstract ..... i

Dedication ..... ii

Acknowledgements ..... iii

List of Tables ..... ix

List of Figures ..... xii

Chapter 1: Introduction ..... 1

The Research Problem ..... 1

Prior Studies ..... 4

Unresolved Issues From Prior Research ..... 5

Significance of Study ..... 6

Purpose Statement ..... 8

Research Questions ..... 8

Key Terms and Definitions ..... 9

Summary ..... 13

Chapter 2: Review of Literature ..... 15

Educational Reform ..... 15

Education and Society ..... 16

Education Finance ..... 17

Federal Role in Education Reform ..... 18

A Nation at Risk: The Imperative for Educational Reform ..... 20

No Child Left Behind ..... 22

Race To The Top ..... 24

Teacher Evaluations ..... 26
Identified Weaknesses Associated With Teacher Evaluations ..... 31
Evaluative criteria ..... 32
Few shared values and assumptions about good teaching ..... 35
Lack of precision in evaluations ..... 37
Lack of communication ..... 39
Limited administrator experience ..... 41
No differentiation between novice and experienced practitioners ..... 43
Principal Preparation Programs ..... 45
Adult Learning ..... 51
Summary ..... 56
Chapter 3: Methods ..... 60
Research Questions ..... 60
Research Design ..... 61
Target Population and Sample ..... 62
Instrumentation ..... 64
Data Validity and Reliability ..... 66
Data Collection ..... 67
Data Analysis and Interpretation ..... 71
Limitations and Delimitations ..... 72
Chapter 4: Results ..... 75
Demographic Profile of the Building Principal Respondents ..... 76
Research Question 1 ..... 83

Research Question 2 ....89
Research Question 3 ....93
Research Question 4 ....96
Summary ....100
Chapter 5: Summary of Findings, Conclusions, and Recommendations ....102
Summary of Findings ....104
Research question 1 ....104
Research question 2 ....107
Research question 3 ....109
Research question 4 ....110
Other findings ....114
Recommendations for Practice ....114
Recommendation 1 ....114
Recommendation 2 ....116
Recommendation 3 ....118
Recommendation 4 ....120
Recommendation 5 ....121
Recommendations for Future Research ....122
Recommendation 1 ....122
Recommendation 2 ....123
Recommendation 3 ....124
Recommendation 4 ....127
Recommendation 5 ....127

Closing Statement ..... 129

References ..... 131

Appendix A: Email Letter Description of Survey With Participant Consent ..... 143

Appendix B: Survey Instrument ..... 145

Appendix C: Survey Reminder Email to Principals ..... 155

List of Tables

Table 1: Demographic Data of the Three BOCES Regions Included in the Study .....63

Table 2: Free and Reduced Lunch Rates of the Three BOCES Regions Included in the Study ...64

Table 3: Student Demographics of the Three BOCES Regions Included in the Study .....64

Table 4: Sample Calculations .....69

Table 5: Survey Response History .....69

Table 6: Frequency Distribution of Respondents' Gender .....77

Table 7: Frequency Distribution of Respondents' District Student Enrollment .....78

Table 8: Frequency Distribution of Respondents' Configuration of Grade Level Supervision ....78

Table 9: Frequency Distribution of Respondents' Experience as a Classroom Teacher .....81

Table 10: Frequency Distribution of Respondents' Experience as an Assistant Principal .....82

Table 11: Frequency Distribution of Respondents' Years of Experience as an Assistant Principal for Those Who Reported Experience as an Assistant Principal .....83

Table 12: Frequency Distribution of Respondents' Perceptions of Their Readiness to Conduct Effective Evaluations of Teachers .....84

Table 13: Frequency Distribution of Respondents' Perceptions of Their Preparedness to Implement Specific Elements of Effective Evaluations of Teachers .....85

Table 14: Frequency Distribution of Respondents' Perceptions of the Effectiveness of Resources in Their Preparation to Conduct Teacher Evaluations .....87

Table 15: Frequency Distribution of Respondents' Perceptions Verbal or Written Feedback That They Received From a Supervisor Regarding the Teacher Evaluations That They Had Written .....88

Table 16: Frequency Distribution of Respondents' Reported Frequency of Teacher Evaluation Training Provided to Administrators in Their Current District....88
Table 17: Correlations for Reported Professional Development and Perceived Readiness to Conduct Effective Teacher Evaluations....90
Table 18: Frequency Distribution by Gender of Respondents' Perceptions of Their Preparedness to Conduct Teacher Evaluations....91
Table 19: Frequency Distribution by Years of Experience of Respondents' Perceptions of Their Preparedness to Conduct Teacher Evaluations....92
Table 20: Frequency Distribution of Respondents' Perceptions of Effective Resources by Years of Experience as a Building Principal....93
Table 21: Frequency Distribution by Participants' Race To The Top (RTTT) Implementation Stage of Respondents' Perceptions of Readiness to Conduct Effective Teacher Evaluations....94
Table 22: Frequency Distribution of Respondents' Reported Timeline for Mandated Implementation of the New APPR Regulations as Articulated by Race To The Top (RTTT)....95
Table 23: Frequency Distribution by Participants' Race To The Top (RTTT) Implementation Year of Respondents' Perceptions of Preparedness to Evaluate....96
Table 24: Frequency Distribution of Respondents' Perceptions of the Effectiveness of Resources in Their Development as Effective Teacher Evaluators....99
Table 25: Frequency Distribution of Respondents' Perceptions of Their Frequency of Use of Resources in Their Development as Effective Teacher Evaluators....100

Table 25: Top Four Resources Identified as Very Effective and the Frequency of Use by Respondents.....114

List of Figures

Figure 1. Developmental stages as defined by Petrie (2011) from the Center For Creative Leadership .....55

Figure 2. Scattergram of the frequency counts of respondents' number of years experience as a building principal.....79

Figure 3. Scattergram of the frequency counts of respondents' number of years experience as a building principal in their current school .....80

Figure 4. Scattergram of the frequency counts of respondents' years of experience as a classroom teacher for those who reported classroom teaching experience .....82

Figure 5. Possible district internal training schedule .....120

## Chapter 1: Introduction

> Change cannot happen unless those who are expected to change are given the opportunity to build the capacity required to make that change. Accountability must be a reciprocal process. When improved performance is expected, there is a responsibility to make investments in developing the skills and knowledge of those who are expected to improve. (Shakman, Breslow, Kochanek, Riordan, & Haferd, 2012, p. 8)

### The Research Problem

Evaluation is a reflection of human performance in organizational settings (Stronge, 1991). Evaluation is not defined under Race To The Top (RTTT), regulations which were designed to reform education, but instead describe the components of evaluation and reference what it can be used for (U.S. Department of Education [USDOE], 2009). Evaluation in the education field is now a central component of instructional improvement that provides rigorous performance measurement and useful feedback (Shakman et al., 2012). The goal of using evaluation as a reflection of one's performance remains the same, but the objectives or outcomes have changed over time. The use of evaluations has changed from a form of documentation to a tool to help teachers improve their practice and support personnel decision-making (New York State Education Department [NYSED], 2012). Evaluations should be designed to promote teacher effectiveness. "In fact, most authors identify the fundamental purposes of teacher evaluation as improving performance and documenting accountability" (Tucker & Stronge, 2005, p.6). In its guidance document the New York Education Department (NYSED) states,

> The purpose of the evaluation system is to ensure that there is an effective teacher in every classroom and an effective leader in every school. The evaluation system will also

> foster a culture of continuous professional growth for educators to grow and improve their instructional practices. (NYSED, 2012, p. 6)

Due to more accountability and the change in objectives, the structure and process for conducting evaluations are beginning to change (Danielson & McGreal, 2000; Marshall, 2009; Marzano, Frontier, & Livingston, 2011; Tucker & Stronge, 2005; USDOE, 2009). "Better defined goals have shifted the educational dialogue from vague opinions about student progress to factual evidence of student performance" (Peterson & Peterson, 2006, p. 91).

This research approaches the issue of effective teacher evaluation from a different perspective than other studies that looked primarily at the process and barriers, in that this study examined what resources and ongoing training building administrators report they need to become effective evaluators of teachers. Krein (1990) writes, "I propose that the problem is not an inability to determine what the components of a good program should be, but the failure to implement them properly" (p. 1).

Prior use of teacher evaluation was primarily to document whether a teacher exhibited the minimum competencies to receive tenure or, on the opposite end of the continuum, it was used as a document to assist in the removal of a teacher (Sawyer, 2001; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985). Teacher evaluation was not seen as a high stakes activity, but as an annual, routine procedure. Historically, school districts were allowed to negotiate their own evaluation process, which often relied heavily on one or two formal classroom observations. Prior research shows that many teachers did not feel that they received enough feedback from their evaluations to tell them how they were doing or to help them grow professionally (Gimbel, Lopes, & Nolan Greer, 2011; Howard & McColskey, 2001; Kimball & Milanowski, 2009; Marshall, 2005; McGrath, 2000; Peterson & Peterson, 2006; Wise et al., 1985).

Legislation aimed at educational reform, such as No Child Left Behind (NCLB) and RTTT, has changed the objective of teacher evaluation (Tucker & Stronge, 2005). "Teacher evaluation is a major component of the educational agenda today" (Tucker & Stronge, 2005, p. 25). The objective of the new evaluation process is to increase student achievement by helping teachers improve their practice and support personnel decision-making (Marshall, 2009; Marzano, Frontier, & Livingston, 2011; Peterson & Peterson, 2006; Shakman et al., 2012; Tucker & Stronge, 2005). RTTT federal legislation has granted funding to states that develop a standard, objective teacher evaluation system. Shakman et al. (2012) define teacher evaluation as a "rigorous system that includes frequent observations, with validated protocols, evidence of teacher practice, and student outcomes, and measures of student learning" (p. 3).

RTTT makes instructional staff and leaders accountable for student outcomes and aims to improve teacher and principal effectiveness based on their performance (USDOE, 2009). As a result of this change, teacher evaluation has transitioned from focusing on teaching to focusing on student learning. Johnson (1997) writes, "by providing the public, administrators, and teachers with data on teachers' skills and performance, a teacher-evaluation system will increase the instructional productivity of teachers, enhance student learning, and ultimately improve the quality of the educational system" (p. 70).

This change requires that building principals have a new set of skills that will lead to teacher self-reflection, growth and improved student achievement. This research examined principals' perceptions of their readiness to conduct effective teacher evaluation and the ongoing professional development support they feel they need in order to develop the skill set necessary to become effective evaluators. Ongoing support and training will be imperative to ensure that principals do not revert to conducting evaluations based on judgment and subjectivity that many

found ineffective and meaningless (Mitgang, 2012; Painter, 2000). "One-day workshops or weeklong trainings in isolation rarely provide the necessary conditions to sustainably influence practice" (Shakman et al., 2012, p. 8).

When looking at principals' perceptions of their readiness to conduct effective teacher evaluation and the ongoing professional development support they feel they need to develop the skill set necessary to become effective evaluators, this study was designed to see if there was a difference between principals at different building levels, specifically those who were required to conduct evaluations under RTTT legislation and those who have not yet been required to do so. In its guidance document NYSED (2012) states,

> For the 2011-2012 school year and thereafter, for teachers and principals in subjects and grades where there is no "value-added" model approved by the Board of Regents for such subject and grade: 20% on student growth on State assessments or comparable measures, and 20% on other locally-selected measures that are rigorous and comparable across classrooms in accordance with standards prescribed by the Commissioner. (p. 6)

Students in grades 3-8 are required to sit for New York State (NYS) assessments in English and Language Arts (ELA), math, and science annually as a way to measure student progress, thus making the value added model not pertinent for those grade levels, and therefore the teachers in those grade levels were to be evaluated under the new guidelines beginning in the 2011-2012 school year. Under the law and regulations, all other grade levels (K-2, 9-12) have to comply with the new evaluation system in the 2012-2013 school year (NYSED, 2012).

**Prior Studies**

Prior studies show that many principals are not adequately prepared to be effective evaluators. This is supported by research that indicates that most evaluations done by

administrators are not accurate reflections of teachers' effectiveness (Halverson, Kelley, & Kimball, 2004; Jacob & Lefgren, 2006; Marzano, Frontier, & Livingston, 2011; Medley & Coker, 1987; Peterson, 2004; Poston & Manatt, 1993). Many of the studies conducted previously show that evaluations done by administrators are often based on perception and are often influenced by factors such as the halo effect or leniency (Haefele, 1993; Hartzell, 1995; Hedge & Kavanagh, 1988; Huffcutt & Woehr, 1994; Wise et al., 1985). There are also other barriers that affect the evaluation process and its outcome, including deficiencies in the instrument tool and in the organization itself (Stronge, 1991). It has often been noted that evaluations are based on a glimpse of the teacher and, therefore, cannot accurately capture the teacher's effectiveness (Howard & McColskey, 2001; Marshall, 2005). "Indeed, much of the published literature on the principal's role in evaluation consists of explanations of procedural and legal aspects of evaluation and calls for better training of principals" (Painter, 2000, p. 2).

**Unresolved Issues From Prior Research**

There have been a number of studies conducted that have looked at the teacher evaluation process. Many of these studies have focused on the process and have identified the most common mistakes that evaluators make (Freeberg, 1969, Gimbel, Lopes, & Greer, 2011; Hartzell, 1995; Johnson, 1997; Kersten & Israel, 2005; Ostrovar Namaghi, 2010; Natriello & Dornbusch, 1980). Previous studies have also concluded that the process is ineffective if it is a tool to help guide personnel decisions and a teacher's professional growth (Halverson et al., 2004; Wise et al., 1985). Few studies have inquired about the professional development, or identified the supports, needed by principals in order for them to become effective evaluators of teachers.

**Significance of Study**

This study expands upon previous studies conducted on the topic of teacher evaluations. This study is significant in that it provides valuable information for state and local educational leaders, as they work to implement a new teacher evaluation system. Mitgang (2012) highlights the importance of training and support for principals, after they are hired, that is tailored to both individual and district needs. This study was undertaken because there are few studies focused on the specific professional development supports needed by principals to strengthen their skills as evaluators. Ongoing support will be needed by principals in order to develop the skills required by the new system. The research over the past 30 years is consistent in reaching the conclusion that effective evaluations should include student data, multiple measures, and evaluator training. Wise et al., (1985) write, "Thus, behavior change requires transformation of belief structures and knowledge in a manner that allows for situation-specific applications" (p. 69).

This study adds to the existing knowledge base of teacher evaluation by providing important insights in the area of supports needed by principals in order to conduct valid and reliable teacher evaluations. This study was designed to provide additional knowledge about ongoing supports needed by principals. It will bring a deeper understanding of what strategies and supports principals feel are effective and beneficial to them as they enhance their skills as evaluators of teachers. In addition to advancing the field of teacher evaluation research, the results could help districts focus on developing internal training that includes practical applications that may increase the validity and reliability of evaluations conducted by principals. Kimball and Milanowski (2009) found that principals who conducted more valid evaluations reported that district training helped build their evaluation skills.

Lack of monitoring the performance of principals as evaluators has contributed to the ineffectiveness of the current system. Poston and Manatt (1993) write, "principals have been poorly monitored in terms of their supervisory skill in conducting evaluations of teachers" (p. 43). They further state,

> However, the effectiveness with which a school organization achieves success in improving teaching with evaluation depends, at least in part, upon the degree to which the process is credible and sufficient to provide usable information. Credibility and sufficiency in turn depend upon administrative skill and competence (Poston & Manatt, 1993, p. 42).

This research has the potential to guide practices and shape regulations regarding the evaluation process of teachers. It is through the findings of this study that practices can be implemented to ensure that principals have the skills and knowledge necessary to be effective evaluators of teachers. Ongoing support will facilitate the growth and development of principals as the new evaluation system changes from a ritualistic paper procedure to a tool for continuous professional growth.

Ongoing supports and training are crucial for a transformational change in the way teachers are evaluated because, as found in the study by Wise et al. (1985), " almost all respondents, even those who believed that principals supported the teacher evaluation program, felt that principals lacked sufficient resolve and competence to evaluate accurately" (p.75). Ongoing supports will facilitate the growth of principals, which will better prepare them to conduct fair and objective teacher evaluations that lead to recognition, encouragement, and improvement in teaching practices. More specifically, the skills acquired by principals, under the new evaluation system will help them facilitate teacher self-reflection, growth, and improved

student achievement. This study captures the perceptions of a select group of building principals as they reflect on what they need and where they obtained training to become effective evaluators of teachers.

**Purpose Statement**

A quantitative design was used to generate and analyze survey data in regards to principals' perceptions of their readiness to conduct effective teacher evaluations and the ongoing professional development support that they need to become effective evaluators that leads to teacher self-reflection, growth, and improved student achievement. More specifically, this research can be used to help identify the ongoing evaluator supports needed to strengthen the skills of a principal who is conducting the teacher evaluation process and help system leaders develop internal systems that will support their building principals as they enhance their skills as evaluators.

This research was based on the responses of building administrators who worked within the following geographic counties at the time of the study: Albany, Columbia, Essex, Hamilton, Rensselaer, Saratoga, Schenectady, Schoharie, Warren, and Washington, comprising the Greater Capital Region of NYS. While their responses to the questions posed may serve all administrators in NYS who must adhere to the new RTTT requirements, it is important to understand that the data came from this convenience sample.

It was through this study that the question of what additional professional development principals report that they need to become effective evaluators of teachers was explored.

**Research Questions**

This study was designed to investigate the following research questions:

1. In what ways do principals report that they have been prepared to be an effective evaluator?
2. What is the relationship between reported professional development and perceived readiness to conduct effective teacher evaluations?
3. Is there a difference in perceived readiness to conduct effective teacher evaluations between principals at different building levels, specifically those who are now required to conduct evaluations under Race To The Top legislation versus those who have not yet been expected to do so?
4. What additional professional development do principals report that they need to become effective evaluators of teachers?

**Key Terms and Definitions**

The list below consists of terms and definitions that are used in this study.

*A Nation At Risk:* A report to the Nation and the Secretary of Education by The National Commission on Excellence in Education (1983). Describing the erosion of American educational foundations as a "rising tide of mediocrity," the Nation at Risk report is often credited with jump-starting the current wave of education reform (National Commission on Excellence in Education [NCEE], 1983, p. 5). The report documented that primary and secondary education in the United States (US) lagged far behind the levels of achievement in other industrialized nations and predicted problems in international competitiveness if our educational system did not improve. The data indicating poor performance of American students were subjected to intense and repeated scrutiny. While legitimate variations in interpretation were possible, an overall troubling picture of achievement provoked widespread concern with K-12 education (Miller, 2000).

*BOCES:* Boards of Cooperative Educational Services (BOCES) in NYS. BOCES is comprised of districts from the same geographic area and exist to provide shared services to those districts. School districts statewide depend on BOCES to meet their educational and financial goals. The BOCES model provides accountability, municipal sharing, efficiency, and equity (Boards of Cooperative Educational Services of New York State [BOCES NYS], 2012).

*Evaluation:* Evaluation is a reflection of human performance in organizational settings (Stronge, 1991). A central component of instructional improvement, that provides rigorous performance measurement and useful feedback. A tool used to help teachers improve their practice and support personnel decision-making (Shakman et al., 2012).

*Feedback:* The goal of feedback is to improve the effectiveness of teaching and promote professional growth. Feedback should be based on the following three criteria: based on descriptive observable data, provide characteristics of effective teaching, and promote reflective inquiry and self-directedness to foster improvements in teaching supported by evidence of student learning (Feeney, 2007, p. 191).

*Highly Effective Principal:* A principal whose students, overall and for each subgroup, achieve high rates (e.g., one and one-half grade levels in an academic year) of student growth. States, LEAs, or schools must include multiple measures, provided that principal effectiveness is evaluated, in significant part, by student growth (as defined in this notice). Supplemental measures may include, for example, high school graduation rates; college enrollment rates; evidence of providing supportive teaching and learning conditions, strong instructional leadership, and positive family and community

engagement; or evidence of attracting, developing, and retaining high numbers of effective teachers (USDOE, 2009).

*Highly Effective Teacher:* It is distinguished by judgment, intuition, insight, creativity, improvisations, and expressiveness (Wise et al., 1985, p. 108). RTTT legislation defines a highly effective teacher as a teacher whose students achieve high rates (e.g., one and one-half grade levels in an academic year) of student growth States, lead education agencies (LEA) or schools must include multiple measures, provided that teacher effectiveness is evaluated, in significant part, by student growth (USDOE, 2009). Supplemental measures may include, for example, multiple observation based assessments of teacher performance or evidence of leadership roles (which may include mentoring or leading professional learning communities) that increase the effectiveness of other teachers in the school or LEA. (p. 12).

*Highly Qualified:* professionally certificated and/or licensed. Other factors include having a bachelor's degree, passing a test in each of the subjects one teaches, and having an academic major or equivalent in each subject taught. Determining if a teacher has a graduate degree or advanced certification results in being *highly qualified* demonstrably (Brimley & Garfield, 2008, p. 209).

*Judgment:* Ability to reach logical conclusions and make high quality decisions based on available information; skill in identifying educational needs and setting priorities; ability to evaluate critically written communications (Witters-Churchhill, 1991, p. 340).

*No Child Left Behind (NCLB):* A reauthorization of the Elementary and Secondary Education Act. NCLB is federal legislation to ensure all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum proficiency on

challenging State academic standards and state academic assessments (No Child Left Behind [NCLB] Act, 2002).

*Observation:* Can be artificial in nature, suggests an inspection approach to supervision, has limited validity based on the skill of the observer, is narrow in scope, and involves a small sample of the teacher's actual work with students (Tucker & Stronge, 2005, p. 7).

*Principal:* The commissioner's regulations require that a full-time principal be employed and assigned to each school. The principal must hold appropriate certification (8NYCRR § 100.2(a)) (New York State School Boards Association [NYSSBA] & New York State Bar Association [NYSBA], 2010, p. 157). A school building leader (SBL) certificate is required for principals in any building level leadership position (8 NYCRR § 80-3.10(a). SBLs must first obtain an initial certificate that will be valid for five years from its effective date (8 NYCRR § 80-3.10 (a)(1)(i)). A professional certificate obtained after an initial certificate will be valid continuously so long as professional development requirements are met (8 NYCRR § 80-3.10(a)(2)(i). The requirements for a SBL initial certificate are a master's degree or higher from a regionally accredited higher education institution or equivalent as determined by the State Education Department (SED); a satisfactory score on the NYS assessment for school building leadership (8 NYCRR § 80-3.10 (a)(1)(ii)(b); and three years of classroom teaching and/or pupil personnel service experience in nursery school through 12th grade (8 NYCRR § 80-3.10(a)(1)(ii)). The requirements for an SBL professional certificate are an SBL initial certificate; three years of school experience in an educational leadership position with at least one of those in a SBL position; and participation in a mentoring program as required in commissioner's regulations (8 NYCRR § 80-3.10(a)(2)(ii)).

*Race To The Top:* A 4.35 billion dollar federal competitive grant program designed to encourage states to and reward states for creating conditions for education innovation and reform (USDOE, 2009).

*Reliability:* Reliability in evaluation refers to the consistency of measurements across evaluators and observations (Wise et al., 1985, p.85).

*Skill:* The ability of leaders to engage in the intended practice (Halverson et al., 2004, p. 8).

*Validity:* The validity of a teacher evaluation process depends on its accuracy and comprehensiveness in assessing teaching quality as defined by the agreed-on criteria (Wise et al., 1985, p.92).

**Summary**

In this chapter, an overview of the literature pertaining to the ongoing supports needed by principals to become effective evaluators of teachers was presented. It was noted through an extensive search of the literature that there is limited research identifying the ongoing supports that principals feel are effective and beneficial in helping them acquire the skills needed to be fair and objective evaluators of teachers. In addition to identifying the problem to be investigated, the purpose statement and the four research questions that guided the study were introduced. The chapter concluded by identifying the significance of the study, followed by a list of definitions used throughout the study.

The remaining four chapters of this study give specifics by going deeper into principal readiness and professional development to conduct effective teacher evaluations that will lead to improved student achievement. Chapter two reviews the literature on teacher evaluation, including the history, purpose, shortcomings, and principal preparation. It is through the research that the need for professional development and ongoing supports for principals is

supported to enable them to conduct valid and reliable evaluations that lead to teacher self-reflection, growth, and increased student achievement. This study builds upon previous research by identifying specific resources and supports that building principals perceive as effective in strengthening their skills to become effective evaluators of teachers. The third chapter describes the selection and rationale for the methods used in this quantitative study to answer the four guiding research questions. In this chapter, the population, instrumentation, and data collection and analysis processes are described. The fourth chapter presents a discussion of the aggregated results from the study. The report concludes with Chapter Five, which includes a summary of the findings as well as conclusions and questions raised as a result of this study with recommendations for both system level leaders and for future research on the topic.

## Chapter 2: Literature Review

Chapter II reviews the literature that examines the history of educational reform with a focus on teacher evaluation. Teacher evaluation is transforming from a rudimentary process of documentation to that of an effective tool to build capacity in teachers to ensure that there is an effective teacher in every classroom. Principal preparation programs and adult learning are also reviewed, as administrators will need to develop a new skill set in order to be effective evaluators of teachers. Research shows that one of the major contributing factors to student achievement is effective school leadership, which is second to having highly effective teachers in the classroom (Clifford, Behrstock-Sherratt, & Fetters, 2012; Dyer & Renn, 2010).

### Educational Reform

For about thirty years, studies have documented the decline in education, and researchers have identified specific areas in need of improvement with teacher evaluation being one of them. The primary means for increasing student performance is to have distinguished or highly effective teachers in the classroom (Oluwole, 2009; Shakman et al., 2012). One method to inform and increase teacher performance is through the effective use of the evaluation process. Evaluations can only improve the performance of classroom teachers if they are valid, reliable, and provide meaningful feedback. A review of the literature shows that researchers have been suggesting the same reforms and practices, aimed at increasing student achievement, consistently over the past three decades. The federal government has made various attempts at educational reform over the years by taking an advisory role, but with little impact. As schools in the US fail to produce students who are college and career ready, educational reform has become a focus for many states and the federal government. Reforming the evaluation system as we know it and helping principals enhance the skills necessary to be effective evaluators of teachers will be

crucial. Waters and Marzano (2006) found in a meta-analysis research project they conducted that when building and district leadership "effectively address specific responsibilities, they have a profound, positive impact on student achievement in their districts" (p. 6).

**Education and Society**

Research shows that it was declared through a 1983 report published by the United States Department of Education (USDOE) that the Nation was falling behind globally, and that our country is not as competitive as it once was (NCEE, 1983). The decline in education has had a negative effect on society both socially and economically. "…the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people" (NCEE, 1983, p. 13). It was not until the passing of NCLB in 2001 and RTTT in 2009 that education began to rapidly change.

Education benefits society by preparing students to be productive contributing members. Education is the core that supports this nation's economic growth, national security, health, and ensures a free democratic society. "Too few people realize the contribution that formal education has made to the social, political, and economic achievement of the United States" (Brimley & Garfield, 2008, p. 34). Data shows that the more educated a person is the less likely they will have to rely on publicly funded services. "President Reagan noted the central importance of education in American life when he said: 'Certainly there are few areas of American life as important to our society, to our people, and to our families as our schools and colleges'" (NCEE, 1983, p. 14). Education not only benefits the individual through the development of skills, but also society in general. Brimley and Garfield (2008) write, "not only does the individual benefit from an investment in an organization, but society as a whole benefits when goods and services are produced for all" (p. 26).

**Education Finance**

> The State FY 2012 Enacted Budget provides $19.64 billion in School Aid for the 2011-12 school year, which includes: $14.9 billion in Foundation Aid; $7.0 billion in other formula-based aid programs (e.g., Building Aid, Transportation Aid, Universal Prekindergarten); $281 million in categorical programs; and a reduction for the $2.55 billion formula-based Gap Elimination Adjustment" (NYS DOB, 2011, p. 200).

With education funding projected to reach $20.45 billion in the 2012-2013 school year, education is one of the biggest investments the community makes collectively in NYS. "It is well established that human capital is more important than natural resources in wealth creation. Fortunate indeed is a nation that has extensive natural resources; however, the nation with highly developed human resources is even more fortunate" (Brimley & Garfield, 2008, p. 19). As the country tries to recover from the current economic downfall, education is recognized as a way for America to recover (Huntley, 2011).

