The Influence of Principal and School Characteristics on Principal Ratings of Teachers Using the North Carolina Teacher Evaluation Instrument

By Stacy Leggett

B.A.in Political Science, May 1991, Mississippi College
M.A.in History, May 1997, Mississippi College
M.S.A.in Educational Leadership, July 2005, East Carolina University

A Dissertation Submitted to

The Faculty of
The Graduate School of Education and Human Development
of The George Washington University
in partial fulfillment of the requirement
for the degree of Doctor of Education

January 31, 2015

Dissertation directed by

Jennifer Clayton
Assistant Professor of Education Administration

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

The Graduate School of Education and Human Development of The George Washington University certifies that Stacy Leggett has passed the Final Examination for the degree of Doctor of Education as of November 21, 2014. This is the final and approved form of the dissertation.

The Influence of Principal and School Characteristics on Principal Ratings of Teachers Using the North Carolina Teacher Evaluation Instrument

Stacy Leggett

Dissertation Research Committee

Jennifer K. Clayton, Assistant Professor of Education Administration, Dissertation Director

Sharon A. Dannels, Associate Professor of Education Leadership, Committee Member

Marguerita DeSander, Department Head of Educational Administration, Leadership, and Research; Western Kentucky University; Committee Member

Acknowledgements

I would like to thank Dr. Jennifer Clayton, my advisor and chairperson, for her patience and the hundreds of pages of varying drafts that she spent countless hours reading and providing feedback. She understood the balance that I needed between prodding me forward and giving me time when the demands of my job became somewhat overwhelming at times. To my methodologist, Dr. Sharon Dannels, thank you for your feedback that constantly "nudged" me to improve my document and for answering my countless methodology questions. Finally, to Dr. Marguerita DeSander for her continued reading and support from Kentucky.

Thanks to Dr. Lee Bolman and Dr. Carol Dweck for permission to use their instruments in my survey: the Leadership Orientations Instrument—Self (Bolman & Deal, 1990) and the Kind of Person Instrument—Others (Dweck, Chiu, & Hong, 1995). Their support of doctoral level research is much appreciated.

Abstract of Dissertation

The Influence of Principal and School Characteristics on Principal Ratings of Teachers Using the North Carolina Teacher Evaluation Instrument

The purpose of this study was to evaluate the influence of principal and school characteristics on principal mean ratings of teachers using the North Carolina teacher evaluation instrument. A review of recent literature identified principal, teacher, evaluation process, and school characteristics possibly influencing principal ratings of teachers, but the studies found explicitly addressing these relationships focused specifically on the relationship between principal ratings and teacher effectiveness as measured by value-added data. A series of simple and multiple regression tests were used in this study to examine the influence of principal characteristics and school characteristics on the distribution of principal ratings for Standards 1 through 5 of the North Carolina evaluation instrument. The predictor variables were principal years' experience as a principal, principal implicit person theory, principal number of dominant leadership orientation frames, school grade span, and school growth status. Principal years' experience as an administrator and teacher were also collected. Exploratory variables included were school Title I status, teacher turnover rate, and the percent of teachers with less than three years' experience.

To evaluate the influence of principal and school characteristics, a stratified, proportional sample of 399 principals were invited to participate in an online survey. Only 73 principals responded with only 68 of the responses meeting the criteria of completion to be used in the study. Principal years' experience as an administrator,

school teacher turnover rate, and school percent of teachers with less than three years' experience significantly influenced principal mean ratings of teachers. The more years' experience as an administrator, the higher the principal mean rating assigned. The higher teacher turnover rate and percent beginning teachers, the lower principal mean rating assigned. This was an exploratory study revealing further opportunities for study on the influence of factors other than teacher effectiveness influencing principal ratings.

Table of Contents

Page

List of Figures

List of Tables

Chapter 1: Introduction

From the 1800s until the end of the twentieth century, teacher evaluations evolved from simple checklists of basic requirements to a combination of checklists and anecdotal records focusing on minimal requirements for teaching proficiency (Danielson, 2000; Peterson, 2004; St. Maurice & Cook, 2005). These traditional assessments, both the checklists and anecdotal records, were only loosely coupled with student learning, teacher growth, and school improvement (Aseltine, Faryniarz, & Rigazio-Digilio, 2006; Darling-Hammond, 1990; Elmore, 2004). As the context of educational accountability evolved through initiatives associated with the No Child Left Behind (NCLB) Act (2001), the Race to the Top (RTTT) initiative (American Recovery and Reinvestment Act, 2009), and other state and national policies, the demands for instructional leadership and evaluation systems connected to teacher effectiveness and student achievement increased. With these reforms, the focus shifted from teaching—the process—to the importance of student learning—the outcome (Brown, 2010; National Governors Association, 2011).

The mission of the North Carolina State Board of Education (NCSBOE) adopted in 2007 stated, "Every public school student will graduate from high school, globally competitive for work and postsecondary education and prepared for life in the 21st Century" (North Carolina Professional Teaching Standards Commission, 2007, n.p.). The NCSBOE charged the North Carolina Professional Teaching Standards Commission (NCPTS), a committee of 16 educators, with considering, "what teachers need to know and be able to do in 21st Century schools" (NCPTS, n.p.). The NCSBOE subsequently adopted the standards developed by this commission as the foundation for the North

Table 1

Ratings of Teachers on the North Carolina Teacher Evaluation Instrument

Rating	Definition
Distinguished	Teacher consistently and significantly exceeded basic competence on standard(s) of performance.
Accomplished	Teacher exceeded basic competence on standard(s) of performance most of the time.
Proficient	Teacher demonstrated basic competence on standard(s) of performance.
Developing	Teacher demonstrated adequate growth toward achieving standard(s) during the period of performance, but did not demonstrate competence on standard(s) of performance.
Not Demonstrated	Teacher did not demonstrate competence on or adequate growth toward achieving standard(s) of performance.

Note: From "North Carolina Teacher Evaluation Process" by the North Carolina Public Schools, 2012, p. 20.

Carolina teacher evaluation process (North Carolina Public Schools, 2010). The new instrument contained five standards to be rated by the principal: "Teachers demonstrate leadership" (North Carolina Public Schools, 2012, p. 8); "Teachers establish a respectful environment for a diverse population of students" (North Carolina Public Schools, p. 9); "Teachers know the content they teach" (North Carolina Public Schools, p. 10); "Teachers facilitate learning for their students" (North Carolina Public Schools, p. 10); and "Teachers reflect on their practice" (North Carolina Public Schools, p. 12). On these five standards, teachers can be rated as *not demonstrated*, *developing*, *proficient*, *accomplished*, and *distinguished*. These ratings are defined in Table 1 above.