A financially overburdened society is beginning to question the effectiveness of its current publicly funded educational system and is skeptical of the return on its investment. This was highlighted by Governor Cuomo, who stated during his 2011 State of the State address: "not only do we spend too much, but we get too little in return. We spend more money on education than any state in the nation and we are number 34 in terms of results" (Cuomo, 2011, para. 13). Public support for education is beginning to wane, and people want better education through an improved system (Brimley & Garfield, 2008). "In addition the recession is causing angst and uncertainty among taxpayers who are acutely aware of the dire economy" (Huntley, 2011, p. 4). This has created a sense of urgency with the public and our elected representatives. Kotter and Cohen (2002) write, "in successful change efforts, the first step is making sure sufficient people

act with sufficient urgency- with on-your-toes behavior that looks for opportunities and problems, that energizes colleagues, that beams a sense of 'let's go'" (p. 15). High property taxes, declining graduation rates, a widening achievement gap, an increase in the number of remedial courses that college freshman require, and a lack of career readiness in high school graduates has created that sense of urgency to which Kotter and Cohen (2002) describe. These factors have caused political leaders to question the value of our educational system and created a sense of necessity where people felt that something needed to be done to change the educational system, as we know it. "Most schools, however will receive far less money as the state implements major curriculum changes and overhauls the way students and teachers are evaluated" (Cooper, 2010, para. 4).

**Federal Role in Education Reform**

"Government institutions, which include public schools, do not react as quickly or as obediently to consumer demand, external pressure, and public criticism as their counterparts in the competitive world!" (Brimley & Garfield, 2008, p. 7). Despite the fact that education is the responsibility of the state governments, according to the Constitution, the federal government felt that intervention was necessary due to the slow change efforts at the state and local levels. As a result of our faltering educational system the federal government, at one time a minor player in the educational system, has begun to directly influence and regulate education to bring about the desired change. The federal government has abandoned its advisory role and is becoming more pervasively active than at any other point in our history. "Provoked by concerns of rising costs, inadequate returns, and dwindling resources, policy makers have turned their attention to initiatives that would not only increase the effectiveness of public education but would increase the responsiveness of the system to the public it serves as well" (Johnson, 1997, p. 69).

The most widely recognized example of the advisory role the federal government played was through the publishing of the 1983 report, *A Nation at Risk: The Imperative for Educational Reform* (NCEE, 1983). More recently, the role of the federal government has shifted from advisory to regulation and mandate. NCLB in 2001 and RTTT in 2009, which hold both states and local school districts accountable for their outcomes, are the latest legislative reforms to be adopted. "A great deal of attention has been given to the profession with the passage of the No Child Left Behind law, resulting in vigorous discussions centered on 'failing schools,' 'accountability through testing,' and the need for 'highly qualified teachers'" (Brimley & Garfield, 2008, p. 367). It is through RTTT legislation that teacher evaluation has become a focus in the quest for educational reform by using it as a tool to inform and increase teacher performance. Evaluations can only improve the performance of classroom teachers if they are valid, reliable, and provide meaningful feedback. In order for this to happen administrators themselves need to enhance their skills to become effective evaluators of teachers.

> Providing evaluators with relatively detailed rubrics or rating scales describing generic teaching behaviors thought to promote student learning, coupled with initial training in applying them, is not enough to ensure that all evaluators' ratings will be positively related to student achievement. If policy makers and program designers want evaluation scores to be more highly related to some criterion such as student achievement, it will take more than specific rubrics and basic training of evaluators in the process to achieve a strong relationship. (Kimball & Milanowski, 2009, p. 65)

Since *A Nation At Risk* was published increasing teacher performance has been a priority (NCEE, 1983). It is well documented that the primary means for increasing student performance, is to have distinguished or highly effective teachers in the classroom (Danielson,

2001; Darling-Hammond, 2000; Oluwole, 2009; Shakman et al., 2012; Smith & Gorard, 2007). This section includes a brief overview of the literature and reform efforts taken that has made teacher performance and effective teacher evaluations a priority in changing the educational system, as they are currently known, in an effort to increase student achievement.

**A Nation At Risk: The Imperative for Educational Reform.** In 1981, the National Commission on Excellence in Education (NCEE) was created to investigate the quality of education in the US. "The Commission was created as a result of the Secretary's concern about 'the widespread public perception that something is seriously remiss in our educational system'" (NCEE, 1983, p. 9). In 1983, the Commission released a report titled, *A Nation At Risk: The Imperative for Educational Reform*. The report was advisory in that it defined the problems plaguing our educational system. Although the report did not contain legislation, many believe that it was the driving force that has launched the wave of recent changes we are currently experiencing in education. The report further asserted that the mediocrity found in schools was cause for the nation's decline in commerce, industry, technology, and innovation. In the cover letter of the report, Gardner, Chair of the Commission, wrote "The Commission deeply believes that the problems we have discerned in American education can be both understood and corrected if the people of our country, together with those who have public responsibility in the matter, care enough and are courageous enough to do what is required" (NCEE, 1983, p.6). The findings in the report struck a cord with many, which has put education on the national forefront and has set the stage for the federal government to intervene through legislative action to bring about vital change.

"Our Nation is at risk" is the opening line of the Commission's report (NCEE, 1983, p. 8). The report goes on to say, "if an unfriendly foreign power had attempted to impose on

America the mediocre educational performance that exists today, we might well have viewed it as an act of war" (NCEE, 1983, p. 8). The commission's report emphasizes the importance for schools to maintain high standards and challenge students so that they reach their full potential (NCEE, 1983). The commission also noted that the educational system had to offer both high quality and equitable opportunities for all students.

> The twin goals of equity and high-quality schooling have profound and practical meaning for our economy and society, and we cannot permit one to yield to the other either in principal or in practice. To do so would deny young people their chance to learn and live according to their aspirations and abilities. It also would lead to a generalized accommodation to mediocrity in our society on the one hand or the creation of an undemocratic elitism on the other. (NCEE, 1983, p. 13)

The report's findings were focused on four areas within the educational process: content, expectations, time, and teaching. The report identified, among many findings, that teachers were inadequately prepared through teacher preparation programs and that many were unqualified to deliver instruction. The reports findings on teaching included the following:

> (1) Too many teachers are being drawn from the bottom quarter of graduating high school and college students; and (2) Half of the newly employed mathematics, science, and English teachers are not qualified to teach these subjects; fewer than one-third of U.S. high schools offer physics taught by qualified teachers. (NCEE, 1983, p. 20)

*A Nation At Risk: The Imperative for Educational Reform* was the start of efforts in school reform after stressing that our public schools were not producing the product that was needed to compete in a world economic setting (Brimley & Garfield, 2008). The need to reform teaching, teacher preparation programs and educational standards were some of the suggestions

made in the commission's report. It is through this report that the importance of having highly qualified teachers in the classroom to increase student achievement began. Since the release of this report, the importance of teacher growth and development as well as the significance of increasing instructional productivity have been identified as two of the primary factors associated with increasing student achievement.

**No Child Left Behind.** NCLB addressed declining student achievement but also "added an assertion that there was a rift in the public school system" (Brimley & Garfield, 2008, p. 394). Rothstein (2004) reports that experts have been aware of an achievement gap, created by social and economic disadvantages, for about 50 years. Given the growing body of knowledge that the primary means for increasing student performance is to have distinguished or highly effective teachers in the classroom legislation, such as NCLB is beginning to acknowledge the importance of teachers (Oluwole, 2009; Rothstein, 2004; Tucker & Stronge, 2005). The NCLB legislation of 2001 was created as a way to increase student performance through accountability, choice, and flexibility in federally funded educational programs (USDOE, 2002). The federal government abandoned its advisory role when it exercised mandates, on an unprecedented scale, with the passing of NCLB (Brimley & Garfield, 2008; Smith & Gorard, 2007).

NCLB required schools to test students annually, using standards-based tests created by each state's department of education, as a way to measure student progress (Catano & Stronge, 2006; Rothstein, 2004; Smith & Gorard, 2007; Tucker & Stronge, 2005). The aggregated results of the annual testing are reported on each district's report card, and parents receive reports consisting of their child's individual scores. "One outcome of the No Child Left Behind Act of 2001 was greater documentation of the extent to which poor and minority students are systematically taught by less highly qualified and less experienced teachers" (Clifford,

Behrstock-Sherratt, & Fetters, 2012, p. 10). NCLB offers equity through its goal of ensuring that all students, including those who have not traditionally done well in school, are taught by highly qualified teachers and have the ability to achieve minimum competency levels in core content areas (Smith & Gorard, 2007). The NCLB legislation required, "Assessment results and State progress objectives must be broken out by poverty, race, ethnicity, disability, and limited English proficiency to ensure that no group is left behind" (USDOE, 2002, p. 1). The progress of students is measured against grade level proficiency according to the state learning standards. The goal of NCLB is to have every school wipe out the achievement gap by race and social class (Rothstein, 2004).

The students' results on the state tests are also used as a way to measure the performance of a school against the proficiency goals set by the state. Brimley and Garfield (2008) write, "The centerpiece of the legislation is to identify schools that are failing to meet student achievement goals and to label them as schools 'in need of improvement'" (p. 209). If a school fails to meet its yearly adequate progress (AYP), it can fall in one of several categories, depending on the number of years it has been identified: improvement, corrective action, or restructuring (USDOE, 2002). "The No Child Left Behind Act includes options for the state to take over schools that fail to meet adequate yearly progress" (Brimley & Garfield, 2008, p. 153).

In its efforts to increase student achievement and close the achievement gap, NCLB also focused on putting qualified teachers in the classroom (Smith & Gorard, 2007; Tucker & Strong, 2005). The highly qualified teacher provision in NCLB "was fueled by 'report after report' of the dire state of education in the nation's public schools, including the fact that 25 percent of teachers were 'not qualified to teach in their subject area'" (Oluwole, 2009, p. 158). This holds true for many teachers, especially at the elementary and middle level, where a majority of their

coursework was focused on methodology and pedagogy rather than content (Smith & Gorard, 2007; Tanner, 2008). It was NCLB legislation that required all teachers who taught core content subjects to be highly qualified (Brimley & Garfield, 2008; Oluwole, 2009; Smith & Girard, 2007; Tucker and Stronge, 2005). The legislation required the use of,

> …practices grounded in scientifically based research to prepare, train, and recruit high-quality teachers. The new program gives States and LEAs flexibility to select the strategies that best meet their particular needs for improved teaching that will help them raise student achievement in the core academic subjects. In return for this flexibility, LEAs are required to demonstrate annual progress in ensuring that all teachers teaching in core academic subjects within the State are highly qualified. (USDOE, 2002, p. 3)

Marzano, Frontier, & Livingston (2011) report that despite the requirements in the NCLB legislation requiring highly qualified teachers, only 14 states require their school systems to do annual evaluations of their teachers.

The federal government has taken more of an active role in education by enacting mandates through legislation. It was the legislation of NCLB that propelled education reform by holding schools accountable, requiring research based practices and data to drive instruction, along with increasing standards for teacher qualification. "The requirements of NCLB, some maintain, have brought considerable federal control to education in the twenty-first century" (Brimley & Garfield, 2008, p. 209).

**Race To The Top.** The most recent legislation aimed at increasing student achievement is the RTTT legislation (2009). RTTT was part of President Obama's American Recovery and Investment Act, which he signed into law in 2009. RTTT was designed to improve education by encouraging sustained change in the educational system in order to increase student

achievement, close the achievement gap, increase graduation rates, and make students college and career ready as well as globally competitive.

The act is structured in four areas to bring about the desired changes. One of these focus areas is "recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most" (USDOE, 2009, p. 2). RTTT makes instructional staff and leaders accountable for student outcomes and aims to improve teacher and principal effectiveness based on their performance.

> In order to compete for the incentives promised in RTTT, states were required to include provisions that evaluations would differentiate effectiveness using multiple rating categories, that a significant portion of educator evaluation ratings would be based on student assessment data, and that all teachers would receive annual evaluations. (Shakman et al., 2012, p. 4)

RTTT transitions teacher evaluation from one based primarily on judgment to evidence based evaluation. The focus of education will shift from instructional process to student learning by using teacher evaluation as a tool for professional development, growth, and accountability. The reform in evaluation will also cause change by putting teachers in "more active and professional roles" in the evaluation process (Danielson, 2001, p. 14). Peterson & Peterson (2006) write, "principals can foster good teacher evaluation by knowing the developments of the past 25 years, taking initiative to support good data gathering for their own teachers and school, and supporting teachers as they become more involved in their own evaluations" (p. 13).

Research shows that the best way to increase student achievement is to have highly effective teachers in the classroom (Clifford, Behrstock-Sherratt, & Fetters, 2012; Danielson, 2001; Poston & Manatt, 1993; Tucker & Stronge, 2005; Wise et al., 1985). There have been

many studies conducted by leading experts, such as Danielson and McGreal (2000), Marzano (2007), and Reeves (2004), identifying what skills are essential to be a highly effective educator. A highly effective teacher, as defined in the *RTTT executive summary,* is

> a teacher whose students achieve high rates (e.g., one and one-half grade levels in an academic year) of student growth. States, LEA's, or schools must include multiple measures, provided that teacher effectiveness is evaluated, in significant part, by student growth. Supplemental measures may include, for example, multiple observation based assessments of teacher performance or evidence of leadership roles (which may include mentoring or leading professional learning communities) that increase the effectiveness of other teachers in the school or LEA. (USDOE, 2009, p. 12)

RTTT requires that every administrator, who will be responsible for conducting teacher evaluations, become certified by attending required trainings. RTTT allows each board of education to decide what the certification process will consist of for the administrators in their district. Although administrators attend these required trainings, in some cases there is no summative evaluation to determine if they have mastered the goals and objectives of the training, which were designed to make principals effective evaluators.

**Teacher Evaluations**

Evaluations, as defined by Stronge (1991), are a reflection of human performance in organizational settings. NYS Commissioners Regulations mandate teacher evaluations (NYSSBA & NYSBA, 2010). "The Commissioner's regulations require that each school board and board of cooperative educational service (BOCES) provide for annual professional performance reviews (APPR) of teachers who provide instructional services or 'pupil personnel services'" (NYSSBA & NYSBA, p. 230). Along with the legislation that mandated teacher

evaluations came the influence from the various unions that have impacted the evaluation process and created a flawed system by limiting the scope and substance of the teacher evaluation. Kersten and Israel (2005) found that, "principals perceive unions as not trusting the more complex, subjective teacher evaluation methods that are currently considered best practices" (p. 62).

Many of the evaluation systems used today were developed in the 1970's based on the clinical supervision model and the work of Hunter and have not changed despite educational research, which has shaped new theories on how students learn and the teaching strategies often required (Danielson & McGreal, 2000; Kersten & Israel, 2005; Marzano, Frontier, & Livingston, 2011). Hunter's work was called mastery teaching and it focused on a seven-step framework for the development of a lesson. Hunter and Russell (1990) wrote, "students learn more when they are well taught, and teaching is the reason for the clinical practice of education" (p. 1). Hunter (1973) also focused on "teaching decisions and behaviors that make the difference between success and frustration for the universe of learners" (p. 62). Hunter and Russell (1990) believed that the two absolutes in teaching were decision-making and maintaining the dignity of each student. "Education is a relativistic, situational profession which requires a constant stream of decisions on the part of the teacher" (Hunter & Russell, 1990, p. 1). Most of the evaluation systems currently in place evaluate teachers based on how well their observed lesson aligns to the Hunter model and, therefore, focus on instructional delivery. Zerihun, Beishuizan, and Van Os (2011) report that most of the evaluation instruments used are influenced by the teacher center model where good teaching is defined by how content is presented to students.

Evaluation systems, as defined by researchers such as Danielson (2000), must align with the complex learning, problem solving, and application of knowledge that we expect from

students. Wise et al. (1985) write, "proper teacher evaluation can determine whether new teachers can teach, help all teachers to improve, and indicate when a teacher can or will no longer teach effectively" (p. 62). The former NYS Commissioner for Education Steiner (2009) wrote,

> We believe teaching well is a deeply complex professional activity, thus the evaluation of teachers must take place along multiple dimensions, but the ability of a teacher to raise the academic performance of her or his students is critical, and that ability – better supported by new models of professional development – must form part of the evaluation system. (para. 5)

Part of the educational reform agenda is to improve teacher effectiveness based on their performance. The RTTT agenda requires states to:

> (ii) Design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that (a) differentiate effectiveness using multiple rating categories that take into account data on student growth (as defined in this notice) as a significant factor, and (b) are designed and developed with teacher and principal involvement.
>
> (iii) Conduct annual evaluations of teachers and principals that include timely and constructive feedback; as part of such evaluations, provide teachers and principals with data on student growth for their students, classes, and schools and
>
> (iv) Use these evaluations, at a minimum, to inform decisions (regarding personnel decisions). (USDOE, 2009, p. 9)

Prior to the recent RTTT legislation mandating reform in the evaluation process, evaluations were often subjective focusing on teacher traits and the instructional process (Danielson & McGreal, 2000; Jacob & Lefgren, 2006; Poston & Manatt, 1993; Smith, 1986).

The written evaluation served as a tenure document that verified that a person met the minimum competencies in order to receive tenure. Many of the elements valued in teacher evaluation were focused on teacher traits such as voice, appearance, emotional stability, trustworthiness, warmth and enthusiasm (Danielson & McGreal, 2000). One of the factors compromising the accuracy of subjective performance evaluations is the value that principals place on teacher characteristics that are unrelated to increased student performance (Jacobs & Lefgren, 2006; Smith & Gorard, 2007). In the 1960s and 1970s, the evaluation process changed its focus to basic skills acquisition and what teachers did to improve the basic skills of their students.

As educational research has advanced and increased our knowledge of student learning so have the goals for student achievement (Danielson & McGreal, 2000). Teacher evaluations must also change if they are to be used as a tool to increase teacher performance and align to the research and skills that students must attain in order to be successful in an ever-changing world. "Research on teaching and learning has yielded much in recent years. Further, our knowledge of how to develop good-teacher evaluation instruments and procedures has likewise increased" (Johnson, 1997, p. 70). With educational reform being on the forefront, many legislatures have become aware of research findings and, as a result, have passed legislation increasing teacher evaluation requirements and mandating teacher evaluation training for certified administrators (Kersten & Israel, 2005).

> On March 14, 2012, the Assembly and Senate passed the revised teacher and principal evaluation law proposed by the Governor (S.6732/A. 9554). On March 27, 2012, the Governor signed the revised teacher and principal evaluation law as Chapter 21 of the Laws of 2012. At its March meeting, the Board of Regents adopted regulations to

implement Education Law 3012-c, as amended by Chapter 21 of the Laws of 2012 (S.6732/A.9554), effective April 4, 2012. (NYSED, 2012, p. 2)

RTTT is transforming teacher evaluations from one of documentation and compliance to that of self-reflection and a professional learning tool. RTTT is aligning to what previous researchers, such as Danielson and McGreal (2000) and Kersten and Isreal (2005), have recommended in order to strengthen teacher evaluation. Evaluation is not clearly defined through RTTT or other guidance documents, such as *Guidance on New York State's Annual Professional Performance Review for Teachers and Principals to Implement Education Law §3012-C and the Commissioner's Regulations* (NYSED, 2012). The RTTT Executive Summary and the APPR Guidance document simply describe the components of evaluation and they reference what it can be used for (NYSED, 2012; USDOE, 2009).

Educators have to stop thinking that observation is the same as evaluation. Classroom observations have been the traditional primary source of evidence used by evaluators in their decisions on teaching performance (Haefele, 1993; Kimball & Milanowski, 2009, Poston & Manatt, 1993). Observation is a method or technique that can be used to obtain information for diagnosis and development of instructional practice (MET Project, 2012). Whereas, evaluation is a process that uses various methods in order to obtain data to judge a teachers total teaching performance. Howard and McColskey (2001) wrote, "when administrators observe teachers, they gain valuable data and feedback, but they don't have enough information to judge a teacher's total teaching performance" (p. 49).

A great emphasis has been placed on research into effective evaluation, and NYS has adopted ten different teacher practice rubrics. "The literature now considers the use of pre-and post- observation conferencing, narratives, rubrics, and portfolios as best practice procedures

within teacher evaluation, with the result being school improvement" (Kersten & Israel, 2005, p. 47). Even though research supports effective teachers and the use of an evaluation system as a tool for professional development, administrators may not be adequately prepared to meet this responsibility (Kimball & Milanowski, 2009; Poston & Manatt, 1993; Wise, et al., 1985).

> Valid, reliable, and helpful evaluation requires evaluators who recognize good teaching (and its absence) and who know how to improve poor teaching when they find it. Evaluator competence is probably the most difficult element of the process. The best supported and most carefully constructed process will flounder if those responsible for implementation lack the necessary background, knowledge, and expertise. (Wise et al., 1985, p. 86)

**Identified Weaknesses Associated With Teacher Evaluations**

The current system of teacher evaluation is plagued with problems. "Despite a heavy reliance on performance ratings, it is generally acknowledged that they are too often contaminated by systematic errors (leniency, central tendency, halo, and contrast errors)" (Smith, 1986, p. 22). Researchers have identified six main areas of deficiency in current teacher evaluation systems: a lack of clearly defined criteria, inconsistent definition of effective teaching, lack of inter-rater reliability, lack of effective feedback, no differentiation between novice and experienced practitioners, and limited administrator experience (Danielson & McGreal, 2000; Poston & Manatt, 1993). The current system offers little value to either the principal or the teacher and therefore little to the student as well. Both principals and teachers view teacher evaluation as a nuisance, routine procedure that must get done annually. Over the years, the current flawed teacher evaluation system has led to, what Danielson and McGreal (2000) describe as a culture of passivity and protection. The current system is more of a documentation

process than one of professional inquiry. Another major shortcoming of the current evaluation system is identified by Huffcutt and Woehr (1994). They point out that the vast majority of performance measurement relies on subjective judgmental measures of performance.

Research studies have shown that rater training is an essential component to help decrease problems associated with subjective performance judgments (Huffcutt & Woehr, 1994). "The more skilled the evaluator, the more likely that she will give ratings that accurately reflects how the teacher actually performs on the dimensions defined by the evaluation system" (Kimball & Milanowski, 2009, p. 39). The dual goals of the reformed teacher evaluation system are to measure teacher effectiveness and to help them continuously perfect their skills. Stronge (1991) says that an evaluation should measure what one does not who they are. There is now an expectation that all teachers increase their expertise from year to year, which produces gains in student achievement from year to year with a powerful cumulative effect.

**Evaluative criteria.** One of the biggest defects in teacher evaluation is the lack of clarity in evaluative criteria. Many people being evaluated do not know what they are being evaluated on, and/or different evaluators may not use consistent criteria to form their judgments. This was supported in a review of the literature over nearly thirty years, when Wise, et al., (1985) found that "narrative evaluation provided insufficient information about the standards and criteria against which teachers were evaluated and resulted in inconsistent ratings among schools – ratings that depended on the judgment of the building principal rather than uniform district objectives for teacher performance" (p. 71). Medley and Coker (1987) found that "educational personnel decisions are based on judgments, which according to the research, are only slightly more accurate than they would be if they were based on pure chance"(p. 243). Evaluators need

to have a clear understanding of the rating system being used and the criteria used to distinguish the different levels within the rating system.

> Teacher evaluation tools are designed to pass judgment on teacher performance. Judgment is uncomfortable for many teachers, and it is difficult to build the trust needed for true change in such a scenario. Yet, one of the nine essential skills that the University Consortium for the Performance-based Preparation of Principals of NASSP, identified as essential for principals was judgment. Many teachers tolerate the process as a necessary annoyance, but question the validity of the judgment passed by the administrator/evaluator. (O'Donovan, 2011, p. 73)

When looking at evaluation validity, Kimball and Milanowski (2009) caution that teachers could receive consequences that are not justified. They argue that "differences in validity across principals are clearly problematic as stakes are raised, and the decision making of those with less valid ratings needs to be improved" (Kimball & Milanowski, p. 36). Wise et al. (1985) reported that teacher satisfaction is strongly related to perceptions that all evaluators share the same criteria for evaluation and the teachers had input in the selection of the criteria used.

As research findings identify best practices and suggest changes in the delivery of instruction, so too should the evaluation tools and procedures that measure these new pedagogical practices be examined. In the 1970s teachers were evaluated based on teacher characteristics thought to be associated with effective classrooms.

> Until recently, teacher performance criteria were more often selected by committees of teachers from a mixture of theory, tradition, and mythology than from research on teaching. Furthermore, the use of a district committee to select criteria for an evaluation instrument results in a type of mass-authorship in which each teacher lobbies for

> inclusion of teacher-behaviors that they themselves do well. (Poston & Manatt, 1993, p. 45)

As research had advanced and defined effective practices many evaluation tools still measure the traits identified in the 1970s. Stronge (1991) writes, the "real concern should be for what a person can do rather than what s/he is. The same argument holds true today for the issue of effective evaluation; that is, performance evaluation systems in education should measure one's ability to perform effectively" (p. 77).

One of the primary ways for districts to increase the accuracy of teacher evaluations is to ensure that principals know the purpose of the evaluation is for professional growth or personnel decision-making. Both of these are valid uses of teacher evaluation but require different criteria to form a judgment. Poston and Manatt (1993) found that the types of teacher evaluation are often based upon the wrong purposes. Wise et al. (1985) state that teacher evaluation needs to be based upon clear criteria that in turn are directly related to the purpose of the evaluation.

> For purposes of accountability, teacher evaluation processes must be capable of yielding fairly objective, standardized, and externally defensible information about teacher performance. For improvement objectives, evaluation processes must yield rich, descriptive information that illuminates sources of difficulty as well as viable courses for change. To inform organizational decisions, teacher evaluation methods must be hierarchically administered and controlled to ensure credibility and uniformity. To assist decision-making about individuals, evaluation methods must consider the context in which individual performance occurs to ensure appropriateness and sufficiency of data. (Wise et al., 1985, pp. 68-69)

**Few shared values about good teaching.** Not only does a lack of clear criteria affect the reliability and validity of evaluations, but so does the evaluator's definition of good teaching. The research by Johnson (1997) shows that the definition of what effective teaching is changes according to the organizational role of the evaluator.

> Most if not all, teacher-evaluation policies are predicated on some working definition of effective teaching. Since the presence of competing conceptions among various roles creates the potential for organizational conflict and dysfunctional behavior, the identification and explication of these conceptions has utility for policy makers and practitioners. (Johnson, 1997, p. 81)

Johnson (1997) discovered that teachers had the most complex view of teaching and as one moved more toward central administration or state agencies the definition of effective teaching became more general and simplistic.

> A larger, more complex view of teaching emerges as one moves toward the organization's core. Given that individuals occupying roles away from the core often exert the greatest amount of influence in crafting teacher-evaluation policies, some concerns exist regarding the content of such policies. Overly simplistic and excessively general conceptions of teaching, as well as undue emphases on particular elements of teaching, are hallmark features of evaluation policies and systems that fail to account for the complexities of teaching. (Johnson, 1997, p. 82)

One must be careful when defining good teaching to allow for flexibility and differentiation.

> Teaching research has demonstrated that effective teaching behaviors vary for different grade levels, subject areas, types of students, and instructional goals. Thus, relative

> teacher competence cannot be assessed on the basis of highly specified, uniform criteria. When a school district adopts a single set of broad criteria, it must differentiate these criteria for specific applications. Excellence must above all be measured by broad, nonstandardized criteria. (Wise et al., 1985, p. 93)

A teacher can be observed for a short period of time and appear to be very effective in their instructional delivery but not necessarily have the same outcome if they delivered the same lesson in another section. A single short observation does not demonstrate that the teacher has the ability to modify her/his lesson based on student needs and learning styles.

As Danielson and McGreal (2000) point out, some evaluators look at teacher characteristics as evidence of good teaching while others evaluate good teaching based on the delivery of the lesson.