North Carolina implemented the new evaluation process in three phases: 13 districts in the summer of 2008, 39 districts in 2009, and 65 districts in 2010 (North Carolina Public Schools, 2008a). A policy brief presented to the NCSBOE after the final

phase of rollout described a discrepancy between the percent of teachers in a school receiving the two highest ratings, accomplished and distinguished, and the school's aggregate student growth (North Carolina Public School, 2011). There was no statistically significant difference in teacher ratings between schools making expected growth and schools not making expected growth based on criteria established by the NCSBOE. See Table 2 for the distribution of ratings. Based on this lack of statistically significant difference, the conclusion of the brief supported, "a need for more training on the teacher evaluation instrument and a more explicit inclusion of student growth data" (North Carolina Public Schools, 2011, p.1).

A sixth standard, "Teachers contribute to the academic success of students" (North Carolina Public Schools, 2012, p. 12) was added to the evaluation instrument in 2012 with the following ratings: *does not meet expected growth*, *meets expected growth*, and *exceeds expected growth* (North Carolina Public Schools). For 2011-2012, teachers with available value-added data received ratings with 70% of the measure based on the teacher individual data and 30% based on the school data; teachers with no available data received ratings solely based on school data (North Carolina Public Schools).
For 2012-2013, only teachers with available value-added data received ratings; the ratings were based solely on individual data (North Carolina Public Schools, 2013).

Research exploring principal characteristics influencing principal ratings of teachers is limited. Most empirical studies of evaluation processes since 2000 have been case studies focusing on district-wide implementation of evaluation processes. Many of these case studies compared principal ratings of teachers to student achievement data. A

Table 2

Percent of Teachers Rated as Accomplished or Distinguished

Standard	Percent of Teachers in Schools not Making Growth Rated as Accomplished or Distinguished	Percent of Teachers in Schools Making Growth Rated as Accomplished or Distinguished
Standard One	67.39	64.06
Standard Two	67.21	66.87
Standard Three	55.17	60.99
Standard Four	63.79	63.93
Standard Five	72.00	63.80

Note: From "Briefing on Educator Effectiveness Policies (Prepared for the Members of the Governor's Education Transformation Commission)" by the North Carolina Public Schools, 2011, p. 2.

review of these studies produced five potential concerns when comparing principal ratings to teacher effectiveness data: (a) Principals can be inconsistent in their abilities to identify effective teachers; (b) Principal ratings can be more strongly correlated to student achievement levels than value-added data; (c) Value-added data based on a single year of data are not as reliable as data collected over multiple years; (d) Students of various backgrounds and achievement levels are not distributed randomly among teachers; and (e) Differences in strength of correlation have been found in most studies across content areas and grade levels (Borman & Kimball, 2005; Eady & Zepeda, 2007; Holtzapple, 2005; Kane & Staiger, 2012; Jacob & Lefgren, 2006; Jacob & Lefgren, 2008; Kimball, White, Milanowski, & Borman, 2004; Weisberg, Sexton, Mulhern, & Keeling, 2009). A limited number of case studies have addressed principal or teacher reactions to and perceptions of evaluation processes (Conley, Muncey, & You, 2005: Kane, Taylor, Tyler,

& Wooten, 2011; Kimball, 2002; Larsen, 2009; Painter, 2000; Rowe, 2004; Sawyer, 2001; Wise, Darling-Hammond, McLaughlin, & Berstein, 1985).

A recent, more comprehensive study published by the New Teacher Project surveyed 15,000 teachers and 1,300 administrators from 12 districts of varying size, location, and evaluation processes in Arkansas, Colorado, Illinois, and Ohio. This study led to the conclusion evaluation systems do not clearly differentiate between effective and ineffective teachers, a phenomenon described as the *widget effect*:

> The failure of evaluation systems to provide accurate and credible information about individual teachers' instructional performance sustains and reinforces a phenomenon that we have come to call the Widget Effect. The Widget Effect describes the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher. (Weisberg et al., 2009, p. 9)

Weisberg et al. focused on evaluation process characteristics including the frequency and duration of observations; the differentiation of the process for veteran, novice, and struggling teachers; the distribution of rewards as a result of the process; and failure of the process to identify professional development needs.

The review of the literature led to the question: If teacher effectiveness is not the only determining factor in principal ratings of teachers, what other factors influence these ratings? Principals observe and provide feedback to teachers to facilitate teacher growth, but factors outside teacher effectiveness can influence principal ratings of teachers including principal training, expertise, and experience (Bryant & Currin, 1995; Kerrins & Cushing, 2000; Sartain et al., 2011; Wise et al., 1985); instructional leadership decisions and leadership style (Davis, Ellett, & Annunziata, 2002; Kimball, 2002; Ovando & Ramirez, 2007; Tuyten, 2009); and professional relationships with teachers (Piggot-

Irvine, 2003; Wise et al). The literature suggests the characteristics of the following as categories of potential factors influencing principal ratings: teachers, students, evaluation systems, and principals.

This study focused on the following principal factors possibly influencing principal ratings of teachers: principal number of dominant leadership frames as measured by Bolman and Deal's (1990) Leadership Orientations Instrument—Self, Section I; principal beliefs about the malleability of others' personal characteristics and abilities as measured by the Kind of Person Instrument (Dweck et al., 1995); and principal years' experience as principal. School grade span and school growth status were also included.

This study addressed gaps in the literature. Limited research analyzed factors influencing principal ratings of teachers. Although substantial literature existed in psychology examining the influence of implicit person theories on reactions to failure and selection of goals, implicit person theories and judgments of others, implicit person theories in non-educational evaluation settings, limited research explored the relationship between implicit person theories and evaluation processes in education (Dweck, 2000). Although studies were found examining the relationship between leadership orientation and educational leadership, no empirical studies were found in searches of ERIC, JSTOR, and PROQUEST databases addressing the influence of Bolman and Deal's (1990, 2003) leadership orientation frames on principal evaluations of teachers.

Problem Statement

The failure of over 30% of North Carolina students to graduate with their cohort in 2008 suggested almost one-third of North Carolina's students were being left behind for academic, social, or other reasons (North Carolina Public Schools, 2008b). The result of a mandate to improve graduation rates, the North Carolina Professional Teaching Standards encouraged teachers to grow professionally and become more adept at meeting the needs of all learners. Evidence of the intention to meet the needs of all learners can be found in the indicator found in the instrument that every teacher, K-12, take "responsibility for the progress of students to ensure that they graduate from high school" (North Carolina Public Schools, 2010, p. 21), and requirements for all teachers to integrate global awareness, differentiate instruction, display knowledge of diverse cultures, and use multiple literacy strategies (North Carolina Public Schools, 2010). Despite efforts to link student learning outcomes to the evaluation process while continuing to capture other aspects of teacher quality, initial results did not demonstrate a significant correlation between principal ratings of teachers and aggregate teacher value-added data (North Carolina Public Schools, 2011) suggesting a need to analyze factors beyond teacher value-added data influencing principal ratings of teachers.