In order to make evaluations valid and reliable there must be a sense of trust and understanding. "A teacher evaluation system must define the teaching task and provide a mechanism for judging the teacher" (Wise et al., 1985, p. 65). There must be a clear understanding of what good teaching is, what the evidence of good teaching is, and how it will be measured. The new evaluation system, mandated through RTTT, incorporates multiple measures of effectiveness to determine a composite score that is then used to differentiate teacher effectiveness using four rating categories (USDOE, 2009). The law specifies that student achievement will comprise 40% of teacher evaluation, while the other 60% will be based on multiple measures of effectiveness consistent with standards prescribed by the Commissioner in Regulations that include the use of surveys, observations, and other options negotiated through collective bargaining agreements (NYSED, 2012).

**Lack of precision in evaluations.** The precision of teacher evaluations and how they align to teacher effectiveness and student achievement has been the focus of past studies (Marshall, 2005; Medley & Coker, 1987; Smith, 1986). Observation has been the primary method used in evaluating teachers, but it is only a snapshot, often 0.1% of the teacher's instruction (Marshall, 2005). Wise et al. (1985) question whether short infrequent observations can allow the principal to make an accurate evaluation that will reward superior teachers, encourage average ones, and improve or terminate the employment of poor ones. Smith (1986) writes, "raters often lack the observation skills necessary to attain observation accuracy" (p. 23). Therefore, it is necessary for teacher evaluations to become more of an objective process. It requires writing more than a narrative of what is observed. In short, the rating of a teacher is often times based on the bias of the evaluator and is not necessarily a true reflection of the teacher. Tucker and Stronge (2005) write, "Too often, personal opinions or biases contaminate the evaluation process and undermine the credibility and trust necessary for meaningful dialogue about instruction" (p. 10). Observations that are part of the evaluation process can be disadvantageous when there are no benchmarks or multiple sources of evidence and the rating is based primarily on the observation of the evaluator. Observation does not give a clear and unbiased reflection of the teacher as Wheatley and Kellnor-Rogers (1999) point out,

> What we see is most influenced by who we have decided to be. Our eyes do not simply pick up information from an outside world and relay it to our brains. Information relayed from the outside through the eye accounts for only 20 percent of what we use to create a perception. At least 80 percent of the information that the brain works with is information already in the brain." (p. 49)

When evaluators give feedback that is not anchored on clear criteria and multiple sources of evidence, it is usually contaminated with bias opinions that compromise the integrity of the process. Kimball and Milanowski (2009) explained that the evaluators used in their study, "may have relied on intuition or gut-level feelings about teachers, without even being completely aware of doing so" (p. 63). Krein (1990) points out that some evaluations are of little value because they are based on emotion and not specific facts. "Some experts call such comments GLOPs (generalized labeling of people). A GLOP is not feedback; it is a biased interpretation of the employee as a person – one that obstructs clear communication about the employee's specific behavior" (Krein, 1990, p. 2). Hartzell (1995) identifies the most common employee performance appraisal errors that affect the precision of evaluations: unwarranted strictness, unwarranted leniency, central tendency, halo effect, recency, contrast, and attribution. "In many districts teachers believe that the present system still depends too much on judgment or predisposition of the principal and leads to different ratings for similar teacher practices in different schools" (Wise et al., 1985, p. 75).

Frase and Streshly (1994) report "A plethora of studies and observations reveal evaluation ratings are grossly inflated" (p. 48). Some researchers such as Hartzell (1995) refer to this as leniency. Hartzell (1995) defines leniency as the tendency to rate a teachers performance higher than it warrants or where others would rank it. Many evaluators do not feel comfortable giving honest feedback to the teacher, as some feel it is not worth the hassle as nothing will come of it. Others do not want to create friction with the people they must work with; some evaluators feel the need to be liked, some believe it is a reflection of their leadership, and others are less candid in their evaluations as an attempt to help the teacher get a job in another school or district (Danielson & McGreal, 2000; Frase & Streshly, 1994; Kersten & Israel, 2005). When teachers

and administrators are polled many will admit that there are ineffective teachers (15 – 20%) in the classrooms (McGrath, 2000; Poston & Manatt, 1993), but Halverson et al. (2004) reported "Praise rather than critique, and high scores rather than low, characterized the written feedback provided by evaluators" (p. 14). Since teachers are often rated on a dichotomous scale and usually receive the highest ranking it has become the culture for teachers to expect it (Danielson & McGreal, 2000). McGrath (2000) reports that "Currently less than 1 percent of the permanent teaching staff nationwide receive anything other than the highest marks on the summary evaluation report" (p. 34). Langlois and Colarusso (1988) believe that the lack of precision in teacher evaluation has led supervision and evaluation to become empty, time-stealing rituals.

**Lack of communication.** In order for teacher evaluations to be meaningful and facilitate professional growth the principal's role needs to change from judge to mentor or coach (Danielson & McGreal, 2000). In order to make evaluations a useful tool for professional growth there must be a sense of trust, which can only be established through the use of honest, unbiased feedback. Feedback, as defined by Seashore, Seashore, and Weinberg (1997), is "information about past behavior, delivered in the present, which may influence future behavior" (p. 3).

Peterson & Peterson (2006) write, "one of the main conflicts in principal roles is between principal as encourager/developer of teachers and principal as summative judge" (p. 68). Often times the communication centered on evaluation are from the principal while the teacher plays a passive role of receiver. In this scenario the teacher may or may not value what is being told of them and react in an emotional manner thus limiting any potential professional growth. "The inherent conflict between the functions of coaching and evaluation. Even the strongest proponents of coaching argue that educators must be very clear about which hat they are wearing

when they provide feedback on teaching" (Danielson, 2001, p. 13). Poston and Mannatt (1993) stress that coaching requires accurate feedback, multiple sources of evidence, and patience in order to improve an individual's performance. "Our observation is that administrators simply can't handle conferencing with evaluatees in ways that contribute to growth and improvement" (Poston & Mannatt, p. 45). If teachers are to grow professionally under the new evaluation system the principal will play a critical role communicating throughout the process. "It is part of principal competence to know who needs reflective talk, who needs encouragement, who needs confrontation, and who works well (or needs to be) left alone" (Peterson & Peterson, 2006, p. 70).

In the current system of evaluation, many teachers look at the system as being top down. A case study conducted by Halverson et al. (2004), which looked at 14 schools within a large school district in Western US, found that the evaluations examined lacked either formative or critical feedback:

> The evidence from our case study schools suggest that evaluators lacked the skills to provide valuable feedback, particularly with accomplished teachers. Evaluators instead used evaluation as an opportunity to work with novice teachers and to build a positive school culture rather than as an opportunity to push instructional practices to the highest levels. (p. 177)

Prior research shows that teachers did not feel that they received enough feedback from their evaluations to tell them how they were doing or to help them grow professionally (Frase & Streshly, 1994; Howard & McColskey, 2001; Kimball & Milanowski, 2009; Marshall, 2005; McGrath, 2000). In looking at effective feedback from evaluations Marshall (2005) wrote,

> These evaluations don't tell teachers where they stand on clearly articulated performances standards, don't give clear direction on the ways in which teachers can improve their performance, and don't answer the question teachers really care about (and often dread): How am I doing? (p. 731)

Seashore et al. (1997) write that giving feedback is not easy and that people are often afraid to offer feedback because they either lack the experience or have given poor feedback in the past that ended with a poor result.

Many of the leading experts on this topic (Danielson & McGreal, 2000; Marshall, 2005; Stronge, 1991) agree that providing effective feedback is essential for a teacher to grow professionally. "Experienced practitioners argue that professional dialogue about teaching, in a safe environment managed and led by teachers, is the only means by which teachers will improve their practice" (Danielson & McGreal, 2000, p. 9). The use of dialogue is also supported by Seashore et al. (1997), who state that, in theory, the most efficient method of learning is through the feedback that is provided by others. Krein (1990) advises principals to communicate what was observed instead of the conclusions drawn from the observation. "Principals can develop their communication effectiveness with additional training, reading, and practice" (Peterson & Peterson, 2006, p. 73).

**Limited administrator experience.** Previous studies show that principals rate their skills in executing the teacher evaluation process very high (Painter, 2000; Poston & Manatt, 1993). This finding is in opposition to research, which has found that teachers do not feel that evaluations are meaningful or lead to their professional growth. This disconnect could come from limited administrator experience. Elmore (2000) believes that leaders are not adequately trained to meet the needs of the 21st century.

The current evaluation process was created in the 19th century when instruction was teacher centered, and books were the primary source of information. Today, with the advancement of technology, how students learn has changed dramatically, and no longer are teachers and textbooks the main sources of information (Westinghouse, 2010). Unfortunately many principals are not prepared to evaluate in the 21st century and lack the pedagogical and technological knowledge necessary to create a vision and support the professional development of teachers.

Danielson and McGreal (2000), Kimball and Milanowski (2009), and Wise et al. (1985) report that principals are at a disadvantage as evaluators if they do not have a command of the content being taught or have pedagogical knowledge. "These limitations on the principal as an evaluator of teachers often seriously impair the effectiveness of teacher evaluation processes" (Wise et al., 1985, p. 110). Not having a background in the subject area could make it difficult for principals to identify inaccuracies in the information being presented, especially in advanced courses (Danielson & McGreal, 2000).

Education in the 21st century is changing with the advancement of technology. As education reform takes place there is a push to integrate technology. This is evidenced in the Common Core State Standards (CCSS), which have been adopted by 43 states, and 3 territories (Common Core State Standards Initiative [CCSSI], 2012). As instructional leaders it is imperative that principals play a key role in technology integration but often lack the knowledge of technology standards and experience with technology integration to be effective (Westinghouse, 2010).

Danielson and McGreal (2000), when writing about limited administrator experience and its impact on evaluation, wrote, "….and they may not be familiar with the proper use of graphing

calculators in a mathematics classroom" (p. 6). Technology has become more sophisticated since graphing calculators, and principals are not prepared to meet this new learning style. The evaluation process is undermined when the administrator lacks experience and knowledge, which leads to the perception that evaluation has little value (Danielson & McGreal, 2000). "The literature does identify preparation and experience as important in evaluator decision making, along with evaluator attitudes and the process followed" (Kimball & Milanowski, 2009, p. 64).

> Another factor is the evaluator's own knowledge or familiarity with job content. Although there is some evidence that familiarity with job content is associated with more accurate rating the research is somewhat mixed, and there has been little attention to whether evaluators with experience in performing the job or who have a knowledge base in evaluatee's occupation rate more accurately. This is a potentially important issue in teacher evaluation because school administrators may not have much knowledge or experience with all academic subjects, particularly at the secondary level. (Kimball & Milanowski, 2009, p. 39)

**No differentiation between novice and experienced practitioners**

The evaluation process established by most districts is the same for all teachers regardless of their experience and previous performance ratings (Danielson & McGreal, 2000). New teachers are more likely to need more support and guidance as they build their skill set compared to the professional needs of a veteran teacher. Peterson and Peterson (2006) report, "There is much agreement that extensive evaluation is a benefit for beginning teachers and is less needed for successful veterans" (p. 11). Therefore, the evaluation system should be differentiated in how teachers are rated, but also in the way it challenges teachers to increase their professional growth (Aseltine, Faryniarz, & Rigazio-DiGilio, 2006; Peterson & Peterson, 2006). In

Danielson's model, there is a professional development track that is designed for most tenured teachers within a school district.

The purpose of the professional growth track is to provide a structured, supportive, and collaborative environment to promote professional learning that will further the district's mission and enhance student learning. This track will provide a continuous cycle of assessment to ensure that all tenured staff continue to meet the district's standards for effective teaching (Danielson & McGreal, 2000, p. 100). "A uniform framework allows those conversations to guide novices as well as to enhance the performance of veterans" (Danielson, 2007, p. 6).

Not only are teachers evaluated under the same standards and process, but they are rated using a system that does not differentiate level of performance (Marzano, Frontier, & Livingston, 2011; Poston & Manatt, 1993). The evaluation system was originally developed to document that teachers met minimum competencies in order to receive tenure, and as a result the rating system was simplistic. Marshall (2009) reports, "the rationale for binary scales is to prevent divisive comparisons and get teachers to read their principals' detailed write-ups. But because virtually all teachers are rated Satisfactorily, a two-level scale doesn't really judge teachers' performance" (p. 30). As a result, an expectation developed among teachers that they would be rated satisfactorily and considered themselves as top performers (Marzano, Frontier, & Livingston, 2011). An evaluation system that does not have rigor and a differentiated rating structure gives the wrong message to teachers by having them believe that it is alright to be status quo and motivate teachers to improve (Marshall, 2009). "Performance appraisal should be valid, reliable, and legally discriminating, meaning it should be truthful, consistent, and have the ability to sort high performance from medium to low performance" (Poston & Manatt, 1993, p. 45). The failure to assess variations in instructional effectiveness also precludes districts from

identifying specific developmental needs in their teachers. The evaluation system under the RTTT reform initiatives is designed to differentiate teacher effectiveness using multiple sources and rating categories that take into account data in student growth. NYS's new system categorizes elements of evaluation: identification and interpretation of data, the use of scoring rubrics, and effective feedback. As of this study, NYS has approved ten rubrics for local adoption. Implementation began in the 2011-2012 school year for educators in grades 3-8, and all teachers will be evaluated under the new system during the 2012-2013 school year. As a result of these new components, principals will need to possess skills to utilize the new process and become effective evaluators.

**Principal Preparation Programs**

Research shows that one of the major contributing factors to student achievement is effective school leadership, which is second to having highly effective teachers in the classroom (Dyer & Renn, 2010; MET Project, 2012). Yet the effectiveness of leadership programs has been in question, as "research thus far has found no correlation between leadership programs and principal effectiveness or indices of effective schools" (Lashway, 2003, p. 1). Kimball and Milanowski (2009), as well as Lashway (2003), report that evaluator training has traditionally been frontloaded in preparation programs. Lashway (2003) continues to state that this is usually followed by, "informal, self-guided, and sporadic professional development" (p. 4). Many principal preparation programs stress management and need to move to a model that emphasizes instructional improvement (Mitgang, 2012). When examining the future trends in leadership development, Petrie (2011), from the Center for Creative Leadership, concluded that the nature of the challenges that leaders are facing is rapidly changing; however, the methods that are being used to develop leaders are staying the same.

The problem with professional development is noted by Dyer and Renn (2010), who state that educational leaders understand the importance of professional development but are reluctant to allocate resources for their own professional development. "A recent Public Agenda survey found that 69 percent of principals and 80 percent of superintendents believed that typical leadership programs are out of touch with the realities of what it takes to run today's school district" (Lashway, 2003, p. 1). This is also supported by Witters-Churchill (1991), who writes that principals are the worst critics of the preparation programs that they had attended. Mitgang (2012), in a report published by the Wallace Foundation, states that University-based programs are flawed because they consist of

> criteria that fail to take into account the needs of districts and diverse student bodies; weak connections between theory and practice; faculty with little or no experience as school leaders; and internships that are poorly designed and insufficiently connected to the rest of the curriculum, and lack opportunities to experience real leadership. (p. 6)

*The Texas Study*, which was conducted by The University Consortium (1991) and cited by Witters-Churchill (1991), found that the respondents in the study felt that their formal preparation programs only prepared them moderately for their administrative roles. The same respondents also felt that programs should focus on specific job related skills such as scheduling, filling out forms, teacher evaluations, budgeting, communication, motivation, managing conflict and managing discipline (Witters-Churchill, 1991). Catano and Stronge (2006) state, "it is important for university programs that prepare students for educational administration to identify skills that school principals need in order to be effective principals" (p. 225). Kimball and Milanowski (2009) found that the training that new principals receive is primarily focused on understanding the procedures of the system and managing the process. Principals need to be

prepared to lead in an environment that is typified by an increased level of complexity and interconnectedness (Petrie, 2011).

Mitgang (2012) identified five key components necessary to ensure that schools have effective principals who are capable of moving districts forward. One of the components is "aspiring principals need pre-service training that prepares them to lead improved instruction and school change, not just manage buildings" (p. 2).

Many preparation programs include a field-based experience, but as Witters-Churchill (1991) states regarding *The Texas Study*, *lecture and discussion* was the most frequently used instructional mode for the development of skills although respondents found it only to be minimally to moderately effective. Principals state that instructional modes, which allow for performance-based skill development, are preferred over other modalities such as lecture and discussion (Witters-Churchill, 1991). Mitgang (2012) reports that "many programs offer internships, but the Stanford research found that they commonly settle for fleeting experiences and passive exercises, such as shadowing a principal" (p. 11). This is supported by Lashway (2003) who cautions that "field-based knowledge has obvious practical value but is oriented around existing practices rather than reforms that may be needed" (p. 3). *The Texas Study* found that the internship was the overwhelming favorite of respondents when asked to select the instructional mode that should be used in the development of skills and they rated it as moderately to highly effective (Witters-Churchill, 1991). In a follow up to the University Consortium study, which focused on the performance-based preparation of principals, California has emphasized performance-based standards in its preparation programs, which has "confirmed our earlier commitment and strengthened our continued move in that direction" (Erlandson,

1986, p. 71). Smith's (1986) research showed that both content and method of delivery influenced the effectiveness of error rater training.

For thirty years, research has addressed the deficit in principal preparation programs. Principals seldom receive the appropriate professional development to be effective evaluators and to make the evaluation process one of professional growth for the teacher (Kersten & Israel, 2005). Poston and Manatt (1993) concurred when they wrote, "current levels of administrative training are insufficient to provide appropriate competence in teacher evaluation or to demonstrate some positive impact upon student achievement" (p. 42). Langlois & Colarusso (1988) described principal preparation programs as "backward schools of education, where professors seldom diverge from their antiquated lecture notes" (p. 33). The research conducted by Kersten and Israel (2005) found that administrators are not adequately trained in the evaluation process due to time constraints, which negatively affects the accuracy of evaluations.

Wise et al. (1985) found that teacher evaluation, properly done, is a difficult undertaking. McGrath (2000) points out that "site administrators have little, if any, training in the human dynamics that allow them to maintain the required level of relatedness while providing constructive feedback on the performance of permanent teachers" (p. 35). Without the proper training and support the evaluation process has been viewed as a tedious task and, once completed, has been filed in the teacher's personnel file with little or no impact on subsequent professional growth. This lack of training of principals is one of the pitfalls in the teacher evaluation process. Marshall (2005) writes,

> It takes experience and savvy for a principal to grasp the subtleties of a classroom; it's even more demanding for a principal to capture them in writing; and it's really challenging to criticize a teacher's performance in a way that is heard. Some principals

> are good at all three- observation, write-ups, and "difficult conversations". Unfortunately, many principals are not, and the training needed to bring them up to speed is woefully lacking. (p. 730)

School leadership does have an impact on student achievement through the recruiting, motivating, and developing of teachers (Dyer & Renn, 2010). The new evaluation system will have a ripple effect on the educational system thus impacting the role and responsibilities of the principal. "Areas of teacher support, such as professional development, induction, and mentoring; areas related to curriculum, instruction, and assessment; and human resource areas, such as recruiting, hiring, and compensation, are all affected by changes to teacher evaluation" (Shakman et al., 2012, p. 6). Mitgang (2012) cautions that "mentoring often takes on a relationship similar to a "buddy system" where the mentor is well-meaning but inadequately trained and the experience is weakly connected to district needs" (p. 24). The reformed evaluation system will now require principals to identify and analyze data, give effective feedback to teachers, and facilitate discussions on research-based instructional practices. In addition principals will have to learn to use the state approved rubrics to evaluate teachers using criterion-referenced measures.

Elmore (2002) reminds us that when expectations and standards change, support must be given to those who are responsible for carrying them out. Therefore, in order for the reformed evaluation system to work, principals must be given the support they need to meet the new expectations.

> Change cannot happen unless those who are expected to change are given the opportunity to build the capacity required to make that change. Accountability must be a reciprocal process. When improved performance is expected, there is a responsibility to make

> investments in developing the skills and knowledge of those who are expected to improve. (Shakman et al., 2012, p. 8)

RTTT calls for improvement in the effectiveness of teacher and principal preparation programs. The need for continuous professional development has been recognized if leaders are to acquire the skills needed to be successful in a rapidly changing educational system. "A growing number of states are supporting induction and professional development by mandating 'second –level' certification that requires formal mentoring, reflection, portfolio development, and/or on the-job demonstration of skills" (Lashway, 2003, p. 4). Dyer and Renn (2010) write that "the public school setting has some unique elements that demand a customized approach to the development of its leaders" (p. 3). In NYS, the Board of Regents has approved new certification standards for both aspiring teachers and building leaders. In order to increase the rigor and ensure that building leaders are qualified and skilled as outlined in the RTTT legislation, the NYS Board of Regents has approved the SBL Performance Assessment, which is grounded in the Interstate School Leadership Licensure Consortium (ISLLC) standards and will emphasize instructional leadership tasks. "Candidates will be required to analyze student achievement data, observe classroom instruction, and provide teachers with feedback and support to improve their effectiveness at delivering Common Core-aligned lessons" (NYSED Office of Higher Education [OHE], 2012, para. 18).

It is perceived that many administrative preparation programs do not adequately prepare aspiring principals to be effective evaluators. "Historically in the United States, leaders at the school level (principals) and school district level (superintendents) began their careers primarily as classroom teachers. This common professional background has generated a population of

educational leaders who share similar motivation, knowledge, and challenges" (Dyer & Renn, 2010, p. 3).

When administrators start their first job, they rarely have any experience or training in conducting teacher evaluation. In fact, most beginning administrators do not have experience evaluating teachers until their first assignment. Once evaluations are completed, they are often placed into the employee's personnel file with no review or feedback to the observer in order to strengthen their evaluation skills (Kimball & Milanowski, 2009). Teacher evaluations are often not reviewed until a performance concern is noted and a teacher may be facing disciplinary charges or being denied tenure (Kimball & Milanowski, 2009). It is during these times that the employee's file is examined and feedback is given, often by a school attorney, as to what the mistakes and weaknesses are with the written documentation. Danielson and McGreal, (2000) state, "those making evaluative judgments must be adequately trained so their judgments are accurate, consistent, and based on evidence" (p. 22). "The ultimate goal of rater training is to bring about performance ratings that are valid (accurate) and reliable" (Smith, 1986, p. 36). The results from *the Texas Study* showed that respondents favored the internship experience but called for higher quality, full time, paid internships that led to on-the job training (Witters-Churchill, 1991). "There are early indications that investing in better leadership training can pay off in higher student performance and lower principal turnover" (Mitgang, 2012, p. 25).

**Adult Learning**

"In order for people in schools to respond to external pressure for accountability, they have to learn to do their work differently and to rebuild the organization of schooling around a different way of doing their work" (Elmore, 2002, p. 5). Student achievement can only increase

if educators are life long learners who do not become static in their professional growth, but who look for ways to increase their own knowledge and skills.

> Conventional wisdom assumes that adults stop developing at around 20 years old – hence the term "grow up" (you have finished growing). However, developmental researchers have shown that adults do in fact continue to progress (at varying rates) through predictable stages of mental development. At each higher level of development, adults "make sense" of the world in more complex and inclusive ways – their minds grow "bigger." (Petrie, 2011, pp. 11-12)

One of the most effective ways to grow is through self-awareness. "Like all other leaders, educational leaders need to understand their individual strengths and development needs. This understanding is arrived at through assessment tools, simulations, experimental activities, and staff and peer feedback" (Dyer & Renn, 2010, p. 6). Adults can learn in one of two ways, horizontal or vertical development (Petrie, 2011). Horizontal development is when competencies are transmitted from an expert. This form of development is for learning technical skills and is best utilized when there are clearly defined problems with known techniques as solutions. Adults also learn and develop through vertical development. Vertical development is when individuals become the facilitators of their own learning and grow by facing limitations that require them to analyze and test their assumptions. Vertical development compel individuals to make sense of the world in more complex and inclusive ways and increases their capacity of thinking (Petrie, 2011).

Professional development is most effective when it is a collaborative effort that is imbedded into the school setting. McCauley and Brutus (1998) note that experience is a key component of learning and has long been a part of adult learning theories. "Getting pre-service

training right is essential. But equally important is the training and support school leaders receive after they're hired" (Mitgang, 2012, p. 24). According to Elmore (2002) professional development that takes place outside of the setting through conferences or workshops are not as effective because you are taking someone else's solution and trying to apply it to your unique situation. "As learning is fundamentally a social process, district's participation in a community of practice has the potential to move the field forward in a way that learning in district silos cannot" (Shakman, et al. 2012, p. 23).

Dyer and Renn (2010) highlight four principles of adult learning: "Adults need to know why they need to learn something, adults need to learn experientially, adults approach learning as problem solving, adults learn best when the topic is of immediate value" (p. 4). Danielson and McGreal (2000) write, "The principles of adult learning show that when people use self-assessment and self-directed inquiry in professional development, they are more likely to sustain their learning, in more disciplined ways, than when outsiders impose professional development requirements" (p. 25). Williams and Szal (2011) reinforce that assessment needs to be part of a principal preparation program if the program is to authenticate learning and help principals revise their practices. They also assert that ongoing evaluation of knowledge skills, dispositions, and overall academic progress are essential. "We learn not so much from our experience, but from our reflection on our experience." (Schon 1983, as quoted by Danielson & McGreal, 2000 p. 24).

> Research with educators has shown the impact of training increases as the design of the program expands beyond presentations and modeling of skills and behaviors to include practice, feedback, and peer coaching. Presentations and modeling generate conceptual

> understanding but little in the way of skill attainment or ongoing application (Dyer & Renn, 2010, p. 5.)

"Generally speaking, the more actively involved raters become in the training process, the greater the outcome" (Smith, 1986, p. 37). Research shows that practice with feedback is effective in building the skills necessary for evaluators to conduct more accurate and valid evaluations (Kimball & Milanowski, 2009; Smith, 1986).

> The evidence suggests that the best way to increase accuracy is to combine the two training approaches. Before raters are asked to observe and evaluate the performance dimensions on which they will be rating. They should also be given the opportunity to practice rating sample performance. Finally, they should be provided with "true" or expert ratings to which they can compare their own ratings (Smith, 1986, p. 37).

Kimball and Milanowski (2009) also suggest that another way to make the training process active for raters is to have them think aloud while making evaluation decisions while being observed. They also suggest that another active approach is to have evaluators keep a log of the evidence that they have collected and how they interpret it. Providing on-going supports to building administrators so that they develop the skills necessary to be effective evaluators of teachers is important because the evaluation of teachers is only as good as those who conduct them.

Various researchers over the years have developed different frameworks to explain the various levels of cognitive development. Petrie (2011), from the Center for Creative Leadership, explains cognitive development through three action logics: dependent conformer (shaped by expectations of others), independent achiever (development of internal compass and beliefs), and interdependent collaborator (reflect on ideology; able to resist polarized thinking). In order to

progress through the various developmental stages, individuals must experience vertical development by analyzing and testing their assumptions when faced with limitations at their current stage. Through this vertical development, the stages are earned for oneself. Figure 1 is an illustration of the three developmental stages, as defined by the Center for Creative Leadership, and how limitations and vertical development both serve as the catapult to the next stage.

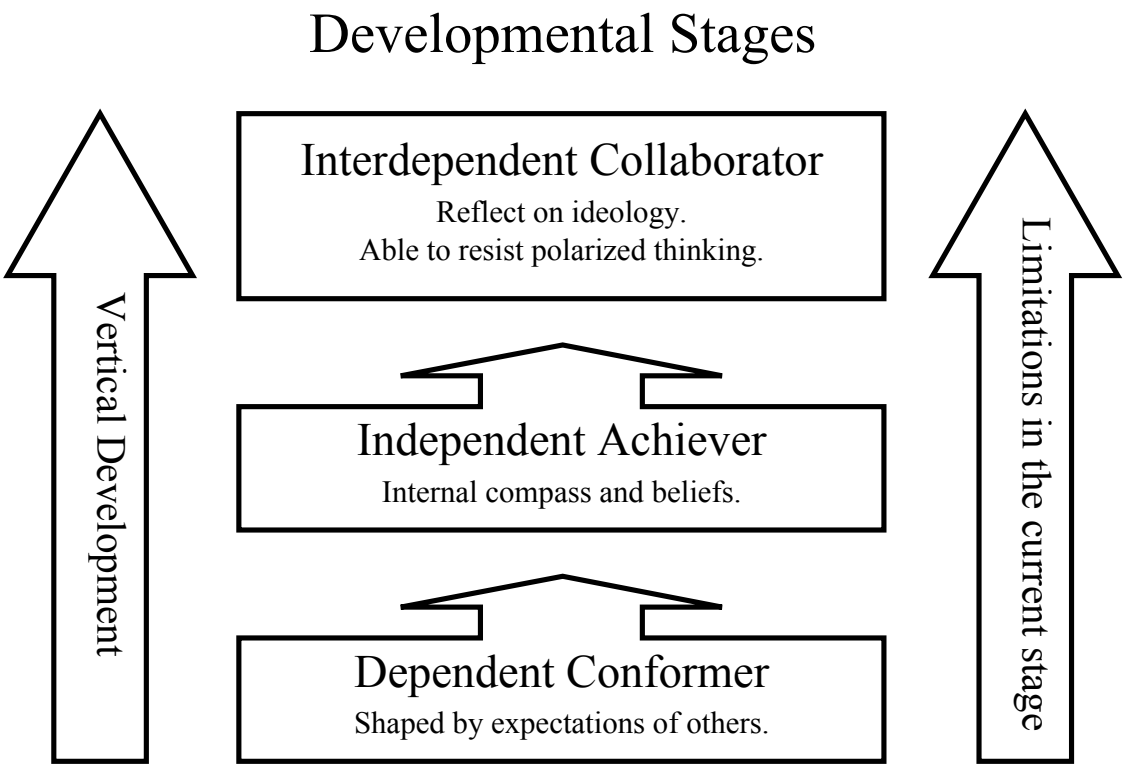


*Figure 1*. Developmental stages as defined by Petrie (2011) from the Center For Creative Leadership.

**Summary**

In this chapter, the literature was reviewed in regards to accountability and how teacher evaluation has been a focus in order to increase student achievement. The review of the literature indicates that having highly effective teachers in the classroom is the primary way to increase student achievement. As a means to encourage teachers to grow professionally and develop into proficient and distinguished educators, teacher evaluation has been identified as the key tool. Transforming the teacher evaluation process from a simple documentation of competency to a tool for professional development requires administrators to possess new skills.

The chapter highlighted the importance of education in this country and the vital role it plays for both individuals and society in general. Education is at the core of this country's existence, yet many in society are questioning its effectiveness. Many question the college and career readiness of the students who graduate from US school systems. The concern over student readiness and global competitiveness has led to a number of reform movements aimed at increasing student achievement.

As Brimley and Garfield (2008) report, school systems are slow to respond to public pressure and adapt. The unresponsiveness of schools to change their systems in order to increase student achievement has forced the federal government to take legislative action to bring about the change necessary for sustainability.

NCLB was the first mandate that focused on the recruiting and retaining of highly qualified teachers in order to increase student achievement. It was through this legislation that there was a focus on standards for both students and educators. NCLB raised the bar challenging both staff and students to reach their capacity.

RTTT is the latest reform legislation aimed at improving the quality of our schools and increasing student achievement. Teacher evaluation is one of the major components found in the RTTT legislation making it the first mandate to design a rigorous evaluation system. It is through RTTT that the teacher evaluation system is being transformed into a valuable tool to strengthen instructional delivery as a means to increase student achievement.

Included in the review of literature was a brief historical overview in the function and purpose of the evaluation system. It was not until recently that teacher evaluations became a *living document* that not only reflect a teacher's competence, but also establishes goals and incentives to grow professionally. The teacher evaluation process, as designed under RTTT, has become an active and collaborative process encouraging communication, and self- reflection about a teacher's performance when aligned with a clear set of criteria. It is through this change that the evaluation system will be more meaningful to teachers by providing them with effective feedback and data that can be used to monitor and guide their performance.

In order to make teacher evaluation a beneficial process that leads to teacher growth administrators must possess a new set of skills. It is the professional responsibility of building leaders to ensure that they have the skills necessary to be effective evaluators of teachers.

Principals can learn to avoid the most common mistakes made by evaluators. These mistakes lessen the value of the evaluation process and have, over time, made it a passive system with little impact on professional growth or on school effectiveness in general. If principals continuously reflect on their work as evaluators, become knowledgeable of the most common errors, and receive feedback about the quality of their evaluations, they will develop into effective evaluators of teachers. The professional development, support, and guidance offered to

principals by districts can no longer be sporadic. It must be ongoing in order to ensure that principals develop the skills necessary to be effective evaluators under the new system.

The review of literature also strongly suggested that principal preparation programs are ineffective in preparing individuals to be instructional leaders. There is a strong call for principal preparation programs to become skills-based with a focus on field placement. "Current administrator preparation programs have an extraordinary opportunity to upgrade the quality of trained administrators. This can be accomplished through the systematic use of performance-based preparation programs" (Erlandson, 1986, p. 73). Even though many programs are incorporating a performance-based model, Lashway (2003) cautions its use. Although performance-based programs teach the skills and methodologies often sought by beginning principals, it can foster a status quo system and suppress change.

The review of literature laid the foundation for this study by confirming the importance of having highly qualified teachers in the classroom to increase student achievement. One of the tools used to facilitate teacher growth is the evaluation process. Principals require a specific set of skills that lead to reliable and valid evaluations that teachers find both effective and meaningful.

This study examined principal readiness and professional development to conduct effective teacher evaluations that lead to improved student achievement. This research adds to the knowledge base of teacher evaluation by providing important insights in the area of supports needed by current and future principals to conduct valid and reliable teacher evaluations. This study was designed with the hope of increasing the knowledge about ongoing supports needed by principals. It results in a deeper understanding of what strategies and supports principals feel are effective and beneficial to them to enhance their skills as evaluators of teachers. The results of

this study can help system leaders create internal systems that include practical applications and increase the validity and reliability of the evaluations conducted by principals.

## Chapter 3: Methods

This quantitative research was a self-report study through the distribution of a survey that examined the perception of principals in Upstate NY on their training to become effective evaluators of teachers. It utilized an exploratory instrument to capture the opinions of principals on the topic. It also investigated the ongoing professional development support that principals identified as beneficial to become effective evaluators that results in teacher self-reflection, growth, and improved student achievement. More specifically, this research was designed to help identify those ongoing evaluator supports necessary to strengthen principals' skills to conduct the teacher evaluation process. Through this study, the evaluation systems and practices that schools are currently using are identified as well as the changes that schools are experiencing due to RTTT legislation. The research was conducted after approval of the Sage Colleges Institutional Review Board (IRB) was secured.

### Research Questions

This study sought to determine which key professional development practices principals report to be associated with their ability to conduct valid and reliable teacher evaluations. It is through this study that the questions of what additional professional development principals report they need to become effective evaluators of teachers was explored. This study investigated the following four questions:

1. In what ways do principals report that they have been prepared to be an effective evaluator?
2. What is the relationship between reported professional development and perceived readiness to conduct effective teacher evaluations?

3. Is there a difference in perceived readiness to conduct effective teacher evaluations between principals at different building levels, specifically those who are now required to conduct evaluations under Race To The Top Legislation versus those who have not yet been expected to do so?
4. What additional professional development do principals report that they need to become effective evaluators of teachers?

**Research Design**

A quantitative approach was used based on research study descriptions given in the literature. A search to locate all possible studies that examined the topic of rater training was conducted using a number of computerized databases. These databases included ERIC, Academic Onefile, Academic Search Elite, Dissertations and Theses: Full Text, and Proquest Central. Creswell (2009) writes, "Quantitative research questions inquire about the relationship among variables that the investigator seeks to know. They are used frequently in social science research and especially in survey studies" (p. 132). This study examined the relationship between principals as evaluators and professional development for professional growth by measuring their perceptions and experiences. According to Babbie (1998) the purpose of survey research falls within one or more of three general objectives: description, explanation, and exploration. The research design used in this study was primarily descriptive. Descriptive research involves collecting data to determine the distribution of certain traits or attributes to describe the current status of the topic being studied (Babbie, 1998). In addition to descriptive statistics, correlations were used to determine if relationships existed between variables, and cross tabulations and chi-square analyses were used to determine if statistically significant patterns of responses were present.

**Target Population / Sample**

All 4,775 NYS public school principals in schools containing grades 3-12 responsible for conducting teacher evaluations were initially considered as the potential target population for this study. This group was chosen because all NYS administrators are mandated to conduct teacher evaluations as prescribed through the RTTT legislation that was signed into law in 2009. Another reason for choosing this group was that the building principal is responsible for the direct supervision of teachers and conducting their evaluations.

The subjects selected for inclusion in the actual sample for this study were the 297 building administrators within the NYS Greater Capital Region who work in districts that fall within the jurisdiction of Capital Region BOCES, Questar III BOCES, or the Washington, Saratoga, Warren, Hamilton, Essex BOCES (WSWHE BOCES). BOCES are Boards of Cooperative Educational Services. BOCES originated and were formed in 1948 by the NYS legislature as a way to help districts meet the needs of their students in a cost effective way. BOCES offer shared educational programs and services to districts. BOCES facilitate the ability of districts to partner and collaborate in order to offer more programming to their students by having them pool resources and share costs, thereby creating opportunities that they might not be able to afford otherwise.

This sample represents approximately 6% of the principals in NYS. The reason this convenience sample was chosen was because the research was an exploratory study on the topic regarding supports and professional development needed by principals to become effective evaluators of teachers. Due to the relevancy of the research, given the new legislation, and the familiarity with the institution conducting the study, it was anticipated that a survey within this area would generate a good response rate, therefore resulting in reliable and valid data. The

findings from this study provided useful information that can be a foundation for further inquiry for a larger population.

The three BOCES regions selected for this study are comprised of 11 counties (Albany, Columbia, Essex, Green, Hamilton, Rensselaer, Saratoga, Schenectady, Schoharie, Warren, and Washington) in Upstate NY. The region includes 78 public school districts, which consist of 290 schools in total. According to the 2010-2011 New York State Report Card Database, there are a total of 297 building principals employed within these districts, and together they supervise 12,196 teachers who are responsible for delivering instruction and increasing achievement to 154,140 students (NYSED, Office of Prekindergarten Through Grade 12 Education [P12], Information and Reporting Services [IRS], 2012). Of the 290 schools included in this study, 61 elementary/middle-level and nine high schools have been identified by NYSED as being school buildings not in good standing. Table 1 shows the demographic data for the three BOCES regions included in this study. Principals and their email addresses were identified using the component directories of the three BOCES and the individual websites of the schools.

Table 1

*Demographic Data of the Three BOCES Regions Included in the Study*

| BOCES Region | Number of Schools | Identified Schools | Enrollment | Number of Teachers | Number of Principals |
|---|---|---|---|---|---|
| Capital Region | 137 | 33 | 77820 | 5969 | 144 |
| Questar III | 73 | 24 | 34984 | 2795 | 74 |
| WSWHE | 80 | 13 | 81336 | 3432 | 79 |
| Total | 290 | 70 | 154140 | 12196 | 297 |

Table 2 gives student demographic information regarding free and reduced lunch rates of the regions included in the study. The purpose of this information is to help give an overview of

the student populations being served in the three regions included in this study, since effective teacher evaluation is aimed at increasing student achievement.

Table 2

*Free and Reduced Lunch Rates of the Three BOCES Regions Included in the Study*

| BOCES Region | Number of Students Receiving Free and Reduced Lunch Rates | Percent of Students Receiving Free and Reduced Lunch Rates |
|---|---|---|
| Capital Region | 11054 | 27% |
| Questar III | 11344 | 32% |
| WSWHE | 21723 | 28% |
| Total | 44121 | 29% |

Table 3 is a chart of student sub-groups who are accounted for thorough NLCB legislation within the three BOCES regions.

Table 3

*Student Demographics of the Three BOCES Regions Included in the Study*

| | Asian | | Black | | Hispanic | | Multiracial | | LEP | |
|---|---|---|---|---|---|---|---|---|---|---|
| BOCES Region | N | % | N | % | N | % | N | % | N | % |
| Capital Region | 5133 | 7% | 10853 | 14% | 3965 | 5% | 682 | 1% | 1450 | 2% |
| Questar III | 665 | 2% | 3287 | 9% | 1597 | 5% | 459 | 1% | 432 | 1% |
| WSWHE | 458 | 1% | 827 | 2% | 579 | 1% | 202 | 0% | 125 | 0% |
| Total | 6256 | 4% | 14967 | 10% | 6141 | 4% | 1343 | 1% | 2007 | 1% |

**Instrumentation**

A design was chosen that allowed for the collection of data in a way that was comfortable to the respondents and to address the study's four research questions. Given the many responsibilities of administrators, the design selected needed to be considerate of their time and encourage them to participate. The survey design was selected because of the economy of the

design and the ability to collect data from a large number of participants in a relatively short period of time. The initial survey request was sent out Wednesday, April 11, 2012, which was during the school spring recess (see Appendix A). It was the researcher's belief that those administrators working that week would have more time to participate in the study since the day-to-day operational requirements of their buildings were assumed to be minimal. The study also fell during the week of the New York State 3-8 math assessments (April 16 – 20, 2012), which could have negatively impacted the response rate because building principals had to focus on overseeing the coordination, administration, and scoring of these mandated high stakes tests, which fell within the same time frame of this study.

The type of research design used was a cross sectional, self administered survey using scaled items so that data about the characteristics, experiences, knowledge, and perceptions in regards to principals and their training on becoming effective evaluators could be efficiently and effectively collected (Creswell, 2009). A cross sectional survey collects data at one point in time. The use of a survey design was selected to "provide a quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population" (Creswell, 2009, p. 145). One of the advantages of using a survey is that it allows the researcher the ability to identify attributes of the large population from the smaller group of individuals who participated.

Due to the limited research on this topic, the survey instrument was a multi-section survey tool created by the investigator (see Appendix B). The survey was comprised of 25 questions requiring approximately 15 minutes to complete. The survey consisted of both perception and demographic questions. Section one was the survey introduction. The introduction was a brief overview of the study and the potential contribution that it could make to

the field. It was in the introduction that candidates were also told that their individual responses would be confidential allowing them to be candid and honest. Section two consisted of questions about respondent demographics that were essential to make the data meaningful for the study. The purpose of the demographic questions was to differentiate traits among the participants such as the grade levels supervised, years of administrative experience, and type of school (rural, urban, city) (see Appendix B, questions 1-11). Section three solicited the respondent's perceptions of readiness to conduct effective evaluations (see Appendix B, questions 12-21). Section four consisted of a scaled question designed to examine effective resources, one matrix question regarding professional development and effectiveness regarding teacher evaluation preparation and two open ended question to gather any information on the topic that may not have been answered in the formal questions (see Appendix B, questions 22-25). The opinion questions were based on professional development and conducting teacher evaluations. The final section of the survey, section five, consisted of a statement of appreciation for completing the survey. Within the survey, prompts were used, such as "You're about 1/3 done" and "You're on the last section of the survey," to show respondents that they were making progress (see Appendix B).

**Data Validity and Reliability**

In an effort to strengthen the face validity of the survey the researcher distributed the survey to a panel of experts and met with each one individually to get feedback. The panel consisted of five experts in the field. The panel consisted of retired and former principals who have a range of building level administrative experience and provided feedback as practitioners in the field. In addition, the researcher was in contact with the director of professional development for the School Administrators Association of New York State (SAANYS).

SAANYS is a NYS administrative association that has served and represented school leaders for over three decades. The researcher also received feedback, electronically through email, from a national expert who has done ongoing research and presented on the topic of school leadership. This individual, as a retired administrator, has also published books and articles on the topic of evaluation. The researcher also sought feedback and recommendations from the data analyst at the WSWHE BOCES during the development of the survey.

The panel of experts was asked to provide feedback on the survey content to be sure that the questions adequately measured principal readiness and professional development to conduct effective teacher evaluations that lead to improved student achievement. The panelists were also asked for suggestions for instrument improvement in both format and in the wording of questions. It was the role of the panelists to ensure that the survey directions and questions were not ambiguous thus helping to increase the reliability of the instrument. In all, five experts with experience in the field of education provided constructive information that helped to craft the survey tool. Overall, the panelists provided positive feedback regarding the efficiency, relevance, and thoroughness of the survey. The modifications, as suggested by the panelists, included re-ordering some questions for clarity and transition purposes and ensuring that response options were clear and exhaustive. The survey instrument developed by the researcher was determined to have face validity and reliability through the use of a panel of experts.

**Data Collection**

Email addresses were identified through the three BOCES component directories and individual district websites. The directories are public and, as a result, were obtained from each BOCES website. A list of all building principals was compiled from these resources. The NYS Report Card Database reported that there were 297 principals within the 11 identified counties,

but when looking at the BOCES directories, 279 principals were listed. This discrepancy could be due to the fact that the NYS Report Card Database identifies by counties and the BOCES may have regional lines that divide a county into more than one BOCES region. There may have been some counties that have districts not covered within the three BOCES regions utilized in this study.

Once approval of the Sage Colleges IRB was secured, participants were recruited by disseminating an email requesting research study participation to all principals of the 77 school districts identified in the sample population. The 279 principals identified from these resources were sent an introductory email with the link to the survey. The email described the study and how the research will help guide leadership practices by identifying the support and on going professional development principals need to become effective evaluators, strengthen classroom instruction, and increase student achievement (see Appendix A).

Survey responses were collected between April 11, 2012 and May 10, 2012 and resulted in a 42% response rate. The district superintendents for the three BOCES also helped publicize the survey by mentioning to the component school superintendents at their monthly April meetings. At these meetings, the local district superintendents were made aware of the study and were asked to encourage their building principals to consider participating (see Table 4).

Participants were asked to complete the survey within four weeks of receipt of the survey request letter. Of the 279 emails, 2 were returned as invalid email addresses, reducing the sample size to 277 recipients. A follow-up reminder was sent to the sample population within one week of the original email. The follow up letter reminded them that they recently received an invitation to participate in a SurveyMonkey study giving thanks to those who sent back a

completed survey and reminding those who did not that their participation is important in order to collect enough data to make the study meaningful (see Appendix C).

Table 4

*Sample Calculations*

| | Number of Principals |
|---|---|
| New York State* | 4775 |
| 3 BOCES Regions in Population* | 297 |
| Attempted to send email invitation | 279 |
| Email invitation returned (invalid address) | 2 |
| Successfully received email invitation | 277 |

*Note:* *2010-2011 data from the New York State report card database by the New York State Education Department, Office of Prekindergarten Through Grade 12 Education, Information and Reporting Services, 2012.

Due to a less than 40% response rate after the first two emails a second reminder was sent two weeks after the original email to everyone who had received an email message, but had not responded. To increase response rates two final reminders were sent, to non-respondents, the last week that the survey was open encouraging those who had not completed the survey to consider doing so. The final reminder was sent the day the survey was scheduled to close to respondents. See Table 5 for a breakdown.

Table 5

*Survey Response History*

| Correspondence | Date | Responses to Date | Change |
|---|---|---|---|
| Original Email Request | 4/11/12 | 0 | 0 |
| 1st Email Reminder | 4/17/12 | 63 | 63 |
| 2nd Email Reminder | 4/24/12 | 84 | 21 |
| 3rd Email Reminder | 5/7/12 | 97 | 13 |
| Final Email Reminder | 5/10/12 | 107 | 10 |
| Final Count | 5/10/12 | 115 | 8 |

SurveyMonkey is a web-based program and anonymity cannot be guaranteed since SurveyMonkey.com is an online instrument and email addresses can be traced. In the introductory letter recipients were made aware of this but also assured that their names and specific information that could identify them would not be used. All reported data were kept confidential. SurveyMonkey collects participant responses over a secure, encrypted connection to ensure that data are sufficiently protected and secure. The survey was also designed to collect only the minimum amount of personal information necessary to achieve the desired purpose of the study and to help protect the confidentiality of the respondents. All data received from the survey were securely kept on the researcher's home computer to which no one else had access to and which was password protected. The only individual who had access to the raw data was the investigator. Participants were also assured that there were no known risks associated with this study.

In the original email and all follow up reminders participants were made aware of their right to opt out of all or part of the study without penalty or loss of benefit to themselves. Each email message consisted of a *remove* link that the recipient could click if they did not wish to receive further emails from SurveyMonkey. If a recipient clicked the remove link they were automatically removed from SurveyMonkey's mailing list and did not receive follow up emails.

Only aggregated data are reported – no individual responses are reported or identified. Participation in this study was voluntary, and all data were stored on the investigator's computer and destroyed at the end of the data collection phase by deleting it from the program and permanently deleting it from the computer's trash. A back up copy was kept on a flash drive in a locked cabinet and was destroyed upon completion as well.

A total of 115 building principals responded to the survey tool, representing a response rate of 42%. Of the total responses, 106 fully completed the survey and nine partially completed it. Of the recipients receiving the original email, four opted out of the survey, and 158 never responded. Due to the firewalls and safety measures put in place in many school districts, some of the surveys could have gone to the recipients spam folders or blocked due to network firewalls.

**Data Analysis and Interpretation**

Once the raw data were collected, the data set was transferred from SurveyMonkey to SPSS. The unit of analysis for this study was individual responses by the building principals. The individual responses were grouped in multiple ways (i.e. building level, experience, gender) to find common responses in regards to professional development necessary for conducting effective teacher evaluations. Through the use of SPSS, descriptive and inferential statistics were used to organize, summarize, analyze, and display the data collected from the respondents. The numbers of responses (valid percent) relative to the total possible (percent) are noted.

Descriptive statistics is a method of analysis that is used to summarize the data under study (Babbie, 1998). Descriptive statistics allow the researcher to take quantitative data and describe the variables or associations between the variables. Descriptive statistics allow the researcher to make representations that describe the sample population or in this case the respondents to the survey.

When examining the data the chi-square test of independence was used, "to determine whether two categorical variables are related" (Weinberg & Abramowitz, 2008, p. 493). Chi-quare is one of the most frequently used nonparametric tests used in social science to analyze data frequencies. Chi-square is based on the assumption that no relationship exists between the

two variables in the total population (Babbie, 1998). The chi-square test is used to determine if statistically significant differences exist between the actual survey responses and the expected frequencies. A finding is statistically significant if the actual survey response is not due just to chance (Vogt & Johnson, 2011). Chi-square gives the probability of the relationship existing between the two variables.

Spearman's rho was also used to determine the strength of the relationship between the variables being measured. Spearman's rho was chosen for bivariate correlations due to the ordinal scales used within the survey.

**Limitations and Delimitations**

The utilization of a survey instrument with closed-ended questions resulted in an approach in which participants were required to select a specific response. This closed question format offered the participants a limited number of pre-determined responses from which they had to select thus preventing them from elaborating on their thought process and scaffolding their answers. This approach was selected because it allowed for a larger number of respondents to participate in the study, within the time frame it was conducted, and it was a more efficient way to collect multiple responses for data. There has not been as much research to identify what specific professional development or resources are effective in helping building principals become effective evaluators of teachers, and it was important to use an exploratory instrument that could capture the opinions of principals on the topic. The use of a survey instrument did limit the feedback to the questions. This was especially constraining to the research questions regarding resources and supports. A qualitative design employing interviews, would have allowed for probing questions, which could have garnered more personal, comprehensive and descriptive responses.

The survey was administered by email. Another limitation outside the control of this study was the possibility of firewalls and security measures that many districts have set up which may have prevented the emails from reaching the intended recipient thus preventing access for them to participate in the study. Some servers may have filters that forwarded the emails to the recipients spam folder again lessening the chance of them becoming aware of the study and participating. The unintended technological barriers, preventing participants from receiving the survey in their inbox where it could be easily found, may have negatively impacted the response rate for this study.

The focus of the research was limited to formal observations and did not include informal observations, such as walk-throughs or mini-observations. Marshall (2009) defines mini-observations as "short, focused classroom visits to get a feel for what is happening…They involve significantly more frequent visits to all the classrooms, individual feedback to each teacher, and individualized ongoing communication about teaching and learning based on the observations" (p. 62-63). Mini-observations are categorized as a form of supervision not evaluation (Marshall, 2009). Mini-observations or walk-throughs are a practice that can be used to help access a teacher but is not designed as a sole evaluative tool (Marshall, 2009; Marzano, Frontier, & Livingston, 2011; Peterson & Peterson, 2006). Mini-observations help the building principal form overall impressions of the teachers and provide more background, which can be used to support the formal evaluation. In short, observations are useful for diagnostic purposes and the development of instructional practice (Marshall, 2009; Measures of Effective Teaching [MET] Project, 2012). Mini-observations help facilitate communication by providing immediate feedback to the teacher that is seen as credible, but does not result in a formal written document like an evaluation. Unlike the observations often negotiated with the evaluation process, mini-

observations are unannounced, do not include a pre-observation meeting, and the follow up should be short and informal. This research did not include informal observations due to the differences in the process, objectives, and required documentation. Building principals are required to conduct formal evaluations but not all building level administrators chose to use mini-observations or walk-throughs as a tool to access teachers and provide feedback.

A delimitation in this study was that teacher evaluations affect all building level administrators in NYS, but only building principals were targeted in this study. Other administrators, such as vice principals or assistant principals, were not included.

The population for this study was the 277 building principals within three Board of Cooperative Educational Services (BOCES) geographic areas in Upstate NY. The 115 respondents to the survey comprised the convenience sample, which further limited the study. The researcher was not aware if the principals who work within the three BOCES regions are representative of the state. Therefore, the generalizations made can only be validly applied to public schools in this specific geographic area. Caution will be needed when applying the findings of this study to principals in geographic areas outside of this study and to principals who lead buildings that are structured with different grade level configurations than those identified in this study.

Further, the convenience sample used in this study may not be representative of the population. Therefore one would need to be careful when considering the findings of this study to develop professional development opportunities or when reviewing district procedures.

Response rate will affect reliability and validity and could be a limitation in the study. Limited respondents could affect which statistical tests were used to determine if there is a statistical significance because criterion is not met.

## Chapter 4: Data Analysis

Prior research found that teacher evaluations do not accurately depict a teacher's effectiveness. Legislation such as NCLB and RTTT have mandated changes in the current teacher evaluation system. These were the primary foci that lead to inquiry and resulted in the formation of this study. This study was designed to use a survey as an exploratory instrument to collect multiple responses to see what specific professional development or resources building principals felt were effective in helping them become effective evaluators of teachers. The following research questions were developed as the framework for this study and served as the foundation for the data analysis.

1. In what ways do principals report that they have been prepared to be an effective evaluator?
2. What is the relationship between reported professional development and perceived readiness to conduct effective teacher evaluations?
3. Is there a difference in perceived readiness to conduct effective teacher evaluations between principals at different building levels, specifically those who are now required to conduct evaluations under Race To The Top Legislation versus those who have not yet been expected to do so?
4. What additional professional development do principals report that they need to become effective evaluators of teachers?

This chapter examines the data collected for each of the four research questions, as well as the data analysis methods used. In answering the four research questions the data were analyzed through the use of descriptive statistics and crosstab analysis to determine if there was

any statistical significance, specifically Pearson chi-square and Spearman-Rho helped to analyze any relationship between the variables.

**Demographic Profile of the Building Principal Respondents**

The convenience sample for this research was the 115 building principals who responded to the survey. This sample represents a 42% response rate. The original survey was sent to 277 building principals within the three BOCES regions in the Greater Capital Region of NYS. Simultaneous variables were studied as a means to determine characteristics pertinent to this study that would describe the respondents who participated. In order to report the demographic findings describing the population, distributions using descriptive statistics were used. Frequency distribution, as defined by Babbie (1998), is "a description of the number of times the various attributes of a variable are observed in a sample" (p. 371). Frequencies are reported in this research as counts and percentages of respondents. Frequencies are rounded and reported to the nearest whole number and include complete data, whereas missing values are excluded. "Descriptive statistics is used when the purpose of an investigation is to describe the data that have been (or will be) collected" (Weinberg & Abramowitz, 2008, p. 2). Knowing the demographic background and professional experiences of the respondents is important as it is from these experiences and perspectives that they answered the questions on the survey for this study.