Research Questions

The research questions were addressed through a survey of a stratified, proportional sample of 226 elementary, 88 middle, and 85 high school principals in North Carolina. The following overarching question guided this study: Do principal and school characteristics influence principal ratings of teachers on the North Carolina teacher

evaluation instrument? These analytic questions were used to answer the overarching question:

1. Do principal characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    a. Do principal implicit person theories as measured by the Kind of Person Instrument (Dweck et al., 1995) influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    b. Does principal use of dominant leadership frames as measured by Bolman and Deal's (1990) Leadership Orientations Instrument influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    c. Does principal total years' experience as a principal influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
2. Do school characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    a. Does the school value-added composite measured by EVAAS data influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    b. Does school grade span as reported by the principal influence principal ratings of teachers on the North Carolina teacher evaluation instrument?

Purpose of the Study

This study investigated factors influencing principal ratings of teachers on the North Carolina teacher evaluation process using a nonexperimental design and a web-

based survey instrument to collect data. Descriptive statistics were used to summarize overall characteristics of participants, and a series of multiple regression tests were applied to analyze the relationships between the predictor and criterion variables. The following were the predictor variables: (a) the number of dominant leadership frames used by the principal in the evaluation process as reported by the principal and measured by Bolman and Deal's (1990) Leadership Orientations Instrument; (b) the dominant implicit person theory of the principal as measured by the Kind of Person Instrument (Dweck et al., 1995); (c) principal total years' experience as a principal; (d) the grade span of students served by the school as reported by the principal; and (e) the school growth status as measured by EVAAS composite. The criterion variables were the principal mean ratings of teachers on each of Standards 1 through 5 with ratings assigned a point value of 0 (*not demonstrated*) to 4 (*distinguished*). Because North Carolina does not calculate mean overall ratings of teachers, this study addressed each standard's mean rating as a separate criterion variable.

Statement of Potential Significance

This quantitative study used multiple regression techniques to determine the influence of principal and school characteristics on principal ratings of teachers. This study is important because no studies could be found analyzing the influence of principal characteristics on their ratings of teachers. The study conducted by the New Teacher Project recognized the failure of evaluation processes to identify effective teachers but focused on evaluation processes and outcomes (Weisberg et al., 2009).

Varying strengths of correlation between principal ratings and value-added data have been found in studies, but no studies were found explicitly exploring the relationships between principal and school characteristics and principal ratings of teachers. Researchers have examined perceptions of evaluation processes, described the evaluation process, and identified problems associated with the process, but they have not focused on a key player in the process, the principal. Data collected through this research could be useful to policy makers, departments of education, and institutions of higher education in providing a better understanding of the principal's role in the evaluation process.

Summary of Methodology

This research was grounded in the premise principal, school, and teacher characteristics influence principal ratings of teachers on the North Carolina teacher evaluation process. This quantitative study employed a nonexperimental, cross-sectional survey and simple and multiple regression techniques to evaluate the relationships among factors possibly influencing principal ratings of teachers using the North Carolina teacher evaluation process. The survey included frequency rating scale items, agreement rating scale items, and questions related to demographic and school information. The survey included the first section from Bolman and Deal's (1990) Leadership Orientations Instrument—Self, the three entity theory items from the Kind of Person Instrument (Dweck et al., 1995), and basic demographic questions about the principals and their education backgrounds. The survey items from the Leadership Orientations Instrument

(Bolman & Deal) and the implicit person theory survey (Dweck et al., 1995; Dweck, 2000) were validated in previous studies. These studies are addressed in Chapter 3.

A stratified, proportional sampling technique was used to select 226 elementary, 88 middle, and 85 high school principals in schools with value-added data for a sample frame of 399. Only principals participating in the summative evaluation process at the school during the previous year were included in the study. Participation in the study was voluntary. Principals received an email describing the study and providing the URL to the survey. The survey was administered online using the Survey Monkey online survey service, and the responses were downloaded into IBM SPSS (Statistical Product and Service Solutions) Statistics GradPack for Windows for analysis.

Theoretical and Conceptual Framework for Forthcoming Study

Previous research focused on identifying teacher characteristics explaining principal ratings of teachers. Kane and Staiger found teacher practice explained only .14 to .37 of variability in principal ratings of teachers. They also found differences among observers. This study explored possible internal characteristics of evaluators in the role of principal and external characteristics of the school setting to identify factors not directly related to teacher practices influencing principal ratings of teachers. Social cognitive theory (Bandura, 1986) was used as the overarching theoretical framework for this study to explore the influence of personal characteristics and environmental factors on the rating process. For this study, personal characteristics focused on characteristics of the principal, and environmental factors addressed characteristics of the school. The conceptual framework for this study incorporated these additional theories and

constructs: (a) implicit person theory (Dweck, 2000); (b) leadership orientation frames (Bolman & Deal, 1990, 2003); and (c) the North Carolina teacher evaluation instrument. Review of the literature indicated the potential influence of internal and external factors including principal leadership style and principal beliefs about the malleability of others' personal characteristics and abilities on their ratings of teachers.

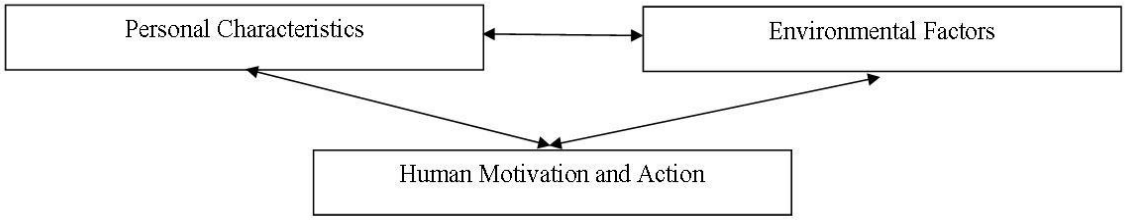
Social Cognitive Theory

Social cognitive theory provides a framework for analyzing reciprocal causal relationships among personal characteristics, environmental factors, and human motivation and action as illustrated in Figure 1. This theory is rooted in the premise human behavior cannot be separated from the characteristics of the individuals involved and the characteristics of the environmental context. "What people think, believe, and feel affects how they behave," (Bandura, 1986, p. 25). Internal forces, external forces, and human behavior interact with each other:

> In the social cognitive view people are neither driven by inner forces nor automatically shaped and controlled by external stimuli. Rather, human functioning is explained in terms of a model of triadic reciprocality in which behavior, cognitive and other personal factors, and environmental events all operate as interacting determinants of each other. (Bandura, p. 18)