Respondents were asked to identify their gender in the building principal demographics section of the survey (see Appendix B, Q. 1). The results are illustrated in Table 6. The data show that 44% were female, and 56% were male. According to the database containing data on all NYS public and nonpublic schools and administrators, the convenience sample is representative of the regions included in this study. Of the 296 reported on the NYSED

database, 129 (44%) were female, and 167 (56%) were male. According to the same NYSED database, 57% of principals in NYS are female, while 43% are male. The results show that the regions included in this study are not aligned with the state, in that there continues to be fewer female administrators. The results from this study can still be applicable when looked at from a gender perspective, as it would be expected that all females and males would respond similarly, regardless of location, since all have similar guidelines and mandates.

Table 6

*Frequency Distribution of Respondents' Gender*

| Gender | n | % |
|---|---|---|
| Female | 51 | 44% |
| Male | 64 | 56% |

Sixty-nine percent of the respondents worked in districts that had a student population of 1,000 students or more. Forty-nine percent of the respondents worked in districts within the two population ranges of either 1,000 – 1,999 or 2,000 – 4, 999 (see Table 7). This is important, as the size of the district that a respondent comes from can affect the number of teachers that they supervise and thus the number of evaluations that they are responsible for conducting. The size of respondents' districts may account for the way they responded to the survey in this study. Prior to the recent mandates, districts had the autonomy to develop their own evaluation system and select their own measurement instruments. Although it was not specifically analyzed in this study, the size of the district could have determined the evaluation systems used prior to RTTT and therefore could have effected respondents' perceptions of teacher evaluation. Student enrollment gives a general sense of the type of district and setting in which respondents work. This could be used for comparison with similar future studies.

Table 8 shows the configuration that best described the grade levels that the respondent currently supervised. Most of the respondents (73%) supervised grades at the elementary level (K-5), whereas 47% were building principals who supervised grades at the secondary level (6-12). The data show that there was a lot of variability in the configurations that the respondents supervised, thus minimizing grade level alignment between respondents. As a result, some of the respondents indicated that they supervised two or more of the grade level bands on the survey (see Appendix B, Q. 5), which accounts for the high frequency counts in Table 8.

Table 7

*Frequency Distribution of Respondents' District Student Enrollment*

| District enrollment | n | % |
|---|---|---|
| 0 – 199 | 2 | 2% |
| 200 – 499 | 11 | 10% |
| 500 – 999 | 23 | 20% |
| 1,000 – 1,999 | 32 | 28% |
| 2,000 – 4,999 | 24 | 21% |
| 5,000 – 9,999 | 18 | 16% |
| 10,000 – 24,999 | 5 | 4% |

Table 8

*Frequency Distribution of Respondents' Configuration of Grade Level Supervision*

| Grade levels | n | % |
|---|---|---|
| K – 2 | 41 | 36% |
| 3 – 5 | 42 | 37% |
| 6 – 8 | 29 | 25% |
| 9 – 12 | 25 | 22% |
| K-12 | 4 | 4% |

Figure 2 is a scattergram showing how many years of experience the respondents had as a building level administrator. The range of experience the respondents had as a building principal ranged from 1 to 25 years with 7.3 years being the mean. Half (50%) of the respondents had six years or less experience as a building level principal. The other 50% of the respondents ranged from 7 to 25 years of experience as a building level principal.

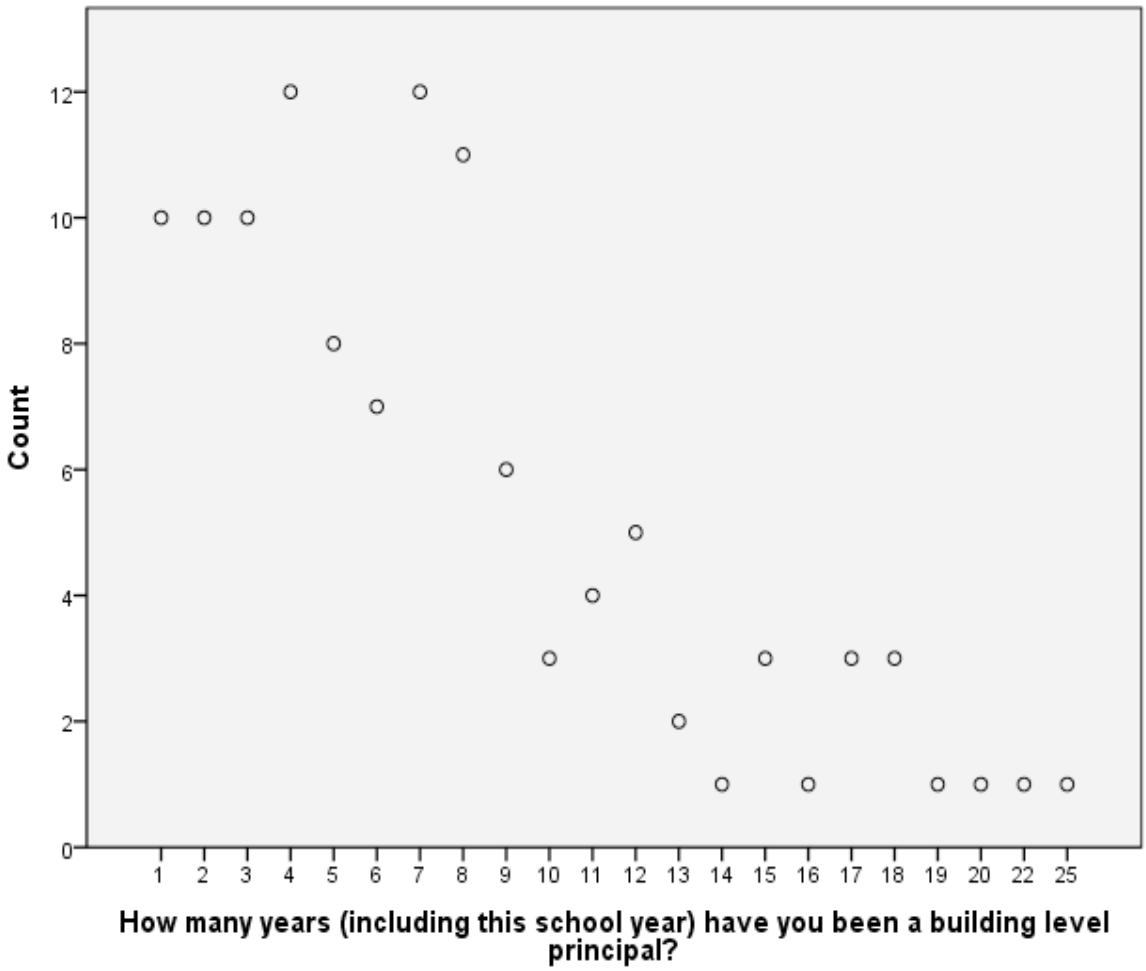


*Figure 2.* Scattergram of the frequency counts of respondents' number of years experience as a building principal.

Fifty-four percent of the respondents were at their current school five years or less, while 7% had between 15 and 22 years of experience in their current school as the building principal.

Figure 3 illustrates the frequency counts for the number of years the respondents had been the building principals in their current schools. The respondents in this study may have been a building principal in multiple districts, thus giving them different experiences since until recently districts had more autonomy over the evaluation system used and how it was implemented. Comparing Figures 2 and 3, it is clear that many of the respondents have had administrative experience as a building principal in more than one district.

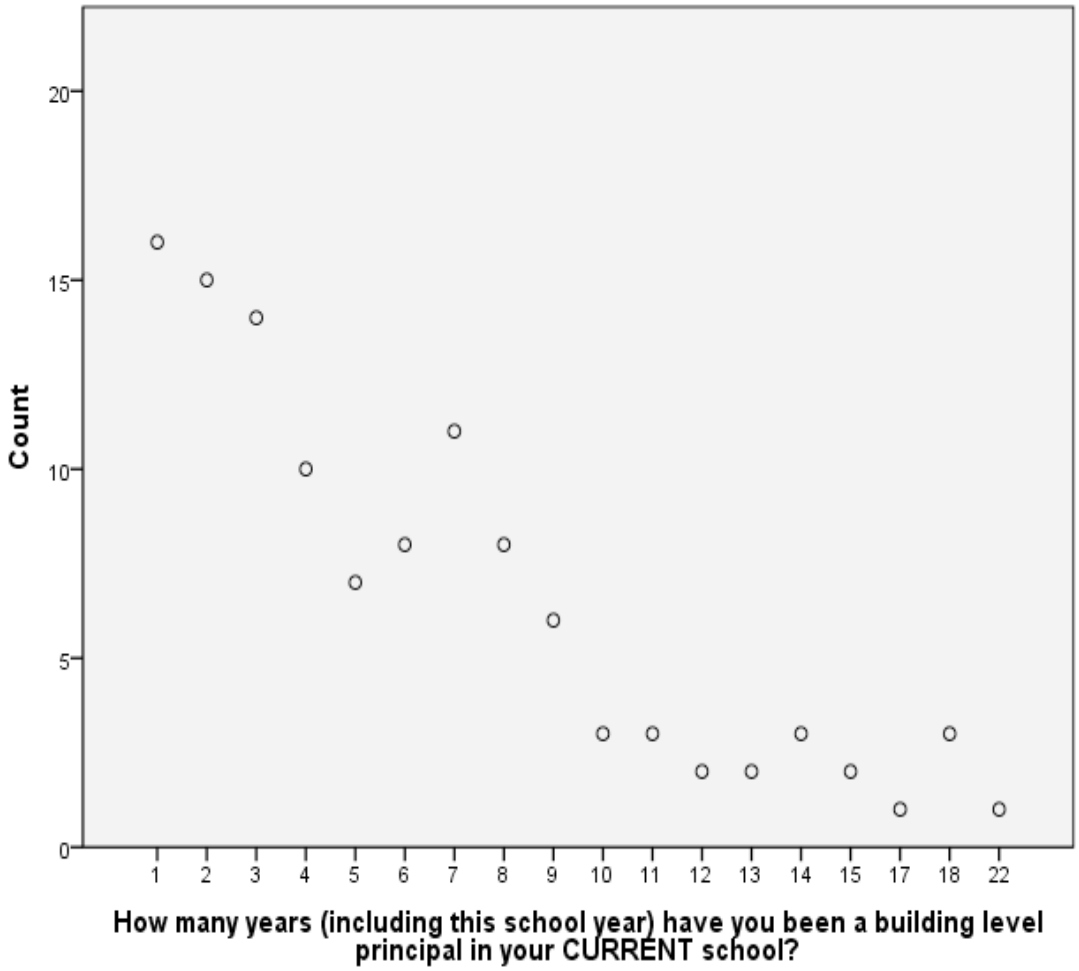


*Figure 3.* Scattergram of the frequency counts of respondents' number of years experience as a building principal in their current school.

Table 9 shows the respondent data regarding experience as a classroom teacher. Of the respondents, 93% had been a classroom teacher, while 7% did not have classroom experience as a teacher. This is important to note because it gives more information about the respondents' backgrounds but more specifically because in several analyses done to answer the four research questions, classroom experience was identified as an effective resource or support in preparing them to conduct teacher evaluations.

Table 9

*Frequency Distribution of Respondents' Experience as a Classroom Teacher*

| Classroom teacher experience? | n | % |
|---|---|---|
| No | 8 | 7% |
| Yes | 106 | 93% |

Of the respondents who reported the number of years that they were a classroom teacher, 51% had 10 or more years' experience as a classroom teacher. The range of classroom teacher experience was from one year to 22 years. Forty-two percent were female, while 58% were male. The mode was 11 years, while the mean was 10 years (see Figure 4).

When looking at previous experience before becoming a building principal, 60% of the respondents reported that they were assistant principals. Table 10 shows the number of respondents who reported to have experience as an assistant principal.

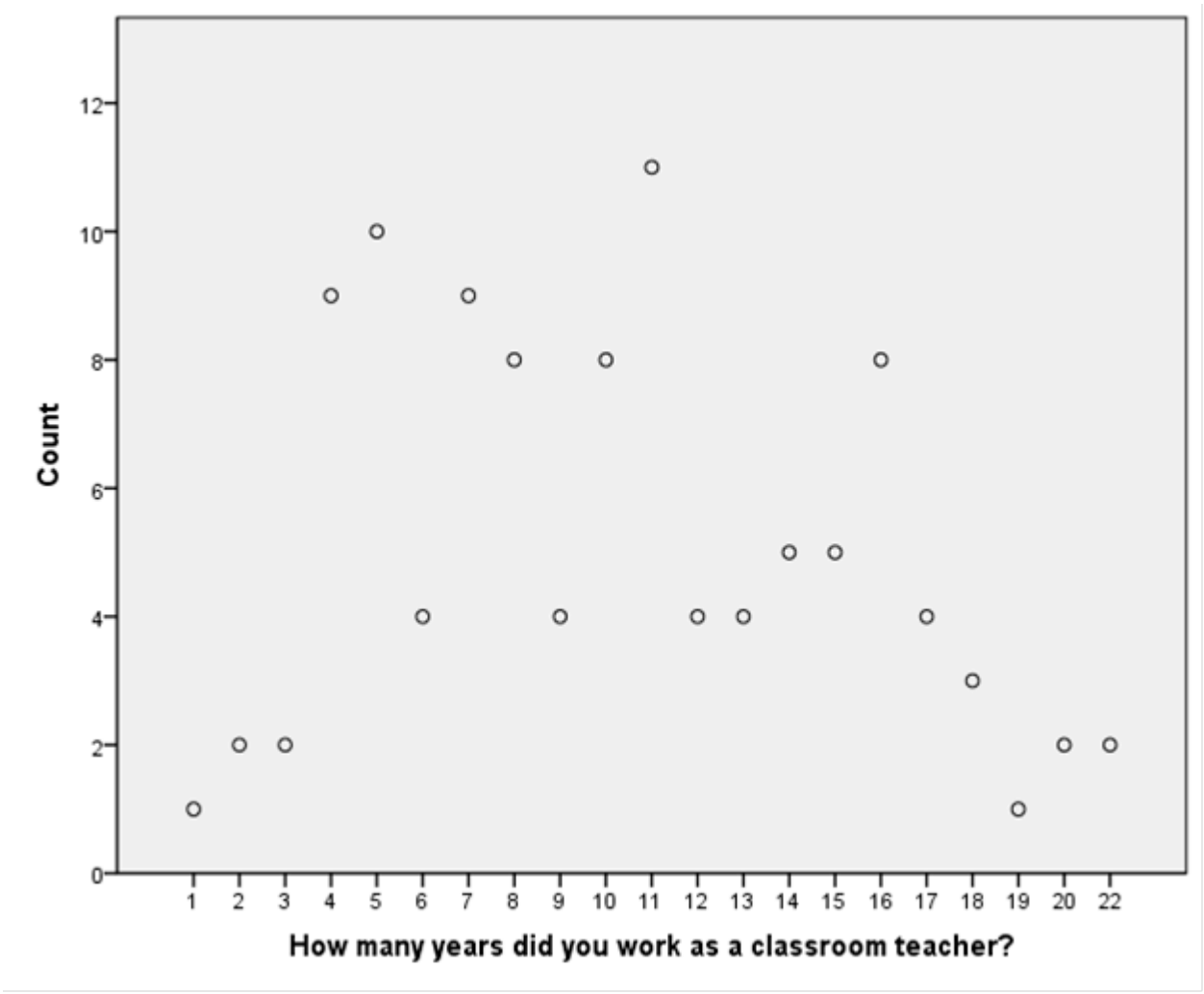


*Figure 4.* Scattergram of the frequency counts of respondents' years of experience as a classroom teacher for those who reported classroom teaching experience.

Table 10

*Frequency Distribution of Respondents' Experience as an Assistant Principal*

| Assistant principal experience? | n | % |
|---|---|---|
| No | 46 | 40% |
| Yes | 69 | 60% |
| Totals | 115 | 100% |

According to the data 68 out of the 69 respondents who reported to have had experience as an assistant principal also reported the number of years of experience that they had in the role. The reported data, in Table 11, shows that 65% of those who reported that they had experience as an assistant principal had three or less years experience in that role. Of the 68 respondents who reported their years of experience as an assistant principal, 13% reported to have been in the role between five and 10 years. The range of experience for the respondents who reported to have been an assistant principal was between one and 10 years with the mean being 3.25 years. The mode was 2 years.

Table 11

*Frequency Distribution of Respondents' Years of Experience as an Assistant Principal for Those Who Reported Experience as an Assistant Principal*

| Years experience as an assistant principal | n | % |
|---|---|---|
| 1 | 13 | 19% |
| 2 | 19 | 28% |
| 3 | 12 | 18% |
| 4 | 9 | 13% |
| 5 | 5 | 7% |
| 6 | 5 | 7% |
| 7 | 1 | 2% |
| 8 | 2 | 3% |
| 10 | 2 | 3% |
| Totals | 68 | 100% |

**Research Question 1:** *In what ways do principals report that they have been prepared to be an effective evaluator?*

When analyzing the respondents' perceptions of their preparedness to evaluate teachers and the resources they felt were effective in their preparation descriptive statistics was used.

When the respondents were asked if they felt prepared to conduct effective teacher evaluations that would result in professional growth for teachers, 13% of them strongly agreed, while 62% agreed. The remaining 25% of the respondents indicated some degree of disagreement with the statement, implying that they did not feel prepared to conduct effective evaluations of teachers (see Table 12).

Because NYS had a differentiated timeline as to when specific grade levels where mandated to implement the new evaluation system, the respondents were asked in the survey as to what year they were mandated to implement. Ten percent of those respondents who will implement in the 2012-2013 school year reported that they strongly agreed with the statement of being very prepared, while 18% of the respondents who implemented in the 2011-2012 school year (the year the study was conducted) reported that they strongly agreed with the statement.

Table 12

*Frequency Distribution of Respondents' Perceptions of Their Readiness to Conduct Effective Evaluations of Teachers*

| Perception of readiness | n | % |
|---|---|---|
| Strongly agree | 15 | 13% |
| Agree | 71 | 62% |

Table 13 shows to what extent the respondents felt prepared with the various elements of the evaluation system. The data for this research question were generated by using a matrix that allowed the researcher to ask the participants to rate a series of elements associated with effective evaluations of teachers.

When looking at all of the respondents, 95% indicated that they perceived some degree of preparedness in implementing the elements of pre- and post-conference of effective evaluations of teachers. Of the 16 elements listed on the survey, at least 90% of the respondents perceived

Table 13

*Frequency Distribution of Respondents' Perceptions of Their Preparedness to Implement Specific Elements of Effective Evaluations of Teachers*

| | Prepared | | | | Unprepared | | | |
|---|---|---|---|---|---|---|---|---|
| | Very | | Somewhat | | Somewhat | | Very | |
| Element | n | % | n | % | n | % | n | % |
| Criteria used for evaluation | 24 | 22% | 79 | 71% | 6 | 5% | 2 | 2% |
| Use of the HEDI scale | 8 | 7% | 66 | 58% | 30 | 27% | 9 | 8% |
| Evaluate assessment data | 21 | 19% | 66 | 58% | 22 | 20% | 4 | 4% |
| Evaluate family and community outreach | 19 | 17% | 51 | 46% | 36 | 32% | 6 | 5% |
| Evaluate lesson plans | 57 | 50% | 50 | 42% | 5 | 4% | 1 | 1% |
| Evaluate student portfolios | 30 | 27% | 57 | 51% | 21 | 19% | 4 | 4% |
| Evaluate student work samples | 38 | 34% | 57 | 51% | 14 | 13% | 3 | 3% |
| Evaluate teacher portfolios | 34 | 30% | 58 | 51% | 16 | 14% | 5 | 4% |
| Evaluation instrument | 25 | 22% | 58 | 52% | 25 | 22% | 4 | 4% |
| Evaluation procedures | 24 | 21% | 66 | 59% | 19 | 17% | 3 | 3% |
| Giving feedback to teachers | 49 | 44% | 53 | 47% | 9 | 8% | 1 | 1% |
| Legal issues regarding evaluations | 7 | 6% | 52 | 46% | 39 | 35% | 15 | 13% |
| Objective observation techniques | 35 | 31% | 67 | 59% | 9 | 8% | 2 | 2% |
| Post conference | 43 | 38% | 64 | 57% | 6 | 5% | 0 | 0% |
| Pre conference | 39 | 35% | 52 | 47% | 5 | 4% | 1 | 1% |
| Use of technology | 36 | 32% | 52 | 47% | 22 | 20% | 1 | 1% |
| Validity/reliability in evaluations | 9 | 8% | 60 | 53% | 33 | 29% | 11 | 10% |

themselves as having some degree of preparedness in six of the elements: pre-conference, post-conference, criteria used for evaluation, evaluating lesson plans, giving feedback to teachers, and

using objective observation techniques. Half of the respondents (50%) felt very prepared to evaluate lesson plans, and at least 50% of the respondents felt some degree of preparedness for all 16 elements measured on the survey. According to the response data, respondents felt less prepared in the use of the new rating scale (HEDI), evaluating family and community outreach, legal issues regarding evaluations, and ensuring validity and reliability in evaluations. In addition to asking respondents for their level of preparedness to implement specific elements of the evaluation process, they were also asked about their perception of resource effectiveness in their preparation to conduct teacher evaluations. Table 14 is a summary of the responses.

The data show that 51% of the respondents felt that on-the-job experience was very effective in their preparation to conduct teacher evaluations, while 98% of the respondents said that on-the-job experience had some degree of effectiveness. At least 90% of the respondents also indicated some degree of effectiveness for collegial discussions and teaching in the classroom.

When principals were asked how they felt about the verbal and written constructive feedback they have received from a supervisor regarding the teacher evaluations that they had written, the results showed that 35% of the respondents said that they had not received any feedback. Of the 70 respondents who reported that they had received feedback, 74% were satisfied, while 20% said that they were very satisfied with the feedback that they received (see Table 15).

Participants were asked how often teacher evaluation training was offered to administrators in their current district (see Table 16). The data show that 56% of the respondents indicated that training was not offered at all, while 21% said it was offered to new administrators only. The remaining 23% of the respondents indicated that some degree of training was offered.

Table 14

*Frequency Distribution of Respondents' Perceptions of the Effectiveness of Resources in Their Preparation to Conduct Teacher Evaluations*

| | Effective | | | | Ineffective | | | |
|---|---|---|---|---|---|---|---|---|
| | Very | | Somewhat | | Somewhat | | Very | |
| Resource | n | % | n | % | n | % | n | % |
| Administrative certification programs | 8 | 8% | 51 | 50% | 26 | 25% | 18 | 18% |
| Books | 7 | 7% | 68 | 65% | 25 | 24% | 4 | 4% |
| Collegial discussions | 46 | 43% | 54 | 51% | 4 | 4% | 2 | 2% |
| Continuing education courses | 12 | 15% | 50 | 63% | 12 | 15% | 6 | 8% |
| Current supervisor | 18 | 18% | 44 | 44% | 24 | 24% | 13 | 13% |
| Experience as assistant principal | 21 | 32% | 37 | 56% | 7 | 11% | 1 | 2% |
| Formal district in-service training | 14 | 17% | 48 | 59% | 12 | 15% | 7 | 9% |
| Former administrative supervisor | 8 | 9% | 47 | 54% | 18 | 21% | 14 | 16% |
| Journal articles | 8 | 8% | 68 | 69% | 18 | 18% | 4 | 4% |
| Mentoring | 14 | 20% | 40 | 58% | 10 | 15% | 5 | 7% |
| On-the-job experience | 54 | 51% | 50 | 47% | 2 | 2% | 0 | 0% |
| Modeling by former admin. when being evaluated as a classroom teacher | 12 | 16% | 39 | 51% | 20 | 26% | 6 | 8% |
| Race To The Top trainings | 13 | 14% | 44 | 47% | 20 | 21% | 17 | 18% |
| Teaching in the classroom | 39 | 39% | 53 | 53% | 8 | 8% | 1 | 1% |
| Videos | 9 | 10% | 52 | 60% | 19 | 22% | 7 | 8% |
| Webinars | 3 | 4% | 41 | 55% | 21 | 28% | 9 | 12% |
| Workshops/conferences | 20 | 21% | 59 | 63% | 13 | 14% | 2 | 2% |

Table 15

*Frequency Distribution of Respondents' Perceptions Verbal or Written Feedback That They Received From a Supervisor Regarding the Teacher Evaluations That They Had Written*

| Perception of feedback | n | % |
|---|---|---|
| Very satisfied | 15 | 20% |
| Satisfied | 55 | 74% |
| No feedback | 40 | 35% |

Table 16

*Frequency Distribution of Respondents' Reported Frequency of Teacher Evaluation Training Provided to Administrators in Their Current District*

| Perception of feedback | n | % |
|---|---|---|
| Throughout the year | 8 | 7% |
| Annually | 12 | 11% |
| Biannually | 5 | 5% |
| New admin. only | 23 | 21% |
| Not offered at all | 60 | 56% |

To explore whether a relationship existed between their level of satisfaction with their initial introduction to the evaluation forms, process, and criteria and their perceived level of preparedness, a nonparametric correlation, Spearman's rho, was used. The relationships were found to be statistically significant: for forms ($r = .3$, $p < .01$), for process ($r = .26$, $p < .01$), and for criteria ($r = .28$, $p < .01$). Using the correlation strengths from Cohen's (1988) work on statistical power analysis, $r = .1$ is weak, $r = .3$ is moderate, and $r = .5$ is strong. Therefore, the relationships between the two variables ranged from moderate to approaching moderate. The data show that as the respondents' levels of satisfaction with their initial introduction to evaluation increased so did their agreement that they felt prepared to conduct effective teacher evaluations that would result in professional growth for teachers. Further, the data show that

those with some form of training (24%) were two times more likely to say that they strongly agreed with the statement that they were prepared to effectively evaluate teachers compared to those who reported that no training was offered at all (12%). The descriptive data further show that 27% of those who reported that no training was offered at all disagreed or strongly disagreed with the statement about being prepared compared to 16% who reported that some form of training was offered.

**Research Question 2:** *What is the relationship between reported professional development and perceived readiness to conduct effective teacher evaluations?*

This research question was addressed through the use of a nonparametric correlational technique, Spearman's rho. Reported professional development and perceived readiness to conduct effective teacher evaluations were the two variables used in the analysis. Correlation strength was determined from Cohen's (1988) work on statistical power analysis as weak ($r$ = .1), moderate ($r$ = .3), and strong ($r$ = .5). Table 17 shows Spearman's rho values for these relationships.

When looking at the data between reported professional development and perceived readiness to conduct effective teacher evaluations, there is an indication that there may be a relationship, although it is not strong. Data reveal a significant relationship of moderate strength for professional development, which includes formal district in-service training ($r$ = .43, $p$ < .01), videos ($r$ = .41, $p$ < .01), current supervisor ($r$ = .31, $p$ < .01), and journal articles ($r$ = .30, $p$ < .01). The professional development activities that had a significant relationship of weak strength included teaching in the classroom ($r$ = .29, $p$ < .01), modeling by former administrators when being evaluated as a classroom teacher ($r$ = .28, $p$ < .05), on-the-job experience ($r$ = .27, $p$ < .01), books ($r$ = .24, $p$ < .05), and collegial discussions ($r$ = .23, $p$ < .05). There is also evidence that

the supervisor plays a role in professional development and that there is a statistical significance with the current supervisor.

Table 17

*Correlations for Reported Professional Development and Perceived Readiness to Conduct Effective Teacher Evaluations*

| Professional development resource | *r* | *p* |
|---|---|---|
| Formal district in-service training | .43** | .000 |
| Videos | .41** | .000 |
| Current supervisor | .31** | .002 |
| Journal articles | .30** | .003 |
| Teaching in the classroom | .29** | .003 |
| Modeling by former administrator when being evaluated as a classroom teacher | .28* | .014 |
| On-the-job experience | .27** | .005 |
| Books | .24* | .015 |
| Collegial discussions | .23* | .020 |

Note. * $p < .05$. ** $p < .01$.

After looking at the group as a whole and discovering that there was a statistically significant correlation between perceived readiness to conduct effective teacher evaluations and specific resources, cross tabulations were performed to determine if there was a pattern of responses between the resources and the respondents' years of experience (1-10 years vs. 11+ years). The chi-square tests performed were not found to be statistically significant, but a pattern emerged which showed that those respondents who had 1-10 years of experience were more likely to answer that the resource had some degree of effectiveness compared to those respondents who had 11+ years of experience.

In order to gain more insight into this relationship, further analyses were conducted using descriptive frequencies and cross tabulation to determine if there were any reported patterns between the variable, perceived readiness, and the variables, gender and years of experience.

The data from this study revealed that men have higher rates of agreement that they felt prepared when compared to females. Whereas, the female respondents indicated that they were more likely to disagree that they are prepared to evaluate teachers (see Table 18).

The Pearson chi-square was performed in SPSS and the relationship between gender and preparedness was found to be statistically significant ($\chi^2 = 3.83$, $p = .05$). This reveals that building principals who are male are more likely to feel prepared to be effective evaluators of teachers than their female colleagues.

Table 18

*Frequency Distribution by Gender of Respondents' Perceptions of Their Preparedness to Conduct Teacher Evaluations*

| | Agree | | Disagree | |
|---|---|---|---|---|
| Gender | n | % | n | % |
| Female | 34 | 67% | 17 | 33% |
| Male | 52 | 83% | 11 | 18% |

Looking at years of experience, the data was at first coded into three categories: early career (1-10 years), mid-career (11-20 years), and late career (21-25 years) but this did not meet the criteria to run further inferential analysis, so the data was recoded into quartiles (1-3 years, 4-7 years, 8-10 years, and 11+ years). This recoding was necessary since 77% of the respondents of the survey would have been identified as early career using the original categories. Table 19 shows the respondents' perceptions of preparedness according to their years of experience as a building principal.

Table 19

*Frequency Distribution by Years of Experience of Respondents' Perceptions of Their Preparedness to Conduct Teacher Evaluations*

| | Agree | | Disagree | |
|---|---|---|---|---|
| Years of experience | n | % | n | % |
| 1-3 years | 22 | 76% | 7 | 24% |
| 4-7 years | 31 | 80% | 8 | 21% |
| 8-10 years | 16 | 80% | 4 | 20% |
| 11+ years | 17 | 65% | 9 | 35% |

The chi-square analysis was not statistically significant, but when looking at the data descriptively, it was revealed that those with 1-10 years of experience had a higher agreement about feeling prepared to evaluate teachers than veteran principals who had 11+ years of experience. Given this revelation, further analysis was done to see which resources the two groups identified as being most effective in preparing them to be evaluators of teachers. The results of this comparison are displayed in Table 20.

The data from this analysis show that veteran administrators with 11+ years of experience reported a higher degree of effectiveness for activities that can be done independently, such as reading (books, case studies, journal articles), watching videos, and reviewing sample evaluations rated as exemplars. Those respondents with 1-10 years of experience rated resources that were more collaborative and formal as being more effective for them. This group rated resources such as participation in a review and discussion of evaluations written by other colleagues as the most effective, but they also indicated a higher degree of effectiveness for participation in workshops and having peers review your evaluations in order to give feedback.

Table 20

*Frequency Distribution of Respondents' Perceptions of Effective Resources by Years of Experience as a Building Principal*

| | 1-10 years | | 11+ years | |
|---|---|---|---|---|
| Resource | n | % | n | % |
| Workshops on specific elements | 63 | 90% | 16 | 76% |
| Workshops on the evaluation process | 61 | 91% | 18 | 82% |
| Conduct tandem evaluations | 37 | 88% | 10 | 71% |
| Have peers review your evaluations | 43 | 90% | 9 | 75% |
| Participate in a review and discussion of evaluations written by other colleagues | 45 | 94% | 14 | 88% |
| Participate in role plays | 21 | 60% | 8 | 53% |
| Participate in webinars | 31 | 65% | 7 | 50% |
| Participate in online learning courses | 13 | 42% | 5 | 42% |
| Read books | 57 | 92% | 19 | 95% |
| Read case studies | 35 | 69% | 12 | 75% |
| Read journal articles | 58 | 89% | 17 | 94% |
| Review exemplars | 52 | 91% | 17 | 100% |
| Take additional college courses | 21 | 66% | 6 | 55% |
| Take courses through a professional organization | 38 | 83% | 10 | 67% |
| Watch videos | 45 | 79% | 17 | 94% |

**Research Question 3:** *Is there a difference in perceived readiness to conduct effective teacher evaluations between principals at different building levels, specifically those who are now required to conduct evaluations under Race To The Top legislation versus those who have not yet been expected to do so?*

In order to analyze the data to answer this research question, responses from the participants had to be coded to create new variables. In the survey, Q. 5 asked the participants to

select all of the configurations that best described the grade level that they currently supervised. The respondents had five different configurations to choose from (see Appendix B). It was necessary to engage in the coding process since the respondents chose from configurations that then had to be composed into the variables being studied (implementing vs. not yet implementing). Under the RTTT regulations, only those teaching grades 3-8 had to be evaluated using the new system in the 2011-2012 school year. All other grade levels, K-2 and 9-12, were not required to implement the new evaluation system until the 2012-2013 school year. Those who supervised any grade in 3-8 were categorized as *implementing* while those who only supervised in K-2 or 9-12 were categorized as *not yet implementing*.

When looking at the participant responses, there were slightly more respondents who were already implementing the new evaluation system under RTTT who strongly agreed in their readiness to conduct effective teacher evaluations when compared to those respondents who were not yet implementing. The data also reveal that there were a higher percentage of respondents who were not yet implementing that had some degree of perceived readiness (93%) as compared to 72% of the respondents who are already implementing (see Table 21).

Table 21

*Frequency Distribution by Participants' Race To The Top (RTTT) Implementation Stage of Respondents' Perceptions of Readiness to Conduct Effective Teacher Evaluations*

| | Strongly agree | | Agree | |
|---|---|---|---|---|
| Implementation stage | n | % | n | % |
| Already implementing RTTT | 9 | 15% | 34 | 57% |
| Not yet implementing RTTT | 2 | 13% | 12 | 80% |

Further analysis needed to be conducted as a means of exploring this research question because the various grades that the participants reported that they supervised varied greatly,

therefore making any cross tabulation difficult to perform. A review of the responses showed there was overlap in principal supervision between the grade levels that are now required to conduct evaluations under RTTT legislation and those grade levels that have not yet been expected to do so. The data would not have been valid if the question was analyzed looking at grade levels alone, since some principals supervised grades that fell between both variables of implementing and not implementing. Further hindering any cross tabulation analysis was the limitation that one-third of the respondents did not respond to the question. The two-thirds who did respond most represented those who were already implementing. The chi-square analysis was not statistically significant.

After considering various ways to code the data that would help answer the research question, it was decided to examine it by focusing on when they were mandated to implement the new APPR regulations as articulated by RTTT (see Appendix A, Q. 6). As shown in Table 22, respondents indicated that 26% were implementing the new evaluation system at the time of this study, and 69% would begin implementing in the 2012-2013 school year.

Table 22

*Frequency Distribution of Respondents' Reported Timeline for Mandated Implementation of the New APPR Regulations as Articulated by Race To The Top (RTTT)*

| Implementation timeline response option | n | % |
|---|---|---|
| This year (2011-2012) | 29 | 26% |
| Next year (2012-2013) | 77 | 69% |
| I have not been informed | 6 | 5% |

The data for this question were examined using chi-square, and no statistical significance was found. Descriptive statistics had to be used to answer this question. The data revealed that there was a higher degree of frequency of preparedness from those who were implementing the

new evaluation system at the time of the study (2011-2012) compared to those who were not yet mandated to implement. See Table 23 for the results.

Table 23

*Frequency Distribution by Participants' Race To The Top (RTTT) Implementation Year of Respondents' Perceptions of Preparedness to Evaluate*

| | Agree | | Disagree | |
|---|---|---|---|---|
| Year of implementation | n | % | n | % |
| This year (2011-2012) | 24 | 86% | 4 | 14% |
| Next year (2012-2013) | 57 | 74% | 20 | 26% |

**Research Question 4:** *What additional professional development do principals report that they need to become effective evaluators of teachers?*

One source of data for this research question was generated by designing an open ended question on the survey. The reason for designing a question of this type was to be able to capture any themes, resources, or supports that were overlooked in the closed ended questions on the survey. An open ended question could also allow for new emerging ideas, which have not been discussed in the literature, regarding supports or resources principals need in order to become effective evaluators of teachers.

The data analysis for this question involved coding the responses. The researcher read through the responses and got a general sense of what the participants had written. Then the researcher identified the key words in each response and attributed it to a label. The researcher then began to categorize the responses that had similar labels by giving it a code. In the final step the researcher grouped each of the codes under a category in order to perform a preliminary analysis.

In the analysis of the responses, three supports and resources clearly emerged: workshops, collegial collaboration, and NYSED guidance and clarification. Forty-seven percent of the respondents to this question identified some type of trainings (workshop, conference, in-service, seminar) as being a needed form of professional development to become effective evaluators of teachers. The three elements identified from the respondents regarding trainings were that they needed to include real evaluation examples (11%), the specific instrument that was adopted by the district (44%), and that the trainings should be specific to the needs of the participants or region (11%).

Collegial collaboration was the second most common form of professional development identified as needed to become effective evaluators of teachers (34%). Three elements emerged as forms of collaboration: discussions, tandem observations, and peer feedback. For those that identified collegial collaboration as a resource, 46% defined it as being associated with collegial discussions, whereas 31% defined it as tandem observations. Of the respondents who identified collaboration as a needed support, 23% of them stated that they wanted more collaborative time to discuss actual evaluations written by them or their colleagues (peer feedback).

The third form of professional development identified by the respondents (11%) was guidance or clarification on teacher evaluations from NYSED.

In addition to analyzing the data from the open ended question to answer what additional professional development principals reported that they need to become effective evaluators of teachers, descriptive statistics were used to analyze Q. 23, which asked the participants to select how frequently they used and how effectively a prescribed list of resources was in their development as effective teacher evaluators. The question was designed as a matrix due to the advantages, especially those of efficiency of space and participant responses. Using a matrix

question allows the researcher to have the participants respond to several sub questions within the general question. This saves space by having the questions compressed versus asking each one individually. It also is more efficient for the participants because it allows them to answer the questions faster and eliminates the need for them to read the same or similar questions several times. By using a matrix design, the participants are able to easily compare their responses for strength of agreement, strengthening both the validity and reliability of their answers. A matrix design also allows the researcher to compare responses to the different questions. Table 24 summarizes the respondents' responses.

Over half (54%) of the respondents reported that tandem evaluations were a very effective resource in supporting them to become effective teacher evaluators. No other resource was rated as being very effective by more than half of the respondents. Having peers review evaluations written by the respondent (peer review) and participating in a review and discussion of evaluations written by other colleagues were both rated as highly effective by 38% of the respondents.

When looking at the data in which the respondents found some degree of effectiveness in the resource, 94% reported that reviewing sample evaluations rated as exemplars would be effective in supporting their professional development to become effective teacher evaluators. This was closely followed by 93% reporting that reading books on effective evaluations and participating in a review and discussion of evaluations written by other colleagues would be an effective resource in developing their skills to become effective evaluators of teachers. In the resources listed, at least 80% or more of the respondents stated that they had some degree of effectiveness. Four of the resources were reported as effective by at least 90% of the respondents.

Table 24

*Frequency Distribution of Respondents' Perceptions of the Effectiveness of Resources in Their Development as Effective Teacher Evaluators*

| | Very effective | | Somewhat effective | |
|---|---|---|---|---|
| Resource | n | % | n | % |
| Review of exemplar evaluations | 19 | 26% | 50 | 68% |
| Books on effective evaluations | 9 | 11% | 67 | 82% |
| Review of colleague evaluations | 24 | 38% | 35 | 55% |
| Journal articles on effective evaluations | 12 | 15% | 63 | 76% |
| Workshop on evaluation process | 18 | 20% | 61 | 69% |
| Specific instructional workshops | 18 | 20% | 61 | 67% |
| Peer review | 23 | 38% | 29 | 48% |
| Tandem evaluations | 30 | 54% | 17 | 30% |
| Videos on effective evaluations | 8 | 11% | 54 | 72% |
| Professional organizational courses | 8 | 13% | 40 | 66% |
| Case studies on effective evaluations | 5 | 8% | 42 | 63% |
| College classes on effective evaluations | 6 | 14% | 21 | 49% |
| Webinars | 2 | 3% | 36 | 58% |
| Role play | 6 | 12% | 23 | 46% |
| Online learning courses | 3 | 7% | 15 | 35% |

When looking at which resources the respondents said that they used most frequently, 15% said that they often read journal articles on effective evaluations, while 14% responded that they often attended instructional workshops focused on specific elements within the evaluation process (pre/post conferencing, collecting evidence, giving effective feedback). When looking at frequency of use, 5% or less of respondents reported often use of 12 of the 15 resources listed.

The data revealed that 71% of the respondents had some degree of use in reading books on effective evaluations. This was followed by 67%, who reported to have used instructional

workshops focused on the overview of the evaluation process to some degree, and 62%, who used to some degree instructional workshops focused on specific elements within the evaluation process (pre/post conferencing, collecting evidence, giving effective feedback). Table 25 is a summary of the respondents' frequency of use for the specified resources.

Table 25

*Frequency Distribution of Respondents' Perceptions of Their Frequency of Use of Resources in Their Development as Effective Teacher Evaluators*

| | Often | | Sometimes | |
|---|---|---|---|---|
| Resource | n | % | n | % |
| Books on effective evaluations | 8 | 8% | 64 | 63% |
| Journal articles on effective evaluations | 15 | 15% | 55 | 54% |
| Workshop on evaluation process | 12 | 12% | 56 | 55% |
| Specific instructional workshops | 14 | 14% | 49 | 48% |
| Videos on effective evaluations | 4 | 4% | 42 | 41% |
| Review of exemplar evaluations | 3 | 3% | 41 | 40% |
| Case studies on effective evaluations | 2 | 2% | 29 | 28% |
| Professional organizational courses | 2 | 2% | 24 | 24% |
| Webinars | 1 | 1% | 22 | 22% |
| College classes on effective evaluations | 5 | 5% | 16 | 16% |
| Peer review | 5 | 5% | 16 | 16% |
| Review of colleague evaluations | 2 | 2% | 18 | 18% |
| Tandem evaluations | 3 | 3% | 14 | 14% |
| Online learning courses | 2 | 2% | 6 | 6% |
| Role play | 1 | 1% | 6 | 6% |

**Summary**

Findings of the data analyses of principal readiness and professional development to conduct effective teacher evaluations that lead to improved student achievement are outlined in

this chapter. There were four research questions that provided the structure and focus for this study. The analyses conducted showed a disconnect between principal training and their ability to effectively evaluate teachers. The findings show that building principals coordinate their own professional development, choosing activities in isolation, but feel that collaborating with others would be effective. Both gender and years of experience played a role in the respondents' perceptions of readiness. As an exploratory study, it was the task of this research to draw conclusions regarding these findings and their implications for building principals, system leaders, and principal preparation programs.

Chapter 5 consists of a thorough discussion of the findings and conclusions. Suggestions for improvement and recommendations for further research related to principal readiness to conduct effective teacher evaluations are included.

## Chapter 5: Summary of Findings and Recommendations

Evaluation is a reflection of human performance in organizational settings (Stronge, 1991). "Teacher evaluation is a major component of the educational agenda today" (Tucker & Stronge, 2005, p. 25). Legislation aimed at educational reform, such as No Child Left Behind (NCLB) and Race To The Top (RTTT), has changed the objective of teacher evaluation (Tucker & Stronge, 2005). The use of evaluations as a form of documentation has changed to that of a tool for professional growth and personnel decision-making (Marshall, 2009; Marzano et al., 2011; NYSED, 2012; Peterson & Peterson, 2006; Shakman et al., 2012; Tucker & Stronge, 2005). Evaluations should be designed to promote teacher effectiveness. "In fact, most authors identify the fundamental purposes of teacher evaluation as improving performance and documenting accountability" (Tucker & Stronge, 2005, p. 6). The *Guidance on New York State's Annual Professional Performance Review For Teachers and Principals To Implement Education Law §3012-c And the Commissioner's Regulations* states, "The purpose of the evaluation system is to ensure that there is an effective teacher in every classroom and an effective leader in every school. The evaluation system will also foster a culture of continuous professional growth for educators to grow and improve their instructional practices" (NYSED, 2012, p. 6).

Shakman et al. (2012) define teacher evaluation as a "rigorous system that includes frequent observations, with validated protocols, evidence of teacher practice, and student outcomes, and measures of student learning" (p. 3). Due to accountability and the change in objectives, how evaluations are structured and conducted are beginning to change (Danielson & McGreal, 2000; Marshall, 2009; Marzano et al., 2011; Tucker & Stronge, 2005; USDOE, 2009).

"Better defined goals have shifted the educational dialogue from vague opinions about student progress to factual evidence of student performance" (Peterson & Peterson, 2006, p. 91).

RTTT makes instructional staff and leaders accountable for student outcomes and aims to improve teacher and principal effectiveness based on their performance (USDOE, 2009). As a result of this change, teacher evaluation has transitioned from focusing on teaching to focusing on student learning. Johnson (1997) writes, "by providing the public, administrators, and teachers with data on teachers' skills and performance, a teacher-evaluation system will increase the instructional productivity of teachers, enhance student learning, and ultimately improve the quality of the educational system" (p. 70).

This change requires that building principals have a new set of skills that will lead to teacher self-reflection, growth, and improved student achievement. It was through this study that the question of what additional professional development principals report that they need to become effective evaluators of teachers was explored. This study was designed to investigate the following research questions:

1. In what ways do principals report that they have been prepared to be an effective evaluator?
2. What is the relationship between reported professional development and perceived readiness to conduct effective teacher evaluations?
3. Is there a difference in perceived readiness to conduct effective teacher evaluations between principals at different building levels, specifically those who are now required to conduct evaluations under Race To The Top legislation versus those who have not yet been expected to do so?

4. What additional professional development do principals report that they need to become effective evaluators of teachers?

Additionally, relationships between principal readiness and professional development to conduct effective teacher evaluations that lead to improved student achievement were explored through principals' perceptions of effectiveness and frequency. Correlations of gender, years of experience as a building principal, and grade level supervision were also conducted.

Data was collected through an Internet-accessed, cross sectional, self administered survey using scaled items so that data about respondents characteristics, experiences, knowledge, and opinions could be collected. The survey tool was designed to collect principals' perceptions of readiness to conduct effective evaluations and effectiveness of resources regarding teacher evaluation preparation.

A convenience sample was used in this study, which consisted of 279 building principals who worked within 11 counties in Upstate NY. The region included 78 public school districts, which consisted of 290 schools in total. A total of 115 building principals responded to the survey tool, representing a response rate of 42%. Using SPSS19, data was analyzed by using frequencies, chi-square, and Spearman's rho to determine statistically significant patterns and correlations.

This chapter consists of three sections: Summary of findings, recommendations for practice, and recommendations for future research on the topic.

**Summary of Findings**

**Research question 1:** *In what ways do principals report that they have been prepared to be an effective evaluator?*

One can imply from the findings of this study that frequency of training does not impact the principals' perceptions of readiness, since the data from this study reveal that few people get regular or ongoing training, yet 75% of the respondents indicated some degree of readiness to conduct effective evaluations of teachers. This finding supports the literature, which reports that previous studies show that principals rate their skills in executing the teacher evaluation process very high (Painter, 2000; Poston & Manatt, 1993). The literature suggests that the strategies in which principals can build their skills to be effective evaluators of teachers can be categorized into four methods: education, relationships, on-the-job experience, and training.

What is worthy of noting is that 60 out of 108 respondents reported that their current district had not offered administrators any teacher evaluation trainings at all. This is alarming since the research shows that the majority of teacher evaluations do not accurately depict the effectiveness of the teacher (Halverson et al., 2004; Jacob & Lefgren, 2006; Marzano et al., 2011; Medley & Coker, 1987; Peterson, 2004; Poston & Manatt, 1993). Painter (2000) states: "Indeed, much of the published literature on the principal's role in evaluation consists of explanations of procedural and legal aspects of evaluation and calls for better training of principals" (p. 2). Ongoing training and supports will be essential if building principals are to enhance their skills as evaluators in order to effectively implement the new evaluation system. The data from this study clearly show that there is a need for better training, starting with the administrative programs that prepare and certify administrators, the mandated trainings offered through RTTT, and the supports and ongoing trainings that are afforded to building principals at the district level.

Fifty-one percent of the respondents said that on-the-job experience was very effective in their preparation to conduct teacher evaluations, while 98% of the respondents said that it had

some degree of effectiveness. Development through job experiences allows individuals to learn, grow, and undergo personal change as a result of the roles, responsibilities, and tasks that they encounter in their jobs (MCCauley & Brutus, 1998). This raises the question of who is monitoring them and providing feedback? If building principals have this perception, then it is essential that system leaders design a procedure to ensure that building principals receive effective feedback. McCauley and Brutus (1998) from the Center for Creative Leadership wrote that "the key emphasis in management development systems should be on learning through job assignments" (p. 6). This is highlighted in the findings of this study, which show that 56% of the respondents reported that there were no in district trainings offered, and 35% said that they received no feedback. If building principals are not receiving training or feedback, how can systems leaders ensure that their building principals are proficient in evaluating teachers effectively? Although it was not a statistically significant finding, it is of interest to note that 0% of the females reported being dissatisfied with the feedback that they had received compared to 9% of the males who reported dissatisfaction with the feedback that they had received regarding the teacher evaluations that they had written. The findings from this study are consistent with the research findings of McCauley and Brutus (1998), who wrote, "It came as no surprise to the task force that about eighty percent of learning came from contact with key people in the workplace and from on-the-job experiences" (p. 13).

The findings in this study indicate that programs that are mandated for administrators to attend, such as administrative certification programs and RTTT trainings, were rated most frequently as having some degree of ineffectiveness. These findings are also similar to those reported by the Center for Creative Leadership, which "found that managers saw most of their learning occurring from the challenges encountered in their jobs and from influential people in

their work settings (such as bosses, mentors, and role models) - formal training programs and nonwork experiences contributed to a less extent" (McCauley & Brutus, 1998, p. 4). The data reveal that those resources that the respondents have little control over such as trainings, preparation programs, and supervisors are the resources that respondents were more likely to indicate had some level of ineffectiveness. Using descriptive statistics, it was revealed that 39% of the respondents indicated some level of ineffectiveness for the mandated RTTT trainings. Since this is the primary means of training in order to prepare administrators to implement the new evaluation system, RTTT facilitators need to determine the best training methods for their region in order to make them more effective.

**Research question 2:** *What is the relationship between reported professional development and perceived readiness to conduct effective teacher evaluations?*

The implication from this research is that the data can be used by the three BOCES regions to design differentiated training opportunities based on the prior experience of the building principals.

The findings of this study indicated that training offered to a new principal might be beneficial, but when the same training is offered to an experienced principal, it may be redundant or not offered at a level of depth to be beneficial. Those respondents with 11 or more years of experience looked different from their less experienced colleagues when looking at the data descriptively, as they report a lower agreement with feeling prepared to effectively evaluate teachers. Although no statistical significance was found, veteran building principals, those with 11 or more years of experience, were feeling less prepared, which calls for tiered professional development or trainings and workshops that need to be designed to target different audiences.

When looking at the resources that each group identified as being effective resources for professional development those respondents with 1-10 years experience indicated that they found formal workshops and in-service trainings as effective, whereas those respondents with 11 or more years of experience preferred activities that were more informal and independent. Both groups indicated that collegial discussions, books, and journal articles were good effective resources.

The findings of this study also showed that the relationship between gender and preparedness was statistically significant. The data from this study reveal that men have higher rates of agreement that they felt prepared to be effective evaluators of teachers when compared to females. When looking at gender and evaluation, there are studies that indicate that there are differences in how women and men evaluate the job performance of those with whom they work. The literature reports that women and men may value different characteristics, and thus may concentrate on a different set of criteria than do men (Shakeshaft, Nowell, & Perry, 2007).

> We did find some differences in the things that women and men focused on. Women were more likely than men to encourage the empowerment of their teachers, establish instructional priorities, attend to the social and emotional development of the students, focus on student relationships, attend to the feelings of teachers, include more "facts" in the evaluation, look for teachers' effects on the lives of children, emphasize the technical skills of teaching, comment on the content and quality of the educational program, provide information gathered from other sources, involve the teacher in decision making, issue directives for improvement, provide immediate feedback on performance, and emphasize curricular programs. Men, on the other hand, were more likely than women to emphasize organizational structure and to avoid conflict. Thus we found that the

> evaluations of teachers written by female principals focused on more items, and particularly more items concerned with teaching and learning. (Shakeshaft et al., 2007, p. 346)

**Research question 3:** *Is there a difference in perceived readiness to conduct effective teacher evaluations between principals at different building levels, specifically those who are now required to conduct evaluations under Race to The Top legislation, versus those who have not yet been expected to do so?*

Descriptive statistics had to be used to answer this research question, since the criteria to analyze using chi-square was not met. In order to determine if there was a statistical significance, a minimum count per cell had to be met and four of the cells did not meet this requirement. A review of the response data showed that the various grades that participants reported that they supervised varied greatly and that one-third of the participants did not answer the question therefore making any cross tabulation difficult to perform. It is recommended in future studies that a larger sample size be used in order to compare the two variables, although the question will have to be modified since all teachers, regardless of grade levels, now fall under the new RTTT evaluation guidelines.

The data revealed that there was a higher percentage of respondents who are not yet implementing that indicated some degree of readiness to conduct effective teacher evaluations compared to those respondents who were currently implementing. This finding is not surprising, since there is a difference between preparation and implementation. When one prepares, they build a skill set as a method to get ready, which can increase ones confidence, whereas implementation is where the rubber meets the road and unforeseen challenges may occur. It is during the implementation phase that one can assess their strengths and weaknesses.

The data used to analyze this question may have been skewed because some districts ran pilot programs, which required implementation from all of the building principals, including those supervising grades K-2 and 9-12, who were not required to implement under RTTT, at the time of this study. When the survey was created, pilot programs were not taken into consideration, and therefore when composing the variables of implementing and not implementing, only the guidelines under RTTT were used. Since this topic is very current and many districts were still in the negotiation phase during the development of the survey, variables such as pilot programs were not considered.

The data also indicated that the closer the respondents were to the implementation date, the more they strongly agreed with the statement about being prepared to conduct effective teacher evaluations. What was of interest is that six of the respondents reported that they had not been informed as to when they would be mandated to implement the new evaluation system as articulated by RTTT. Additionally, the data showed that 33% of those who had not been informed of an implementation date strongly agreed with the statement that they were prepared to conduct effective evaluations of teachers.

**Research question 4:** *What additional professional development do principals report that they need to become effective evaluators of teachers?*

A small number of participants (n = 38) responded to the final question of the survey asking them what other professional development would help them to become more effective teacher evaluators (see Appendix B). This can be interpreted in one of several ways: 1) that there were no additional forms of professional development that were not already identified on the survey, 2) that the participants became fatigued by the time they came to question number 24 and did not have the desire to respond in narrative form, and/or 3) that the participants did not

feel comfortable answering an open ended question that solicited them to answer in their own words.

In the future this could be avoided by designing a similar question using skip logic. The initial question could be: *Is there any other professional development that would help you to become a more effective teacher evaluator?* with the choice of responses being *yes* and *no*. For the respondents who selected *no* skip logic could be used and have them go onto the next question. For the respondents who selected *yes* the next question could have been a follow up, using an open ended question format, asking them to identify the professional development that would help them to become more effective teacher evaluators. Asking the question through the use of skip logic could help identify why only a small number of participants responded to the question.

The responses to this question could have been affected by the matrix design that was used. A question using a matrix format can result in participants responding in a particular way or pattern rather than answering the questions based on content. Instead of considering each question or element individually before responding the participant may view all of the elements as being analogous and therefore give similar responses to each one (Babbie, 1998). Due to this question having fifteen resources that the participants were asked to rate, a matrix design appeared to be the most efficient and appropriate method.

Of interest when analyzing the response data from the final question is that 47% of the respondents had mentioned collegial discussion as a form of professional development when the researcher looked at all of the data labels before they were coded and categorized. This shows that collegial discussions that are specific to the instrument and real examples are an important

form of professional development that principals have identified as needing in order to become effective teacher evaluators.