An examination of the literature yielded the following as possible internal factors influencing principal ratings of teachers: (a) principal leadership style; (b) principal expertise and competence in supervision; (c) principal background prior to educational administration; and (d) principal implicit person theory beliefs. Environmental or external factors potentially influencing a principal's ratings were (a) the evaluation process itself; (b) teacher characteristics including teacher effectiveness as measured by

Figure 1



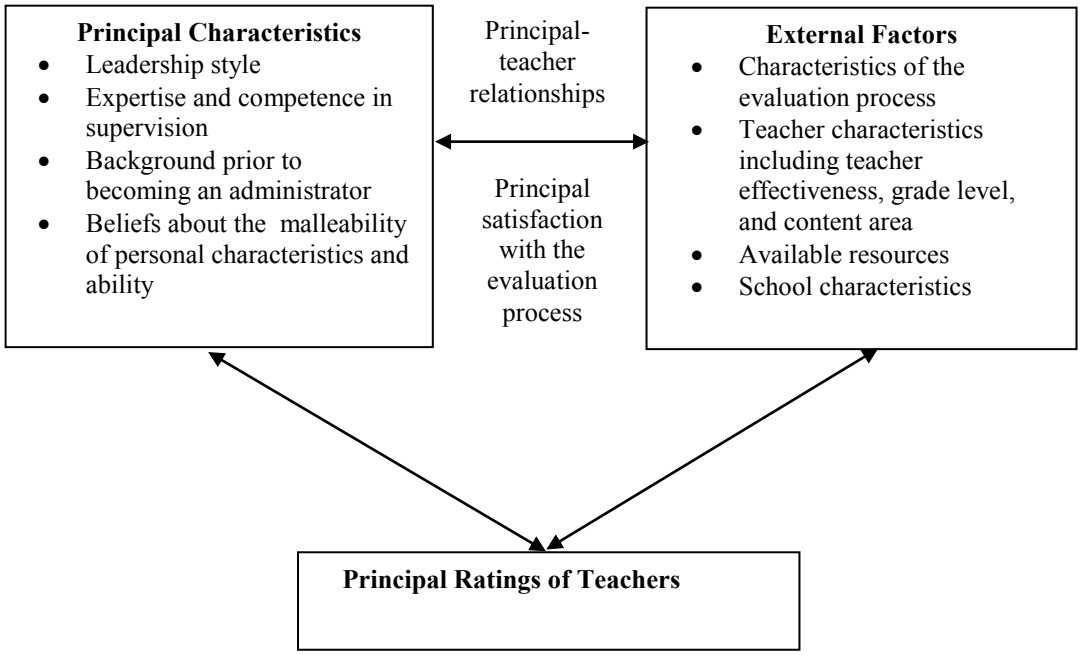
Social Cognitive Theory (Bandura, 1986)

Note: An illustration of Bandura's (1986) social cognitive theory's bidirectional interactions among personal characteristics, environmental factors, and human motivation and action. Based on theory published in *Social Foundations of Thought and Actions: A Social Cognitive Theory* by Prentice-Hall, Inc., Englewood Cliffs, NJ.

value-added data, grade level, and content area; (c) barriers and problems encountered in the evaluation process; (d) resources including time and support; and (e) school characteristics such as grade span and accountability designation.

The literature also supported the hypothesis that the interactions of external and internal factors influence principal ratings; for example, principal relationships with teachers influence principal ratings of teachers (Bryant & Currin, 1995; Piggot-Irvine, 2003). Another example of interaction of factors is the influence of principal satisfaction with the evaluation process on their implementation of the process. Limited research was found addressing the influence of teacher and supervisor satisfaction but no empirical studies were found addressing the influence of principal satisfaction (Conley et al., 2005; Kimball, 2002). Figure 2 adapts the basic graphic provided in Figure 1 of social cognitive theory to provide a visual outline of factors possibly influencing principal ratings of teachers focusing on principal characteristics as "internal characteristics" and school characteristics as "external characteristics" as they are external to the principal.

Figure 2



Possible Factors Influencing Principal Ratings of Teachers

Note: Factors extrapolated from review of literature with potential for influencing principal ratings of teachers.

Implicit Person Theory

Implicit person theories influence actions in a variety of ways including how individuals address complex academic problems, interact with others, react to successful and unsuccessful outcomes, and perceive themselves (Dweck & Leggett, 1988). Individual beliefs about the malleability of personal characteristics and ability can be described as entity or incremental theories (Dweck, 2000). Entity theorists believe personal attributes such as intelligence, personality, and social skills are fixed. Incremental theorists believe individual traits such as intelligence, personality, and social skills can grow and develop over time (Dweck, 1986). Early implicit person theory research focused on self-perceptions related to intelligence. When faced with complex

problems in academic settings, entity theorists focused on how much intellectual ability they believed they possessed and attributed failure to lack of intellectual ability. Incremental theorists focused on developing their ability and improving performance by exerting effort and applying problem-solving strategies (Dweck & Leggett).

Individuals can hold both incremental and entity theories as they do not typically systematically search for discrepancies among beliefs, but studies show about 85% of individuals hold a dominant implicit theory (Dweck et al., 1995). After excluding the 15% of individuals who typically do not fall clearly into either category, participants were evenly distributed as entity and incremental theorists unrelated to level of education, cognitive ability, self-esteem, or optimism (Dweck et al., 1995).

Implicit person theories affect social interactions including reactions to rejection and perceptions of others (Chiu, Hong, & Dweck, 1997). Entity theorists are more likely than incremental theorists to attribute long-term negative traits to oneself when rejected, stereotype others, agree more strongly with stereotypes, more closely associate groups with stereotypes, attribute traits to members of groups based on group identity, use stereotypes to reflect group differences, group individuals based on stereotypes, and use more extreme qualifiers to describe traits of groups. This tendency to stereotype and use traits as the primary measure in evaluating others is attributed to entity theorist beliefs that traits are inherent. Entity theorists more quickly make judgments of groups because they believe a person demonstrating trait-related behavior in one situation means the person was more likely to demonstrate trait-related behavior in subsequent situations (Chiu et al., 1997; Levy, Plaks, Hong, Chiu, & Dwek, 2001; Levy, Stroessner, & Dweck,

1998). Individuals with incremental theories were more likely to request additional information before making judgments (Gervey, Chiu, Hong, & Dweck, 1999).

Leadership Orientation Frames

Bolman and Deal (1990, 2003) provided leaders four lenses for framing organizations and recommended all four frames be used for successful leadership. Leaders operating in the structural frame work to, "achieve established goals and objectives," "increase efficiency and enhance performance," "align goals of organization and individuals," "design structures to meet organizational needs," and "address problems through careful analysis and restructuring" (Bolman & Deal, p. 45). Leaders operating in the human resource frame work to "meet human needs within the organization" and "ensure a good fit between individuals and the organization" (Bolman & Deal, p. 115). Leaders operating in the political frame must be able to "manage conflict," "develop coalitions," and "allocate scarce resources" (Bolman & Deal, p. 115). Finally, leaders operating in the symbolic frame "make meaning within the organization," "unite the organization around shared values and beliefs," and "create symbols to guide the organization" (Bolman & Deal, p. 123-124). Bolman and Deal (2003) described the following interpretations of evaluation processes through the four frames: Evaluating in the structural frame is a "way to distribute rewards or penalties and control performance" (Bolman & Deal, p. 306); evaluating in the human resource frame is a "process for helping individuals grow and improve" (Bolman & Deal, p. 306); evaluating in the political frame is an "opportunity to exercise power" (Bolman & Deal, p. 306); and

evaluating in the symbolic frame is an "occasion to play roles in shared drama" (Bolman & Deal, p. 306).

North Carolina Teacher Evaluation Instrument

A commission of 16 educators developed five standards aligned with the mission of the North Carolina State Board of Education adopted in 2007, "Every public school student will graduate from high school, globally competitive for work and postsecondary education and prepared for life in the 21st Century" (North Carolina Professional Teaching Standards Commission, 2007). The standards are: (a) Teachers demonstrate leadership; (b) Teachers establish a respectful environment for a diverse population of students; (c) Teachers know the content they teach; (d) Teachers facilitate learning for their students; and (e) Teachers reflect on their practice (North Carolina Professional Teaching Standards Commission). This instrument was implemented in three phases from 2008 to 2011 (North Carolina Public Schools, 2008a).

The initial purposes of the new evaluation system were to assess teacher performance and guide the development of professional growth plans. The process relied on data from teacher self-assessment, reflection, collected artifacts, and classroom observations. Teachers were classified as (a) beginning teachers during their first three years; (b) probationary teachers if they were teachers with more than three years' experience without career status; and (c) career-status teachers if granted this status by their current school district. The North Carolina teacher evaluation process required nine components for beginning teachers annually, probationary teachers annually, and career status teachers in their licensure renewal year: training, orientation, teacher self-

assessment, pre-observation conference, observation, post-observation conference, summary evaluation conference and scoring the teacher's summary ratings. These components are outlined in Figure 3. This study focused on the ratings assigned during the summary evaluation conference. The evaluation requirements for career-status teacher evaluations during nonrenewal years could be met through an abbreviated evaluation process. For abbreviated evaluations, the evaluator rated the teachers only on Standards One, Four, and Six with a minimum of two twenty-minute informal observations (North Carolina State Board of Education, TCP-C-004, 2011).

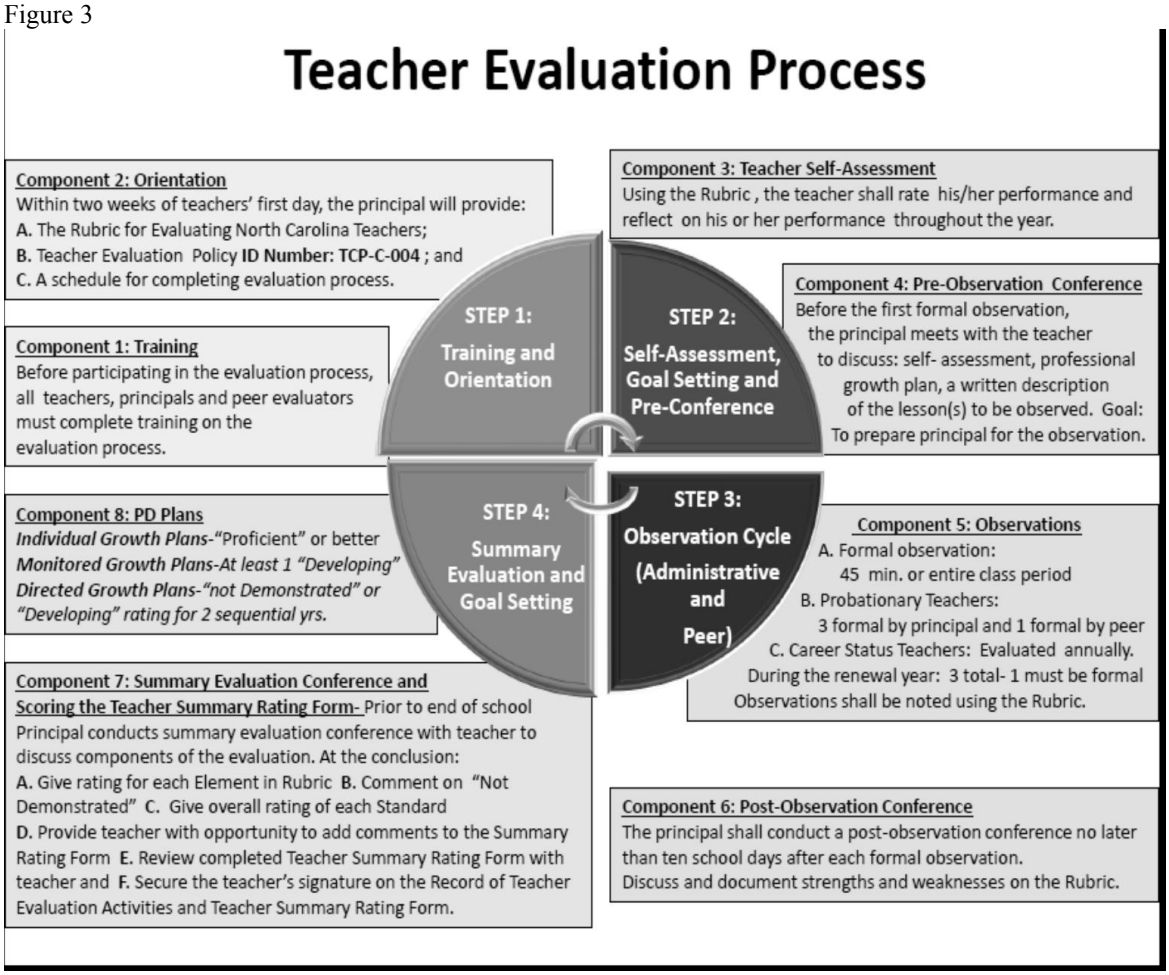
The eight purposes of the North Carolina teacher evaluation process as stated by North Carolina Public Schools (2008b) include the following:

> Serve as a measurement of performance for individual teachers; serve as a guide for teachers as they reflect upon and improve their effectiveness; serve as the basis for instructional improvement; focus the goals and objectives of schools and districts as they support, monitor, and evaluate their teachers; guide professional development programs for teachers; serve as a tool in developing coaching and mentoring programs for teachers; enhance the implementation of the approved curriculum; and inform higher education institutions as they develop the content and requirements for teacher training programs. (North Carolina Public Schools, p. 1-2)