When looking at how the respondents rated the effectiveness of resources in supporting their development to become effective evaluator of teachers, it was interesting to note that, although conducting tandem evaluations with a colleague received the highest percentage of respondents rating it "very effective" (54%), of all of the resources listed it was rated seventh in its overall effectiveness (84%) by the respondents. It is also interesting to note that respondents rated participating in a review and discussion of evaluations written by other colleagues as having a high degree of effectiveness (93%), as it was ranked the second most effective resource, but when asked about having their own evaluations reviewed by peers or colleagues, the degree of effectiveness decreased (86%) ranking it as the sixth resource having some degree of effectiveness.

When comparing the effectiveness of resources, it is worth noting that, although the respondents rated participating in a review and discussion of evaluations written by other colleagues (93%), having peers review their written evaluations (86%), and conducting tandem evaluations with a colleague (84%) as having some degree of effectiveness, the frequency that the respondents reported some degree of use was 21% or less. Although the respondents had stated that there was a degree of effectiveness with these resources in their development as effective teacher evaluators, they are not utilizing them.

Comparing the respondents' perceptions of effectiveness to their reported rate of use raises further questions due to the incongruities. Is the reported difference between effectiveness and use due to the level of participation required? When using books and journal articles as a resource, the respondents have a greater degree of flexibility as to when that can be done, and

there is no required reflection of work or follow up, whereas participating in a role play, taking courses, or participating in a tandem observation requires more structure, higher expectations, and greater participant involvement. The frequent use of books as a resource is also an indication that many building principals view teacher evaluation as an individual versus a collective process.

The frequency of use by the respondents could also be affected by accessibility. Building principals have greater access to books and journal articles then they have to evaluation exemplars and the time to participate in tandem observations or collegial and peer reviews. The frequency of use could also be potentially affected by whether a resource is sponsored (paid for) by the district or is the financial obligation of the respondent. Books and journal articles are economical to building principals and can often be obtained through professional organizations, discounted through online purchase, or obtained free through online databases. Districts are often willing to pay for building principals to attend job related conferences and workshops.

Through RTTT, building principals are required to attend several workshops aimed at conducting effective teacher evaluations in order to become a certified evaluator, so one might have predicted that the frequency of use would have been rated higher than 62% or 67%. The mandated training might have also contributed to response bias by having the respondents' rate the use of conferences higher than in the past when there was no mandate for building level administrators to become certified evaluators of teachers.

It is through the data in Table 26 that it becomes clear that, although the respondents know what resources are very effective, they are not utilizing them. Table 26 lists the top four resources identified as *very effective* by the respondents and shows frequency of use.

Table 26

*Top Four Resources Identified as Very Effective and the Frequency of Use by Respondents.*

| | Very Effective | | Often Used | |
|---|---|---|---|---|
| Resource | n | % | n | % |
| Tandem evaluations | 30 | 54% | 3 | 3% |
| Peer review | 23 | 38% | 5 | 5% |
| Review of colleague evaluations | 24 | 38% | 2 | 2% |
| Review of exemplar evaluations | 19 | 26% | 3 | 3% |

**Other findings.** A finding that emerged, when analyzing the data for the open-ended question about any additional professional development that respondents felt would help them in becoming more effective teacher evaluators, was that 11% of the respondents mentioned time. Although not a form of professional development time was an emerging theme when the data was coded and categorized. What is of particular interest is how the respondents defined time in their responses. Half of them wanted more hours in the day in order to accomplish all of their responsibilities, including teacher evaluations, whereas the other half wanted more time to practice the new evaluation process in order to become more efficient and effective in its use. Petrie (2011) from the Center for Creative Leadership wrote, "According to social psychologists, people's motivation to grow is highest when they feel a sense of autonomy over their own development" (p. 17).

**Recommendations for Practice.**

Based on the findings of this study the following recommendations are provided:

**Recommendation 1:** *Principal preparation programs need to strengthen their offerings by differentiating their classes in order to make them more effective.*

It is recommended that principal preparation programs strengthen their offerings by examining the courses required in the program by using data from the intake process. It is clear from this study that those going into administration have varying backgrounds and experience. Ninety-three percent of the respondents had previously worked as a classroom teacher, while 60% had experience as an assistant principal before becoming a building principal. The respondents rated these experiences as effective resources. Therefore it is recommended that preparatory programs offer differentiated course offerings to those with and without these experiences. It is further recommended that the programs preparing individuals to become building principals design classes knowing where the gaps are to help strengthen their programs and make them more effective.

The competencies that will be most valuable to future leaders appear to be changing. Preparation programs that focus on content to prepare leaders for the 21st century have become dated and redundant. "While these were relatively effective for the needs and challenges of the last century, they are becoming increasingly mismatched against the challenges leaders currently face" (Petrie, 2011, p. 10). Preparation programs need to be designed in a way that causes individuals to take ownership for their learning and development. As programs look at redesigning themselves to meet the developmental needs of future leaders, they have to transition from horizontal learning, in which competencies are transmitted from an expert, to vertical learning, where students have to take control of their learning and make sense of the world in more complex and inclusive ways. Courses need to be designed in a way that forces the learner to question their assumptions, puts them in uncomfortable situations, and involves self-reflection and identification of a sense of importance to the topic.

> Many leadership programs operate on the assumption that if you show people how to lead, they can do that. However, the most difficult challenges that people face in their work lives are often associated with the limitations of the way they "make meaning" at their current level of development. When a person surfaces the assumptions they have about the way the world works, they get the chance to question those assumptions and allow themselves the opportunity to start to make meaning from a more advanced level. (Petrie, 2011, p. 16).

Preparatory programs need to use the four trends for the future of leadership development as identified by the Center for Creative Leadership as guidelines when designing their classes and program requirements: more focus on vertical development, transfer of greater developmental ownership to the individual, greater focus on collective rather than individual leadership, and much greater focus on innovation in leadership (Petrie, 2011). Preparation programs need an overhaul as the environments that today's principals must lead in have changed.

**Recommendation 2:** *RTTT coordinators and system leaders need to offer differentiated professional development and resources to building principals.*

When differentiating and designing workshops for building principals regarding evaluating teachers effectively, RTTT coordinators and facilitators should also get the experience level, grade levels of supervision, and gender of the participants prior to the workshop in order to make it more meaningful to the participants. The type of instruction or guidance that a new administrator will need is much different than that of an administrator who has several years of experience. Gender differences, as well as the grade levels that the participant supervises, may

also affect the type of training that should be offered. Therefore, RTTT trainings should be differentiated based on prior experience and not a canned approach for all principals.

The results of this study indicate that *one size fits all* trainings are not effective, as there is a difference in what building principals with 1-10 years of experience report as effective resources for professional development compared to those reported by building principals with 11 or more years of experience.

The respondents to this survey who had 11 or more years of experience showed through their responses that the resources that they find most effective are ones that can be done independently (journal articles, books, webinars). They also indicated that they find collegial discussions effective. Based on these findings a small professional learning community may work best for them versus a large workshop or conference. It is recommended that RTTT facilitators and system leaders help pull those resources together and coordinate a professional learning community that will allow these administrators to discuss what is working and not working through round table discussions. It will be important to have a facilitator who can answer questions, allowing building principals to bring what they learned back to their districts. A small professional learning community will meet the needs of these administrators by allowing them to participate in independent work, but at the same time, it also allows for follow up in a format that they report is effective.

The results of this study indicate that building principals, regardless of years of experience, found collegial discussions, books, and journal articles as good effective resources for professional development. It is recommended that system leaders incorporate these into any district wide professional development activities that are aimed at increasing the skills of their building principals to become effective evaluators of teachers. The overall perception of the

effectiveness of these resources and the frequency of their use can also be viewed as an indication that building principals want on-going professional development and have facilitated their own professional development through the use of resources that are convenient and economical. Professional development that is general and front-loaded is not a valuable method in supporting building principals in becoming more effective teacher evaluators.

System leaders need to identify emerging leaders and provide appropriate on-the-job experiences that will help build or strengthen their skills necessary to become effective leaders in a complex organization. It will be important for system leaders to be aware of the roles within their system that have higher developmental potential, so that they can use those assignments to help emerging leaders grow by developing the skills and complex frameworks needed for making decisions and guiding actions.

> Sternberg and his colleagues developed the notion of tacit knowledge - knowledge that is developed through direct experience as one moves from a novice to an expert in a given field or domain - as a type of knowledge distinct from that acquired through formal education. (McCauley & Brutus, 1998, p. 5)

To effectively help emerging leaders grow through their job experiences, it is important to put them in new and unfamiliar situations. It is essential that system leaders recognize the power of non-training development opportunities for aspiring principals and future system leaders. Growth occurs when routines are disrupted, situations call for new tactics and behaviors, and the emerging leader has to successfully complete the task (McCauley & Brutus, 1998).

**Recommendation 3:** *Race to the Top Coordinators need to collaborate with system leaders, building principals, and state agencies to design workshops that support the needs of their specific region.*

Based on the responses from this study, it is recommended that RTTT coordinators and facilitators survey the building principals, using a design that is effective and efficient so as to not limit the number of participants, within their respective region to determine what their specific concerns are and then design workshops around those specific needs. Instead of having general front loaded informational sessions, workshops should be tailored to the needs of the participants.

The formal trainings for the region could be information that is front loaded and focused on the three elements (process, forms, and criteria), where only one-half of the respondents were satisfied with the training that they received in these areas. Ongoing training sessions, through the use of collegial discussions, might be targeted in the use of the rating scale (HEDI), evaluating evidence of family and community outreach, legal issues regarding evaluations, and validity and reliability. These were the areas in which respondents expressed greatest weakness.

RTTT coordinators and systems leaders need to account for this and help facilitate by providing the time, space, and the target resources on key topics based on where the region is in the implementation of the new evaluation system. The recommended training might start with Student Learning Objectives, using a data analyst as a resource, since that is what districts are currently responsible for developing. In a month or two, the trainings might change and focus on the use of the rubrics that districts within the region have adopted. The mid-year training focus might be on observer bias and inter-rater reliability as building principals begin their multiple observations, and at the end of the year, the trainings could focus on scoring.

Figure 5 illustrates a possible district internal training schedule that aligns the topic for trainings with the steps in the evaluation process as well as the school calendar. The proposed model suggests a frontloaded introduction of the process to be used, the criteria used for

evaluative purposes, and a chance to become familiar with the forms to be used. After the initial introduction, training should be ongoing and aligned to what principals will be required to do during the school year.

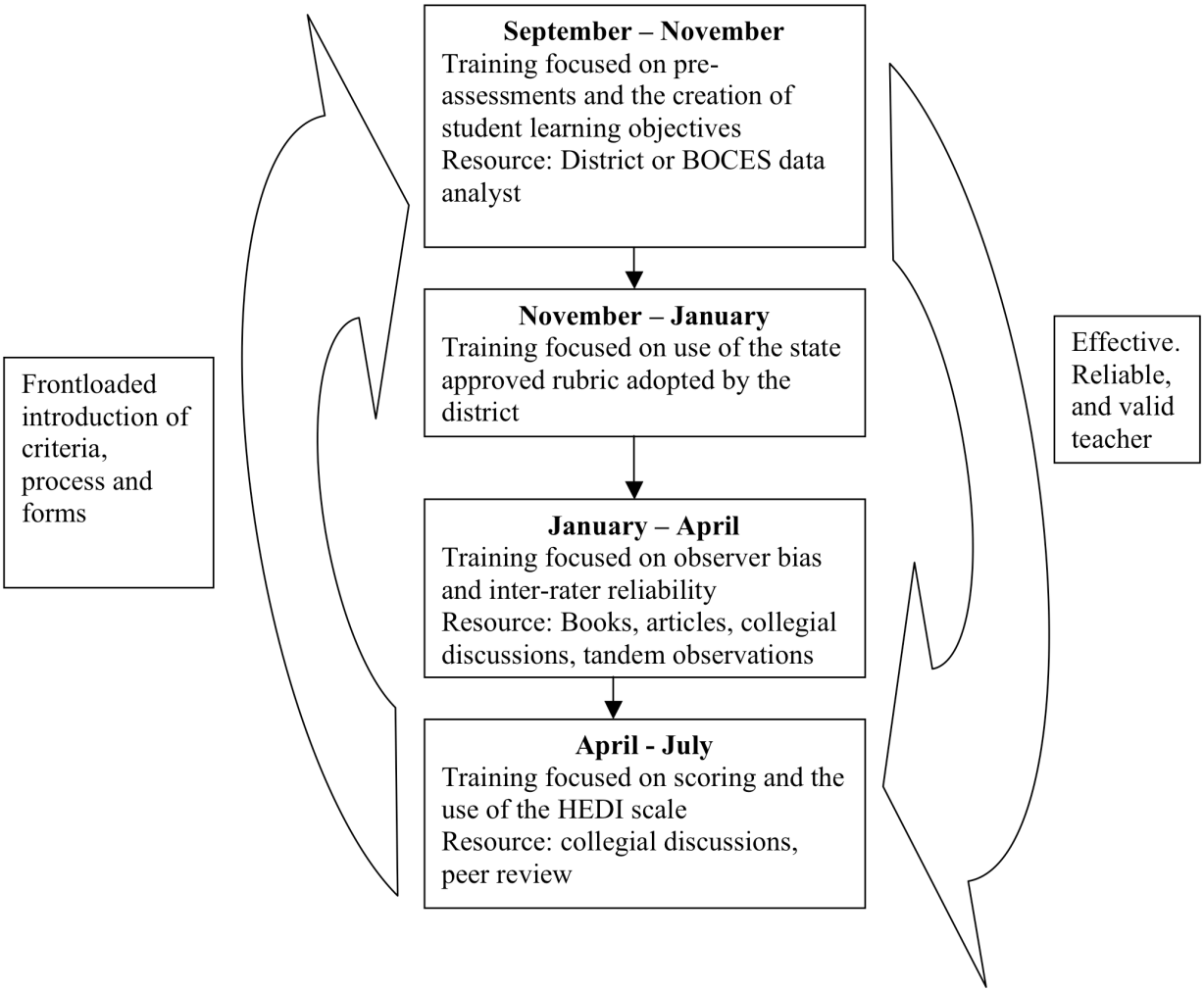


*Figure 5*. Possible district internal training schedule.

**Recommendation 4:** *Regional or building wide focus groups that focus on effective teacher evaluation should be created.*

As system leaders, it is important for superintendents to create learning communities for their administrators, specifically building principals, that focus on their development as effective teacher evaluators. The training for teacher evaluation can not only consist of front loaded

training, where the focus is on the process, but must be ongoing where evaluations done by principals are reviewed and specific feedback is given leading to skill enhancement.

When establishing these learning communities, it will be important to establish group norms that facilitate collaboration, peer feedback, active participation, analysis, and reflection regarding the evaluations that are written by each group member. System leaders should establish a set schedule in order for this group to meet, so that it remains a priority and does not dissipate due to busy schedules, other tasks, and responsibilities. The training must be ongoing to create sustainability and reinforce the skills needed to become effective evaluators of teachers.

Respondents rated the effectiveness of both books and journal articles high in their preparation as evaluators of teachers. Therefore, it is recommended that system leaders use this resource, which was often done in isolation, to facilitate book studies within their administrative teams. It might be worthwhile for the system leader to purchase highly recommended book titles for their administrators to read in order to discuss at regularly scheduled administrative meetings. System leaders can also take the same approach by distributing copies of a journal article for the administrators to read in order to be able to discuss at the next administrative meeting. This approach can help system leaders create a transitional step by taking an activity that principals previously did in isolation and making it a collective process.

**Recommendation 5:** *Districts need to create policy regarding internal procedures on teacher evaluation that measures the accuracy of written evaluations and provides feedback to the evaluators.*

When reauthorizing legislation involving teacher evaluation, the design of internal procedures should be considered. All of the mandated trainings focus on the process, but there is little to encourage or help districts set up internal procedures that will provide building principals

with feedback, guidance, or evaluation of the validity and reliability of the evaluations that they have written. School districts have to assume responsibility of providing ongoing supports and training as well as developing procedures that help establish a culture leading to sustained transformational change in teacher evaluations. Just as principals need to provide effective feedback to teachers that consist of descriptive data, characteristics of effective teaching and reflective inquiry, and self-directedness to promote professional growth, so too must principals if they are expected to grow and enhance their skills.

It will be the responsibility of individual school districts to create systems that will provide feedback and guidance to building principals as they continue to become familiar with the new evaluation process and hone their skills to become effective evaluators of teachers. It is recommended, through the findings of this study, that districts create a process that allows for inter-rater reliability through collegial discussions, tandem evaluations, peer review, and other resources that building principals have identified as effective resources.

**Recommendations For Future Research**

**Recommendation 1:** *Further study of relationships between principals' readiness and professional development to conduct effective teacher evaluations that lead to improved student achievement.*

For much of the data analyses, no statistical significance was found, although patterns of interest were discovered when examining the descriptive frequencies. This could be due to the fact that there was not a large enough sample to meet the criteria necessary to answer the question of statistical significance through the use of inferential statistics. Due to the patterns that were discovered in this study, it is recommended that further studies try to obtain a larger sample size that will allow tests of statistical significance to be conducted to determine the

strength of correlations and patterns between the variables used in this study. Since this study was conducted within a three-region area, it might be more informative if future studies expanded to include more of the 37 BOCES regions within NYS or used all building principals in NYS as their target population. When reviewing at the level of preparedness, as reported by the respondents, and their perceptions of the effectiveness of resources and supports designed as forms of professional development to help them become effective evaluators of teachers, the data showed evidence of a relationship, but it was not strong. It is recommended that future studies continue investigating this potential relationship by designing a qualitative research approach that utilizes a focus group, so more data related to the topic can be collected.

It was of interest to note that in this study, respondents with 11 or more years of experience indicated that they felt less prepared to conduct effective evaluations of teachers. It is recommended that a future qualitative study be conducted with this target group as to why they feel inadequately prepared. The findings from this study indicate that future studies should more closely examine the resources needed by principals as they go through the phases of development (entry, skill-building, performance, and mastery).

The findings of this study also revealed that the relationship between gender and perceptions of preparedness was statistically significant. This is supported in the literature, but future studies could explore why this relationship exists. It is recommended that future research explore why males are more likely to feel prepared to be effective evaluators of teachers than their female colleagues.

**Recommendation 2:** *An examination of district policies regarding the evaluation process for principals.*

School districts have to assume responsibility for providing ongoing supports and training, as well as developing procedures that help establish a culture leading to sustained transformational change in teacher evaluations that lead to teacher growth and improved student achievement. Just as principals need to provide effective feedback to teachers that consist of descriptive data, characteristics of effective teaching, reflective inquiry, and self-directedness to promote professional growth, so too must principals receive similar types of effective feedback if they are expected to grow and enhance their skills. In order for transformational change to occur, the procedures and policies within districts must also change. A future exploratory study could examine the procedures that districts use to provide feedback to their evaluators and the effectiveness of those procedures. Future studies could also compare the internal procedures of schools that are identified as highly effective versus those in need of improvement. A quantitative study examining district board policies or administrative regulations that facilitate and guide accountability, feedback, and ongoing supports for building principals that are aimed at increasing their skills as effective teacher evaluators would be another recommended study.

**Recommendation 3:** *Further examination of principal preparation and readiness to conduct effective teacher evaluations between principals who are now required to conduct evaluations under Race To The Top legislation versus those who have not yet been expected to do so.*

When examining the correlations between the perceived readiness of the respondents and the year they were expected to implement the new evaluation system as mandated under RTTT legislation, interesting trends emerged. Those who were required to implement at the time of the study felt better prepared compared to those who were expected to implement the following year. The data also showed that there were slightly more respondents who were already implementing

the new evaluation system under RTTT who strongly agreed in their readiness to conduct teacher evaluations compared to those respondents who are not yet implementing. Future research could further examine what created these results, as well as having participants identify what resources they felt were most effective in their preparation. Examination of principal preparation and readiness to conduct effective teacher evaluations based on the grade levels they supervise was not possible in this study due to the large variability in the grade configurations that respondents indicated that they supervised.

Since all districts within NYS are now mandated to implement the new evaluation system, future studies may examine this research question by comparing states. There are 37 states that currently have legislation pertaining to teacher evaluation. Some of those states have already implemented the new teacher evaluation system as designed under RTTT, while others are just beginning to implement. Future studies could use the research questions posed in this study to examine principal readiness by comparing states who are implementing the new evaluation system under RTTT versus those that have chosen not to or by comparing states based on the number of years that they have been implementing the new system.

It will be important to continue to examine teacher evaluation and principal preparedness, as the stakes are higher than ever for all of the stakeholders involved. With any new initiative comes anxiety and uncertainty, and many have already expressed concerns over the potential outcomes. When looking at evaluation validity, Kimball and Milanowski (2009) caution that teachers could receive consequences that are not justified. They argue that "differences in validity across principals are clearly problematic as stakes are raised, and the decision making of those with less valid ratings needs to be improved" (p. 36). The concern regarding the lack of

preparation for principals is highlighted by states and districts, such as Tennessee and Washington D.C., who have already implemented more rigorous evaluation systems.

In the Washington D.C. School district, under an evaluation system titled IMPACT, the "chief assessment tool is five 30-minute observations by administrators and master educators of teachers each year as they work in the classroom. That's a total of 2 ½ hours a year of observation" (Strauss, 2011, p. 2). According to an article in the *Washington Post,* in July 2011, 206 teachers, 5% of total teachers in the district, were fired because they were rated as ineffective (Turque, 2011).

In the state of Tennessee, which was one of the first two states to be selected for the RTTT grants, the new evaluation system titled Teacher Education Acceleration Model (TEAM) continues to be under scrutiny. It is reported by the State Collaborative on Reforming Education (SCORE) that 29% of the teachers in Tennessee feel that the new system would not have a positive impact on their teaching ("SCORE releases report," 2012). Whereas the old evaluation system was based on observations and teacher self-reflection, the new system calls for multiple measures including at least four observations, student achievement, and growth. Wording in the Tennessee State law calls for schools to use test data or something *comparable* to test data in the evaluation of teachers, which would account for 50% of their performance. The problem is that over half of the teachers in the state of Tennessee do not have an end of the year state test and are therefore being evaluated on district-wide averages (Farmer, 2012). The Tennessee Department of Education (TDOE) (2012) stated that after the first year of implementation, one in six teachers falls significantly short of expectations in advancing student learning. It was also noted that the observation scores that these teachers received were not aligned with their scores on student achievement growth. This led the TDOE (2012) to believe that districts were ignoring their most

struggling teachers and not providing the appropriate feedback that educators need to improve their performance and, ultimately, student outcomes. The first recommendation by SCORE, which was established to review the newly implemented evaluation system, called for improved evaluator training ("SCORE releases report," 2012).

**Recommendation 4**: *A further study of the relationships between the training of evaluators in a business model to that of building principals in schools in need of improvement and those deemed highly effective.*

As education reform takes place with a focus on teacher evaluation, it is recommended that future studies examine the correlations between the training and preparedness of evaluators in education compared to those in the business industry. The business industry has used evaluations as a way to measure employee performance, which often results in whether or not it earns a profit at the end of the fiscal year. For many years, companies such as Anheuser-Busch have invested in studies that look at performance appraisal (Smith, 1986). With student achievement and outcomes becoming the focus of teacher evaluations, it might be of interest to compare this new system to that of the business industry, where outcomes have also been the driving force behind employee performance appraisal. Future studies could look at the training programs between both and help to identify those resources or supports that are identified as being effective. Future studies could also examine and compare the training models that both use to prepare and support those who conduct evaluations.

**Recommendation 5:** *A further study of relationships between written performance reviews and the point of view and value terms used.*

Based on the literature and the findings in this study, many feel that evaluations do not provide adequate feedback that leads to increased performance. Since the new evaluation system

under RTTT is aimed at increasing teacher performance, future studies could assess the quality of feedback that teachers receive. It is suggested that a qualitative study be conducted to look at teacher evaluations written under the new legislation that focuses on the role of the evaluator as well as the value of the written feedback.

Research questions could focus on examining if the value terms used are general with no descriptive meaning, which negatively impacts professional development, or if they are specific and descriptive, denoting a certain set of characteristics. Do principals use factual or evaluative assertions in their written narratives? The use of evaluative assertions such as biased, good, effective, and/or desirable could prejudice the evaluation. Further research could also look at the conferencing language used by principals to see how it correlates to the written language on the evaluation document. It is clear that meaningful and effective feedback that leads to professional growth is missing in the evaluation process.

The research by Johnson (1997) also shows that the role that the evaluator plays within the organization impacts the validity of an evaluation. Future studies could study the validity of evaluations conducted by system leaders, such as superintendents, compared to those conducted by building principals. With the new evaluation system calling for multiple observations, many districts are employing the assistance of all administrators, regardless of the distance of their role from the organization's core of working with children. Future studies could look at the impact that one's role has on effective feedback under the new evaluation system. What would be of interest is to examine the criteria used when doing observations and the focus of the feedback. For example, would the feedback from a building principal differ in focus from that of a central office administrator?

Due to the anxiety that teachers feel during the post-conference of the evaluation process, much of the feedback that is given may be lost, forgotten, or even misinterpreted due to emotions. Therefore, written feedback will be imperative in allowing teachers the chance to once again reflect on what was verbally communicated to them. In order to help teachers grow and become *highly effective*, the quality of feedback will be essential and therefore should continue to be the focus of future studies.

**Closing Statement**

The conclusions and recommendations offered through this study are intended to lead to conditions that will support principals as evaluators of teachers and increase student achievement. There is a transformation in teacher evaluations that is currently starting to take place. The purpose of evaluations is no longer to document teachers' competence but to measure teachers' effectiveness and help them improve over time. The expectation is that all teachers will increase their expertise from year to year, which produces gains in student achievement from year to year with a powerful cumulative effect.

As teachers develop and grow based on research-based practices, the tools used to help them develop professionally should also change. Teacher evaluation should not become status quo as other aspects of education refine and change.

Ongoing support will be needed by principals to strengthen their skills as evaluators under the new system. As this study indicates, changes in procedures should include ongoing training and feedback for those responsible for implementing the process. Training programs need to be tailored to the needs of the population, and general front loaded sessions have not been judged to be an effective method of preparation. Many of the respondents in this study found that their on-the-job experience best prepared them to evaluate teachers. If this is the case,

how can evaluators know if they are effective if they receive no feedback or have nothing to measure themselves against? Most importantly, how can administrators rate themselves as effective evaluators of teachers when those they are evaluating feel that the process is meaningless and does not lead to growth.

Looking at the literature and the findings of this study, it is evident that ongoing training needs to occur if the new evaluation system is going to result in the desired change. It will be up to NYSED, system leaders, and RTTT coordinators to consider the findings of this study and to develop policies, regulations, and trainings that establish structures that offer support as well as meet the needs of their administrators as they strive to be effective teacher evaluators. The research is clear that the best way to increase student achievement is by having highly effective teachers in the classroom. "Improving supervision and evaluation will improve the education kids are getting and that's why the school system exists in the first place" (Langlois & Colarusso, 1988, p. 33). Principals cannot help teachers grow if they are not proficient in using one of the primary tools aimed at assisting them and students may not achieve if they do not have *highly effective* teachers instructing them every day.

References

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

Babbie, E. R. (1998). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth.

Boards of Cooperative Educational Services of New York State. (2012). *Welcome page*. Retrieved from http://www.boces.org

Brimley, V., & Garfield, R. R. (2008). *Financing education in a climate of change* (10th ed.). Boston, MA: Pearson, Allyn, and Bacon.

Catano, N., & Stronge, J. H. (2006). What are principals expected to do? Congruence between principal evaluation and performance standards. *NASSP Bulletin, 90*(3), 221-236. doi:10.1177/0192636506292211

Clifford, M., Behrstock-Sherratt, E., & Fetters, J. (2012, May). *The ripple effect: A synthesis of research on principal influence to inform performance evaluation design*. Retrieved from American Institutes for Research website: http://www.air.org/files/1707_The_Ripple_Effect_d8_Online.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Common Core State Standards Initiative. (2012). *Homepage*. Retrieved from http://www.corestandards.org/

Cooper, R. K. (2010, November 12). School superintendents wince at local share of Race to Top grants. *The Business Review*. Retrieved from http://www.bizjournals.com/albany/

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods* (3rd ed.). Los Angeles, CA: Sage.

Cuomo, A. M. (2011, January 5*). State of the state address* [Transcript]. Retrieved from the New York State Governor's Office website: http://www.governor.ny.gov/sl2/stateofthestate2011transcript

Danielson, C. (2001). New trends in teacher evaluation. *Educational Leadership, 58*(5), 12-15. Retrieved from http://www.ascd.org/publications/educational-leadership.aspx

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

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

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1). Retrieved from http://epaa.asu.edu/ojs/

Dyer, K. M., & Renn, M. T. (2010). Getting an education: School leaders need specialized development. *Leadership in Action, 29*(6), 3-7. doi:10.1002/lia.1311

Elmore, R. F. (2000, Winter). *Building a new structure for school leadership*. Washington, DC: Albert Shanker Institute. Retrieved from http://www.ashankerinst.org/Downloads/building.pdf

Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Washington, DC: Albert Shanker Institute. Retrieved from http://www.ashankerinst.org/Downloads/Bridging_Gap.pdf

Erlandson, D. A. (1986). NASSP's Consortium for the Performance-Based Preparation of Principals: An update. *NASSP Bulletin, 70*(493), 70-71. doi:10.1177/019263658607049312

Farmer, B. (2012, June 11). Non-tested subjects pinpointed as evaluation trouble spot. *WPLN: Nashville Public Radio*. Retrieved from http://wpln.org/?p=37943

Feeney, E. J. (2007). Quality feedback: The essential ingredient for teacher success. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas, 80*(4), 191-198. doi:10.3200/TCHS.80.4.191-198

Frase, L. E., & Streshly, W. (1994). Lack of accuracy, feedback, and commitment in teacher evaluation. *Journal of Personnel Evaluation in Education, 8*(1), 47-57. doi:10.1007/BF00972709

Freeberg, N. E. (1969). Relevance of rater-ratee acquaintance in the validity and reliability of ratings. *Journal of Applied Psychology, 53*(6), 518-524. doi:10.1037/h0028622

Gimbel, P. A., Lopes, L., & Nolan Greer, E. (2011). Perceptions of the role of the school principal in teacher professional growth. *AASA Journal of Scholarship and Practice, 7*(4), 19-31. Retrieved from http://www.aasa.org/jsp.aspx

Haefele, D. L. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education, 7*(1), 21-31. doi:10.1007/BF00972346

Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In W. K. Hoy & C. G. Miskel (Eds.), *Research and Theory in Educational Administration Series (Vol. 3): Educational administration, policy, and reform: Research and measurement* (pp. 153-188). Greenwich, CT: Information Age.

Hartzell, G. N. (1995). Helping administrators learn to avoid seven common employee performance appraisal errors. *Journal of Staff Development, 16*(2), 32-37. Retrieved from http://www.learningforward.org/publications/jsd

Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology, 73*(1), 68-73. doi:10.1037//0021-9010.73.1.68

Howard, B. B., & McColskey, W. H. (2001, February). Evaluating experienced teachers. *Educational Leadership, 58*(5), 48-51. Retrieved from http://www.ascd.org/publications/educational_leadership.aspx

Huffcutt, A. I., & Woehr, D. J. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*(3), 189-205. Retrieved from http://www.bpsjournals.co.uk/journals/joop/

Hunter, M. C. (1973, February). Appraising teaching performance: One approach. *National Elementary Principal, 52*(5), 60-62. Retrieved from http://www.naesp.org/Publications.aspx

Hunter, M., & Russell, D. (1990). *Mastering coaching and supervision.* Thousand Oaks, CA: Corwin Press.

Huntley, D. W. (2010). *An examination of the extent to which the use of social media by school districts to prepare and communicate the annual school district budget impacts the passage or failure of that budget* (Doctoral Dissertation). Retrieved from http://library.sage.edu

Jacob, B., & Lefgren, L. (2006, Spring). When principals rate teachers: The best -- and the worst -- stand out. *Education Next, 6*(2), 58-64. Retrieved from http://educationnext.org/

Johnson, B. L., Jr. (1997). An organizational analysis of multiple perspectives of effective teaching: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 69-87. Retrieved from http://www.springer.com/education/assessment,+testing+&+evaluation/journal/11092

Kersten, T. A., & Israel, M. S. (2005). Teacher evaluation: Principals' insights and suggestions for improvement. *Planning and Changing, 36*(1-2), 47-67. Retrieved from http://www.planningandchanging.ilstu.edu

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

Kotter, J. P., & Cohen, D. S. (2002). *The heart of change: Real-life stories of how people change their organizations*. Boston, MA: Harvard Business School Press.

Krein, T. J. (1990, May). Performance reviews that rate an 'A'. *Personnel, 67*(5), 38-40. Retrieved from http://ioma.bna.com/Newsletter.aspx?id=2147483662 &libID=2147483662

Langlois, D. E., & Colarusso, M. R. (1988, May). Don't let teacher evaluation become an empty ritual. *Executive Educator, 10*(5), 32-33. Retrieved from ERIC database. (EJ370306)

Lashway, L. (2003). *Transforming principal preparation: ERIC digest* (Digest No. 165). Eugene, OR: ERIC Clearinghouse on Educational Management. Retrieved from ERIC database. (ED473360)

Marshall, K. (2005, June). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan, 86*(10), 727-735. Retrieved from http://www.kappanmagazine.org/

Marshall, K. (2009). *Rethinking teacher supervision and evaluation: How to work smart, build collaboration, and close the achievement gap*. San Francisco, CA: Jossey-Bass.

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

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

McCauley, C. D., & Brutus, S. (1998). *Management development through job experiences: An annotated bibliography*. Greensboro, NC: Center for Creative Leadership. Retrieved from http://www.ccl.org/leadership/pdf/research/MgmtDevelopmentBib.pdf

McGrath, M. J. (2000, October). The human dynamics of personnel evaluation. *School Administrator, 57*(9), 34-39. Retrieved from http://www.aasa.org/SchoolAdministrator.aspx

Measures of Effective Teaching (MET) Project. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from author's website: http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Leadership, 80*(4), 242-247. Retrieved from http://www.tandf.co.uk/journals/titles/00220671.asp

Miller, R. (2000, June). Greater Expectations National Panel: A summary of efforts in school reform since 1983. Retrieved from Association of American Colleges and Universities website: http://www.greaterexpectations.org/briefing_papers/SchoolReform.html

Mitgang, L. (2012). *The making of the principal: Five lessons in leadership training.* New York, NY: Wallace Foundation. Retrieved from http://www.wallacefoundation.org/knowledge-center/school-leadership/effective-principal-leadership/Documents/The-Making-of-the-Principal-Five-Lessons-in-Leadership-Training.pdf

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the Secretary of Education* (Research report). Retrieved from U.S. Department of Education website: http://www2.ed.gov/pubs/NatAtRisk/

Natriello, G., & Dornbusch, S. M. (1980). Pitfalls in the evaluation of teachers by principals. *Administrator's Notebook, 29*(6), 1-4. Retrieved from ERIC database. (EJ261536)

New York State Department of the Budget (2011, November). *Economic, revenue and spending methodologies*. Retrieved from http://www.budget.ny.gov/pubs/supporting/MethodologyBook.pdf

New York State Education Department. (2012, August). Guidance on New York State's annual professional performance review for teachers and principals to implement Education law §3102-c and the Commissioner's Regulations. Retrieved from http://engageny.org/wp-content/uploads/2012/05/APPR-Field-Guidance.pdf

New York State Education Department, Office of Prekindergarten Through Grade 12 Education, Information and Reporting Services. (2011, April 26). *Public school enrollment & staff - 2009-10: Districts and pupils by size and district*. Retrieved from http://www.p12.nysed.gov/irs/statistics/enroll-n-staff/TABLE2.pdf

New York State Education Department, Office of Prekindergarten Through Grade 12 Education, Information and Reporting Services. (2012). *New York State report card database*. Retrieved from https://reportcards.nysed.gov/databasedownload.php

New York State Education Department, Office of Higher Education. (2012, February). *Conceptual approval of schedule for new certification examinations for teachers and school leaders*. Retrieved from http://www.highered.nysed.gov/http/www.regents.nysed.gov/meetings/2012Meetings/February2012/212hed2.pdf

New York State School Boards Association, & New York State Bar Association. (2010). *School law* (33rd ed.). Albany, NY: LexisNexis.

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2008).

Oluwole, J. O. (2009). Tenure and the "highly qualified teacher" requirement. *Whittier Journal of Child and Family Advocacy, 8*(2), 157-190. Retrieved from http://www.law.whittier.edu/index/centers-programs/ccr/whittier-journal-of-child-family-advocacy/

Ostovar Namaghi, S. A. (2010). A data-driven conceptualization of teacher evaluation. *The Qualitative Report, 15*(6), 1504-1522. Retrieved from http://www.nova.edu/ssss/QR/index.html

O'Donovan, E. (2011, July-August). Teacher evaluation is not synonymous with teacher quality: We should focus on coaching and collaboration, not evaluation. *District Administration, 47*(7), 73-74. Retrieved from http://www.districtadministration.com

Painter, S. R. (2000). Principals' efficacy beliefs about teacher evaluation. *Journal of Educational Administration, 38*(4), 368-378. Retrieved from http://www.emeraldinsight.com/info/journals/jea/jea.jsp

Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin, 88*(639), 60-79. doi:10.1177/019263650408863906

Peterson, K. D., & Peterson, C. A. (2006). *Effective teacher evaluation: A guide for principals*. Thousand Oaks, CA: Corwin Press.

Petrie, N. (2011, December). *Future trends in leadership development* [White paper]. Greensboro, NC: Center for Creative Leadership. Retrieved from http://www.ccl.org/leadership/pdf/research/futureTrends.pdf

Poston, W. K., Jr., & Manatt, R. P. (1993). Principals as evaluators: Limiting effects on school reform. *International Journal of Educational Reform, 2*(1), 41-48. Retrieved from https://rowman.com/page/IJER

Reeves, D. B. (2004). *Accountability for learning: How teachers and school leaders can take charge*. Alexandria, VA: Association for Supervision and Curriculum Development.

Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the Black-White achievement gap*. Washington, DC: Economic Policy Institute.

Sawyer, L. (2001, February). Revamping a teacher evaluation system. *Educational Leadership, 58*(5), 44-47. Retrieved from http://www.ascd.org/publications/educational_leadership.aspx

SCORE releases report on new teacher evaluation. (2012, June 12). *Dyersburg State Gazette*. Retrieved from http://www.stategazette.com/story/print/1859046.html

Seashore, C. N., Seashore, E. W., & Weinberg, G. M. (1997). *What did you say?: The art of giving and receiving feedback.* Columbia, MD: Bingham House Books.

Shakeshaft, C., Nowell, I., & Perry, A. (2007). Gender and supervision. In *The Jossey-Bass Reader on Educational Leadership* (2nd. ed., Revised ed.) (pp. 339 -348). San Francisco, CA: Jossey-Bass.

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

Smith, D. E. (1986). Training programs for performance appraisal: A review. *Academy of Management Review, 11*(1), 22-40. doi:10.2307/258329

Smith, E., & Gorard, S. (2007). Improving teacher quality: Lessons from America's No Child Left Behind. *Cambridge Journal of Education, 37*(2), 191-206. doi:10.1080/03057640701372426

Steiner, D. (2009, December 14). *An agenda for education reform in NYS* [Statement]. Retrieved from the New York State Education Department's website: http://www.oms.nysed.gov/press/AgendaforEducationReform.html

Strauss, V. (2011, July 20). By IMPACT standards, Ms. Frizzle would fizzle [Web log post]. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/post/miss-frizzle-fails-impact-evaluation/2011/07/18/gIQA3QyaOI_blog.html

Stronge, J. H. (1991). The dynamics of effective performance evaluation systems in education: Conceptual, human relations, and technical domains. *Journal of Personnel Evaluation in Education, 5*(1), 77-83. Retrieved from ERIC database. (EJ435167)

Tanner, L. (2008). No Child Left Behind is just the tip of the iceberg. *Social Studies, 99*(1), 41-48. Retrieved from ERIC database. (EJ787821)

Tennessee Department of Education. (2012, July). *Teacher evaluation in Tennessee: A report on Year 1 implementation*. Retrieved from www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf

Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

Turque, B. (2011, July 15). 206 D.C. teachers fired for poor performance. *The Washington Post*. Retrieved from http://www.washingtonpost.com/

U.S. Department of Education. (2002, January 7). *The No Child Left Behind Act of 2001: Executive summary*. Retrieved from http://www2.ed.gov/nclb/overview/intro/execsumm.pdf

U.S. Department of Education. (2009, November). *Race to the Top program executive summary*. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Vogt, W. P., & Johnson, B. (2011). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences* (4th ed.). Thousand Oaks, CA: Sage.

Waters, J. T., & Marzano, R. J. (2006). *School district leadership that works: The effect of superintendent leadership on student achievement*. Denver, CO: Mid-continent Research for Education and Learning.

Weinberg, S. L., & Abramowitz, S. K. (2008). *Statistics using SPSS: An integrative approach* (2nd ed.). Cambridge, NY: Cambridge University Press.

Westinghouse, D. (2010). *Technology and the changing face of education in the 21st century: Are educational leadership programs preparing future leaders for this change?* (Doctoral Dissertation). Retrieved from http://library.sage.edu

Wheatley, M. J., & Kellner-Rogers, M. (1999). A simpler way. San Francisco, CA: Berrett-Koehler.

Williams, H. S., & Szal, D. (2011, March). Candidates' assessment of a principal preparation program. *Education, 131*(3), 481-485. Retrieved from http://www.projectinnovation.biz/education_2006.html

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

Witters-Churchill, L. (1991). University preparation of school principals. *School Organisation, 11*(3), 339-345. doi:10.1080/0260136910110309

Zerihun, Z., Beishuizen, J., & Van Os, W. (2011). Conceptions and practices in teaching and learning: Implications for the evaluation of teaching quality. *Quality in Higher Education, 17*(2), 151-161. doi:10.1080/13538322.2011.582793

*Appendix A*

Email Letter Description of Survey with Participant Consent

Dear Colleague,

I would like to start by thanking you for taking the time out of your busy schedule to read this message. I know that time is precious for administrators and as a result you must prioritize to make the use of it most beneficial. I am hoping that you review the following research and consider your participation both worthwhile and a valuable contribution to the field.

I am a doctoral candidate at Sage College in Albany in the Educational Leadership Program conducting quantitative research on the types of ongoing professional development and supports building level administrators will need to become effective evaluators of teachers, which will guide their professional growth and increase student achievement. The existing research shows that many administrators receive very little preparation on how to conduct effective teacher evaluations. The research also shows that most of the preparation that principals receive is front loaded with a focus on the process. Very few administrators receive feedback or support throughout their career to ensure that their evaluations are valid, reliable and meaningful. As education undergoes a major transformation and shifts to an era that focuses on accountability and learning the role you play, as an instructional leader, is crucial. It is also documented in the research that the number one way that schools can increase student achievement is by having highly effective teachers in the classroom. One-way to help assist teachers to grow and become proficient is by using teacher evaluation as a professional development tool. As a result you have been selected to participate in a survey of public school principals responsible conducting teacher evaluations within grades K-12. The goal of this research is to better understand what supports principals feel would best help them hone their skills as evaluators and ensure that they do not experience a model drift from the training that you are currently receiving through the Race To The Top (RTTT) initiative. This study should provide valuable information to the profession as we transition into the age of accountability where there is a strong focus on teacher evaluation and a shift on its purpose and use.

I know your time is valuable, so I am requesting approximately 15 to 20 minutes of your time to complete a simple online survey that will help in a study that could provide information that will positively impact principals. Your name and that of your school or district will not be

collected and all information is confidential. All data will be kept on the researcher's computer until the conclusion of the data-collecting phase, at which time all data will be destroyed. All data will be reported only in aggregate form. As a result, confidentiality of all participants is assured. This study is voluntary and all participants have the option to withdraw from the study at any time without penalty or loss of benefit to themselves. If a participant opts to withdraw, any data collected up to that point would be destroyed. By clicking the SurveyMonkey link, you are giving consent to participate in this survey and study. Please answer the questions honestly and to the best of your ability so that the most accurate results can be gleamed from the study. If at any point you become uncomfortable with the survey and your participation in the study, you may withdraw from the survey at any time. There are no known risks associated with this study and no deception was used.

Your participation in this study will add to the literature where there is a strong focus on teacher evaluation but little research on the professional development and supports that principals need in order to perfect their skills as evaluators. With Race To The Top and the common core standards being an initiative Nationwide, directly impacting each public school building and classroom in New York State, the data from this study may help districts, superintendents, Race To The Top coordinators, and principals to develop a system that will benefit students by supporting and strengthening the principals skills to become effective evaluators thus resulting in highly effective teachers. Teacher evaluation is a relevant topic for all administrators that will require ongoing training and support. This shift will have an impact on education and make a difference in the lives of the children you have the privilege to work with.

*Appendix B*

Survey Instrument

## Section 1: Survey Introduction

Regional Survey on Principals and Their Training On Becoming Effective Evaluators.

Dear Colleague,

Thank you in advance for completing this survey. As a building level administrator, during a time of educational reform, your feedback on principals and their training to become effective evaluators is valuable.

A shift is occurring in the purpose and process of teacher evaluations This research examines principals perspective on their preparation as evaluators.
The brief survey will take about 15 minutes to complete.

Your survey responses and contact information are strictly confidential and will not be shared. All results will be reported in aggregated form. The deadline for completing the survey is April 30, 2012.

I appreciate your feedback. It will provide the education community with added insight into the role and training of building level principals as evaluators. Aggregated results of this study will be reported in the researcher's final paper and presented at the Sage College Doctoral Colloquium in the fall of 2012.

If you have any questions concerning this research study, or wish to receive a copy of only the aggregated results please email me at hunzis@sage.edu.

Thank you for your cooperation. I hope you are able to participate in this study.

By clicking the "Next" icon you are giving consent to voluntarily participate in the survey and study.

Sincerely,

Shawn Hunziker
Doctoral Candidate
Sage Graduate School
hunzis@sage.edu

## Section 2: Building Level Principal Demographics

Please complete the following questions to the best of your ability.

**1. What is your gender?**

- ○ Female
- ○ Male

**2. Select the enrollment size interval that best describes your district?**

- ○ 0-199
- ○ 200-499
- ○ 500- 999
- ○ 1,000 - 1,999
- ○ 2,000 - 4,999
- ○ 5,000 - 9,999
- ○ 10,000 - 24,999
- ○ 25,000 - 38, 999
- ○ Greater or equal to 39,000

Other (please specify)

**3. How many years (including this school year) have you been a building level principal?**

**4. How many years (including this school year) have you been a building level principal in your CURRENT school?**

**5. Please select the configuration that best describes the grade levels that you currently supervise. (Check all that apply).**

- ☐ K-2
- ☐ 3-5
- ☐ 6-8
- ☐ 9-12
- ☐ K-12

Other (please specify)

**6. When are you mandated to implement the new APPR regulations as articulated by Race To The Top?**

- This Year (2011-2012)
- Next Year (2012 - 2013)
- I have not been informed
- This does not apply

Other (please specify)

**7. How many teachers are you responsible for evaluating this school year?**

**8. Have you ever been an assistant principal?**

- No
- Yes

**9. How many years were you an assistant principal?**

**10. Have you previously worked as a classroom teacher?**

- No
- Yes

**11. How many years did you work as a classroom teacher?**

## Section Three: Building Level Principals and Experience with Teacher Evalua...

You're about 1/3 done! Thank you so much for participating. Completing this survey will make the research findings more valuable.

Directions: When answering the following questions please answer based upon your CURRENT role in your CURRENT district.

**12. Overall how would you rate the verbal or written constructive feedback you have received, from a supervisor, regarding the teacher evaluations that you have written?**

- ○ Very Satisfied
- ○ Satisfied
- ○ Dissatisfied
- ○ Very Dissatisfied
- ○ No feedback received

**13. Prior to this year, how often was teacher evaluation training provided to administrators in your CURRENT district?**

- ○ Offered throughout the year
- ○ Offered annually (every year)
- ○ Offered biennially (every other year)
- ○ Offered to new administrators only
- ○ Not offered at all

Other (please specify)

**14. Did you participate in any of the evaluation trainings provided to administrators in your current district?**

- ○ No
- ○ Yes

**15. What type of teacher evaluation instrument are you CURRENTLY using? (Check all that apply.)**

- ☐ Checklist
- ☐ Narrative
- ☐ Rubric

**16. Which of the following teacher rubrics does your district CURRENTLY use for teacher evaluations?**

- Classroom Assessment Scoring System™ (CLASS™)
- Danielson's Framework for Teaching
- Danielson's Framework for Teaching (2011 Revised Edition)
- Marshall's Teacher Evaluation Rubric
- Marzano's Causal Teacher Evaluation Model
- Marzano's Teacher Practice Rubric
- NYSTCE Framework for the Observation of Effective Teaching
- NYSUT Teacher Practice Rubric
- Thoughtful Classroom Teacher Effectiveness Framework
- The Teaching and Learning Framework
- No rubric is being used at this time

Other (please specify)

**17. Which of the following NYSED approved teacher rubrics does your district plan to adopt to comply with the new APPR regulations?**

- Classroom Assessment Scoring System™ (CLASS™)
- Danielson's Framework for Teaching
- Danielson's Framework for Teaching (2011 Revised Edition)
- Marshall's Teacher Evaluation Rubric
- Marzano's Causal Teacher Evaluation Model
- Marzano's Teacher Practice Rubric
- NYSTCE Framework for the Observation of Effective Teaching
- NYSUT Teacher Practice Rubric
- Thoughtful Classroom Teacher Effectiveness Framework
- The Teaching and Learning Framework
- There have been no discussions regarding the selection of a rubric for teacher evaluations
- There have been discussions regarding the selection of a rubric for teacher evaluations but no decision made

**18. To what extent do you agree with the following statement: I feel very prepared to conduct effective teacher evaluations that will result in professional growth for teachers.**

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

**19. When first employed in your CURRENT district how would you rate the initial introduction to the following? (If you have not received an introduction choose N/A for not available).**

| | very satisfied | satisfied | somewhat unsatisfactory | not at all satisfactory | N/A |
|---|---|---|---|---|---|
| Teacher Evaluation system (process) | | | | | |
| Teacher Evaluation instrument (forms) | | | | | |
| Teacher Evaluation criteria | | | | | |

If you chose not at all satisfactory as a response to any of the above please explain.

## 20. What is your level of comfort with each type of evaluation instrument?

| | Very comfortable | Comfortable | Somewhat uncomfortable | Not at all comfortable |
|---|---|---|---|---|
| Checklist | ○ | ○ | ○ | ○ |
| Narrative | ○ | ○ | ○ | ○ |
| Rubric | ○ | ○ | ○ | ○ |

## 21. To what degree do you feel that you are prepared to implement the following elements of effective evaluations?

| | Very Prepared | Somewhat Prepared | Somewhat Unprepared | Very Unprepared |
|---|---|---|---|---|
| Criteria used for evaluation | ○ | ○ | ○ | ○ |
| Determining an overall accurate rating of teacher performance (HEDI) | ○ | ○ | ○ | ○ |
| Evaluate assessment data | ○ | ○ | ○ | ○ |
| Evaluate evidence of family and community outreach | ○ | ○ | ○ | ○ |
| Evaluate lesson plans | ○ | ○ | ○ | ○ |
| Evaluate student portfolios | ○ | ○ | ○ | ○ |
| Evaluate student work samples | ○ | ○ | ○ | ○ |
| Evaluate teacher portfolios | ○ | ○ | ○ | ○ |
| Evaluation instrument | ○ | ○ | ○ | ○ |
| Evaluation procedures | ○ | ○ | ○ | ○ |
| Giving feedback to teachers | ○ | ○ | ○ | ○ |
| Legal issues regarding evaluations | ○ | ○ | ○ | ○ |
| Objective observation techniques (e.g. scripting, checklist, note taking) | ○ | ○ | ○ | ○ |
| Post conference | ○ | ○ | ○ | ○ |
| Pre-Conference | ○ | ○ | ○ | ○ |
| Use of technology | ○ | ○ | ○ | ○ |
| Validity/Reliability in evaluations | ○ | ○ | ○ | ○ |

## Section Four: Principal Professional Development to Become Effective Evalua...

You are on the last section of the survey. Thank you.

Directions:

1)Please choose one response for each resource.

2)If any of the options listed below were not made available please select N/A for not available.

3)If you select N/A for any response please comment on whether you feel the resource would have been helpful if it was available.

### 22. Please choose a response for each resource or support that best describes its effectiveness in preparing you to conduct teacher evaluations

| | Very Effective | Somewhat Effective | Somewhat Ineffective | Very Ineffective | N/A |
|---|---|---|---|---|---|
| Administrative certification programs | ○ | ○ | ○ | ○ | ○ |
| Books | ○ | ○ | ○ | ○ | ○ |
| Collegial discussions | ○ | ○ | ○ | ○ | ○ |
| Continuing education courses (e.g. CAS, doctorate) | ○ | ○ | ○ | ○ | ○ |
| Current supervisor | ○ | ○ | ○ | ○ | ○ |
| Experience as an assistant principal | ○ | ○ | ○ | ○ | ○ |
| Formal district inservice training | ○ | ○ | ○ | ○ | ○ |
| Former administrative supervisor | ○ | ○ | ○ | ○ | ○ |
| Journal articles | ○ | ○ | ○ | ○ | ○ |
| Mentoring | ○ | ○ | ○ | ○ | ○ |
| On- the- job experience | ○ | ○ | ○ | ○ | ○ |
| Modeling by former administrator when being evaluated as a classroom teacher | ○ | ○ | ○ | ○ | ○ |
| Regional/state wide Race To The Top Trainings (RTTT) | ○ | ○ | ○ | ○ | ○ |
| Teaching in the classroom | ○ | ○ | ○ | ○ | ○ |
| Videos | ○ | ○ | ○ | ○ | ○ |
| Webinars | ○ | ○ | ○ | ○ | ○ |
| Workshops/Conferences | ○ | ○ | ○ | ○ | ○ |

Other resource or support (please specify and give its effectiveness)

**23. For each of the following resources please select how frequently you use and how effective you find each one for supporting your development as an effective teacher evaluator.**

| | Frequency | Effectiveness |
|---|---|---|
| Attend instructional workshops focused on specific elements within the evaluation process (pre/post conferencing, collecting evidence, giving effective feedback) | | |
| Attend instructional workshops focused on the overview of the evaluation process | | |
| Conduct tandem evaluations with a colleague | | |
| Have peers review your evaluations | | |
| Participate in a review and discussion of evaluations written by other colleagues | | |
| Participate in role plays on effective evaluations | | |
| Participate in webinars on effective evaluations | | |
| Participate with online learning courses | | |
| Read books on effective evaluations | | |
| Read case studies on effective evaluations | | |
| Read journal articles on effective evaluations | | |
| Review sample evaluations rated as exemplars | | |
| Take additional college courses on conducting effective evaluations | | |
| Take additional courses through a professional organization | | |
| Watch videos on effective evaluations | | |

**24. What other professional development would help you to become a more effective teacher evaluator?**

**25. Are there any comments you would like to share regarding principals and their training to become effective teacher evaluators?**

## Section Five: End of Survey

You are done! Thank you so much for participating in this study.

*Appendix C*

Survey Reminder Email to Principals

Research Survey: Principal readiness and professional development to conduct effective teacher evaluations that lead to improved student achievement

This is a follow up to the email that was sent on April 11, 2012 regarding a survey that I am conducting for my doctoral research. I know you are busy as an administrator and that you receive a lot of emails per day. If you have already responded to the survey THANK YOU. I appreciate your input and contribution to the field. If you have not had time to fill out the survey I ask that you consider participating. If you have any questions or are unable to open the survey from the link provided please email me to let me know.

Here is a link to the survey:
https://www.surveymonkey.com/s.aspx

This link is uniquely tied to this survey and your email address. Please do not forward this message.

Thanks for your participation!

Please note: If you do not wish to receive further emails from us, please click the link below, and you will be automatically removed from our mailing list.
https://www.surveymonkey.com/optout.aspx