Synthesis of Theories and Construct into a Conceptual Framework

There are numerous possible factors influencing principal ratings of teachers. This study focused on the following principal characteristics: (a) principal leadership styles as measured by Bolman and Deal's (1990, 2003) Leadership Orientation Frames; (b) principal implicit person theories as measured by the Kind of Person Instrument (Dweck et al., 1995); and (c) principal total years' experience as a principal. The following teacher and school characteristics were included: (a) school growth status

Figure 3

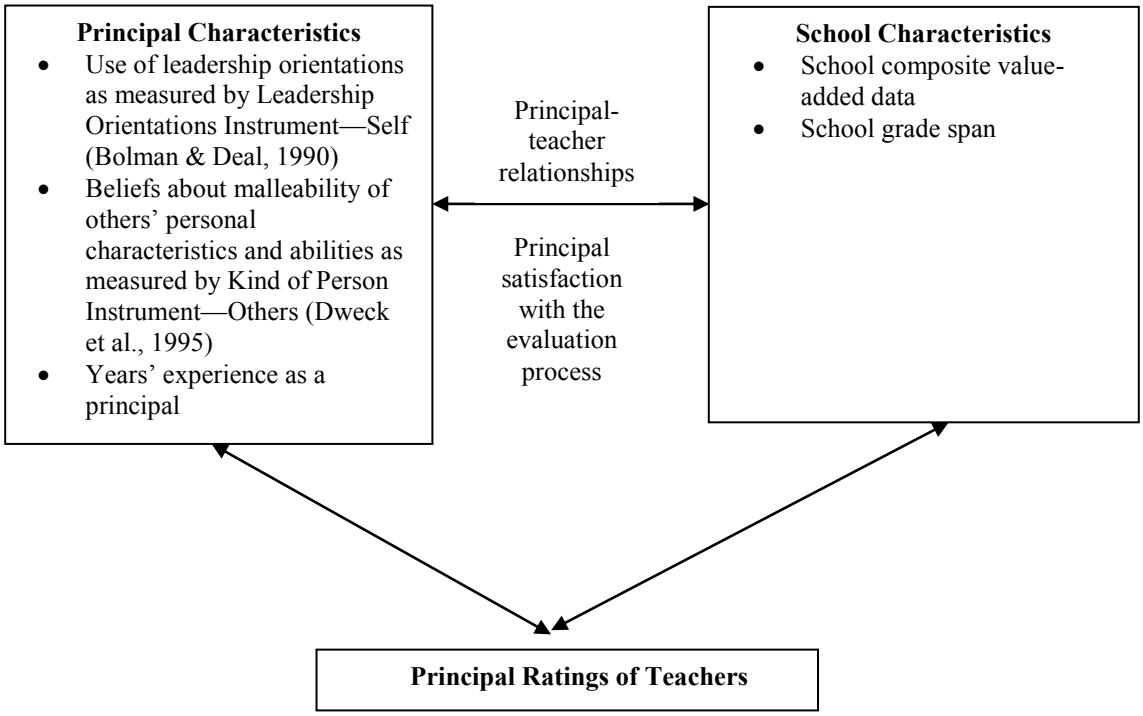


Components of the North Carolina Teacher Evaluation Process

Note. Description of the nine components and the four steps of the North Carolina Teacher Evaluation Process by North Carolina Public Schools. Retrieved from http://ncees.ncdpi.wikispaces.net/NC+Teachers on April 6, 2013.

based on teacher value-added data and (b) grade span of the school. Figure 4 illustrates these relationships. There were possible factors that were not addressed in the study. Their omission is described in the following section, "Delimitations." The original design of this study examined the influence of principal implicit person theories, principal leadership orientation frames, principal years' experience as a principal, school grade span, and school growth status on principal ratings of teachers on Standards 1

Figure 4



Conceptual Framework of Principal and School Characteristics Influencing Principal Ratings of Teachers

Note: Conceptual framework illustrating the relationships among social cognitive theory, principal leadership orientation, principal implicit person theory, and the North Carolina teacher evaluation instrument.

through 5 and used simple and multiple regression analysis to determine the factors which best predicted principal ratings of teachers. Because separate tests were conducted for each standard, all available ratings were used. Standards 1 and 4 include ratings for all teachers. Standards 2, 3, and 5 include ratings for beginning teachers, teachers without career status, career status teachers during their licensure renewal year, and possibly career status teachers selected by the principal to receive ratings on all standards.

Delimitations

This study was limited to principals in North Carolina who have used the North Carolina teacher evaluation process to assign teachers their most recent summative rating.

Ratings were assigned at the end of the 2012-2013 school year, but the survey was not conducted until January 2014. Principals in their first school year as principal or in their first school year as principal of that school were not included in this study. The findings may not be generalizable to principals in other states or schools with different evaluation processes. Only principals were surveyed in this study, so the findings were not generalizable to others participating in evaluation processes including teacher leaders, mentors, assistant principals, curriculum coaches, and central office personnel.

An Excel pivot table of the schools and grade levels in North Carolina generated 57 grade level configurations. For the purposes of this study, elementary schools were limited to schools with Grade 4 and/or Grade 5 students because value-added data were only available for Grade 4 through high school. Elementary schools did not include schools serving students beyond Grade 6. Middle schools only included schools serving students in Grades 6, 7, and/or 8. High schools only included schools serving students in Grades 9-12. Schools designated as early colleges and serving students in Grades 9-13 were not included. Schools serving multiple grade spans such as 7-12 schools with both middle school and high school students were not included.

Numerous factors with potential for influencing principal ratings of teachers were not included in this study. The evaluation process was not addressed as the process was mandated by the North Carolina State Board of Education. Individual teacher value-added data, grade levels, and content areas were not addressed because this study only examined aggregate teacher data. Because each district and regional education services agency received Race to the Top funds and services from Department of Public Instruction consultants to provide similar regional professional development and support,

barriers, problems, and available resources were not addressed as districts were implementing the same process with similar regional professional development and support. School characteristics were collected in the survey, but only grade span was included as a research question as there were no empirical studies related to Title I status and teacher evaluations or accountability designation and teacher evaluation. Only student achievement related to school growth status as designated by accountability standards based on value-added data were included; other student characteristics were not included. This information was collected from the school report cards. Although social cognitive theory focuses on the interaction of personal characteristics, environmental factors, and motivation and action, interaction of factors including teacher-principal relationships and principal perceptions of the process were not addressed. This study only examined interaction between principal characteristics and principal mean ratings and school characteristics and principal mean ratings.

Limitations

There are limitations to this study associated with the use of self-reported data, the principal as the unit of study, and multiple regression techniques. Principal leadership orientation and implicit person theory were based on self-reported data. The following limitations are associated with the use of self-reported data: (a) Principal dominant leadership orientations were based on principal perceptions of their own leadership not on teacher perceptions; (b) Only using self-evaluations decreased the validity of the Leadership Orientations Instrument; and (c) Principals possibly reported their implicit person theory based on their perceptions of the correct answer not their true theory. Focusing on the principal as the unit of study instead of the student or teacher meant

student characteristics such as socioeconomic status and prior achievement or teacher characteristics such as content area, grade level, or career status were not included. Finally, the design of this study only determined if predictive relationships, not causal relationships, existed. Regression techniques can identify relationships but not the underlying causes. Because the independent variables could not be manipulated for this study, causal relationships could not be identified. Due to the size of the study and confidentiality of teacher evaluations, this study only examined principal aggregate ratings of teachers, not their ratings of individual teachers. Not including individual teacher ratings also limited this study's ability to identify if principal ratings of teachers within their building correlated to individual teacher value-added data.

Definition of Terms

Beginning teachers are teachers with less than three years' teaching experience in North Carolina. Beginning teachers are required to participate in all components of the evaluation process each year including a self-evaluation, three observations by principal, one observation by peer, and other components described previously in Figure 3 (North Carolina Public Schools, 2010).

Student growth "is the change in student achievement for an individual student between two points in time" (NCDPI, 2013, p. 2). In North Carolina student growth is measured by Education Value-Added Assessment System (EVAAS) value-added models (VAM).

Value-added models (VAM) in North Carolina use univariate and multivariate response models to project student performance using prior performance. Student observed scores are then compared to predicted scores; the difference between these two

scores is the teacher's value-added data (Wright, White, Sanders, & Rivers, 2010).

Expected growth for schools in North Carolina is a formula developed by the North Carolina Public Schools and adopted by the North Carolina State Board of Education based on prior performance of students within the school; statewide performance of students with similar abilities; and statistical adjustments to account for differences across subjects and grade levels. Expected growth is the amount of growth reasonably expected over the course of the year and is predicted using EVAAS value-added models with student prior achievement. Models are created for each subject and grade level. These models apply previously available data to multivariate models when possible. Univariate response models are used when data is not available to use multivariate response models.

Growth index is a standardized measure of effectiveness calculated by dividing the school's value-added composite by the standard error of measure (SAS, n.d.).

Meets expected growth is a status assigned by North Carolina to schools with a growth index between -2 and +2 (SAS, n.d.).

Exceeds expected growth is a status assigned by North Carolina to schools with a growth index greater than +2 (SAS, n.d.).

Does not meet expected growth is a status assigned by North Carolina to schools with a growth index less than -2 (SAS, n.d.).

The *Rubric for Evaluating North Carolina Teachers* is the evaluation instrument for completing the self-assessments, administrative observations, peer observations, and summative evaluations*;* it is a matrix of standards, elements, and descriptors based on the North Carolina teacher evaluation process (North Carolina Public Schools).

Social cognitive theory provides a framework for analyzing the bidirectional causal relationships among personal characteristics, environmental factors, and human motivation and action (Bandura, 1986).

Implicit person theories are beliefs held by individuals about the potential for themselves and others to change that are difficult to articulate. This study focuses specifically on implicit person theories about the malleability of others' personal characteristics and abilities (Dweck et al., 1995).

Incremental theorists believe in the malleability of others' personal characteristics and abilities. They believe a person's characteristics and abilities can grow and change over time (Dweck et al., 1995).

Entity theorists believe that a person's characteristics and abilities are fixed and do not develop much over time (Dweck et al., 1995).

Chapter 2: Review of Literature

There are significant differences in teacher effect on student achievement: the adjusted standard deviation for teacher effect on ninth grade math scores was 1.75 percentile points (Aaronson, Barrow, & Sander, 2007), and differences of 50 percentile points were found among students having a series of effective teachers versus ineffective teachers for three years (Sanders & Rivers, 1996). Although goals of teacher evaluations include the improvement of teaching and learning, there is only a loose coupling between teacher evaluations and teacher professional growth. Evaluation processes most effective in improving student achievement connect evaluations to school improvement and teacher professional development: "Meaningful teacher evaluation in schools can be an important catalyst for organizational learning and school improvement when it is linked to broader conceptions of leadership in school" (Davis et al., 2002, p. 289). To strengthen the connection between teacher evaluation processes and student achievement, the focus of evaluation systems has shifted over the last decade from the process and pedagogy of teaching to the desired outcome of teaching: student achievement.

To better capture teacher effectiveness, multiple data sources should be used in evaluations, and effective evaluators consider multiple data sources and are willing to vary data sources collected from teacher to teacher (Peterson, 2004; Piggot-Irvine, 2003). These data sources can be objective or subjective with the most common subjective data being teacher evaluation rating scales and the most common objective data being some measure of student achievement, but these data do not always correlate (Bommer et al., 1995; Morsh & Wilder, 1954; Sartain et al., 2011). For example, 91% of Chicago Public School teachers received superior and excellent ratings in 2007-2008, but 66% of

Chicago Public Schools failed to meet state accountability standards (New Teacher Project, 2009, in Sartain et al., 2011).

As a result of the perceived ineffectiveness of teacher evaluation processes to accurately measure teacher effectiveness and impact student learning, states began enacting legislation to require multiple measures of teacher performance: student achievement data and principal ratings. Among the Southern Regional Education Board (SREB) states, 7 of 16 passed legislation linking teacher evaluations to student achievement through value-added growth models and expanding performance rating systems in 2010-2011. These 16 states include Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia (SREB, n.d.). States weighted value-added data as the following percent of the overall evaluation model: Arkansas, Florida, and Maryland, 50%; Oklahoma and Tennessee, 35%; and Virginia, 40% (Dixon, 2011). North Carolina State Board of Education policy TCP-C-004 (2011) adopted a model of six standards for teachers with ratings on five standards assigned by the principal and the rating on one standard calculated based on student growth and eight standards for principals with ratings on seven standards assigned by the superintendent and the rating on one standard calculated based on student growth, but did not develop a formula for assigning an overall rating.

Although numerous studies were found examining the relationship between value-added data and teacher evaluation ratings, limited research examined other factors possibly influencing this relationship. Most research examining other factors focused on evaluation process (Tuyten, 2009), the nature of feedback given (Bryant & Currin, 1995;

Chow, Wong, Yeung, & Mo, 2002; Kerrins & Cushing, 2000; Kimball, 2002), and decisions to implement with fidelity and rate accurately (Sagona, 2012; Sartain et al., 2011; Tziner, Murphy, & Cleveland, 2002).

The online databases of The George Washington University's Gelman Library were used to search for most of the research in this literature review. Electronic databases searched included ERIC and JSTOR. These databases were searched between June 13, 2010, and July 22, 2013, for the key term "teacher evaluation." The search for "teacher evaluation" on eric.gov returned 82,370 articles. The list was narrowed using thesaurus descriptor selection, "teacher evaluation," and education level, "elementary and secondary education." Narrowing the list using the descriptor selections eliminated higher education studies, program evaluations, teacher evaluations of students, and other studies not related to K-12 teacher evaluations and provided a list of 572 studies since 2000 and 737 from 1990-1999. JSTOR was searched for "teacher evaluation" and produced 118,958 sources. Beta search was used to narrow research articles from 1997-2013 to produce a list of 3,130 articles. The researcher used study descriptions and abstracts when available to eliminate studies related to evaluation of students, evaluations of teacher candidates, student evaluations of teachers, and parent evaluations of teachers. Study dates were used to identify studies from the last 15 years. Studies included were the most cited recent and seminal studies, the most relevant to addressing principal ratings of teachers, and the most comprehensive studies on teacher evaluation. For the literature related to the theoretical frameworks, "implicit person theory" and "leadership frame" were used on eric.gov and JSTOR to search for articles related to evaluation processes. A search of ERIC database for Bolman and Deal found 71 sources. Limiting

the results to ERIC documents and academic journals for the last 15 years (1997-2012) reduced the number of articles to 46. A search for "leadership frame" on the ERIC database produced 303 hits. These studies were further filtered using the abstracts to identify studies most relevant to this project.

The organization of the literature review originated in the theoretical and conceptual frameworks described in Chapter 1. Social cognitive theory (Bandura, 1986)—rooted in reciprocal interactions among internal characteristics, external factors, and motivations and actions—provided the overarching framework for analyzing the literature. In reviewing the literature for external factors possibly influencing principal ratings of teachers, the following three categories emerged: teacher, student, and school characteristics. Internal characteristics, limited in this study to internal characteristics of the principal, include leadership decisions, expertise, leadership frames, and beliefs about the malleability of others' personal characteristics and abilities. Motivation and action in this study focused on the motivations and actions of the principal in rating teachers. The literature review is organized into the following sections: (a) a review of literature suggesting teacher characteristics related to principal ratings of teachers; (b) a review of literature suggesting student characteristics related to principal ratings of teachers; (c) a review of literature suggesting evaluation characteristics related to principal ratings of teachers; and (d) a review of literature suggesting principal characteristics related to principal ratings of teachers. The section on principal characteristics includes research related to leadership orientations (Bolman & Deal, 1990) and implicit person theories (Dweck et al., 1995).

Teacher Factors Related to Evaluation Ratings

Recent empirical research addressing evaluation processes has been dominated by case studies examining the relationship between teacher evaluation ratings and teacher value-added data. These studies found mixed results regarding the relationship between principal ratings and teacher value-added data in Chicago, Illinois; Washoe County, Nevada; and Cincinnati, Ohio, and identified potential teacher factors precipitating these inconsistent relationships (Holtzapple, 2005; Kane et al., 2011; Kimball et al., 2004; Milanowski, 2004). This section of the literature review is organized around the following teacher factors suggested by recent literature to influence principal ratings of teachers: teacher effectiveness, teacher ability, teacher experience, and teacher content area and grade level.

Teacher Effectiveness

Recent studies regarding teacher effectiveness used value-added data as a measure of teacher effectiveness. A limited number of studies used other measures such as student and/or parent surveys (Kane & Staiger, 2012; Peterson, Wahlquist, & Bone, 2000), but this review of literature focused on teacher effectiveness as measured by value-added data. This section of the literature review briefly examines concerns related to the validity and reliability of value-added measures, temporal effects on teacher value-added data, and the relationships among teacher ability, teacher experience, teacher value-added data, and principal ratings of teachers.

Statistical concerns with value-added models. Three areas of concern were identified with the Tennessee Value-Added Assessment System (Ballou, 2008a, 2008b): bias, imprecision, and inadequate models. Theoretical concerns with value-added models

include the nonrandom assignment of students to teachers may contribute to bias and imprecision. Parents' decisions on where to live and administrators' decisions on how to place students in teacher classrooms do not create random distributions of students across all classrooms or schools. Test measurement errors including alignment with curriculum, ceiling and floor effects, measurement errors, and timing of tests contribute to imprecision. Finally, value-added models do not adequately capture the complex relationships among current year schooling inputs, nonschooling inputs, and past schooling inputs (Ballou, 2008a, 2008b). There are also concerns with other measures of teacher effectiveness including observation data, surveys, and samples of students work (Ballou, 2008a).

Assumptions of statistical testing apply to value-added models. A simulation was conducted by Reardon and Raudenbush (2009) to test the assumptions of value-added models. This study attempted to control for the following assumptions not met with value-added models: (a) Every student could be potentially assigned to every school; (b) The assignment of other children to a school does not affect the impact of attending a school; and (c) A school's effect is the same for all subpopulations. This study simulated 500 schools with 500 students in each school by comparing the results of four models. Models A and B assumed each school's effect was homogeneous for all subpopulations. Model A assumed a linear relationship between potential outcomes and school effect, and Model B assumed a quadratic relationship between potential outcomes and school effect. Models C and D allowed heterogeneity of school effects. Model C assumed a random relationship between potential outcomes and school effect, and Model D assumed both linear and quadratic relationships between potential outcomes and school effects. Models

C and D more accurately estimated true school effects suggesting the importance of modeling the heterogeneity of school effects (Reardon & Raudenbush).

Temporal variability of teacher effectiveness. When using value-added models to measure teacher value-added effect, the temporal variability in teacher effectiveness should be considered. Teacher effectiveness varies from year to year, and teacher effect based on low-stakes tests is more accurate when measured over a period of years according to a longitudinal study in Florida (McCaffrey, Sass, Lockwood, & Mihaley, 2009). Florida has two types of achievement tests: the Sunshine State Standards Florida Comprehensive Achievement Test which is a high-stakes test and Florida Comprehensive Achievement Test Norm-Referenced Test (FCAT-NRT), a low-stakes test. To determine teacher effect on student achievement, data were used from the Florida Education Data Warehouse associated with teachers in five large school districts (n = 6,303, 3,512, 4,421, 4,025, and 3,801). Achievement tests were administered in third through tenth grades to allow observations of lagged achievement in fourth through tenth grades. One-third of teachers in the top quintile of class achievement on the FCAT-NRT remained in the top quintile the next year. Ten percent of the teachers in the top quintile one year fell to the bottom quintile the next year (McCaffrey et al.). The stability coefficients for elementary teachers increased from .157 to .442 with a single-year estimate to .268 to .612 with a two-year estimate. Three years of data produced stability coefficients of .55 in middle schools and .66 in high schools (McCaffrey et al.). Although teacher effectiveness varies from year to year, the results of a study addressed under "Principal Characteristics Influencing Principal Ratings of Teachers" suggest principals may not lower ratings to reflect weaker performance to build relationships (Kluger & DeNisi, 1996). Studies

addressed under "Implicit Person Theories" suggest evaluators holding an entity theory may not raise ratings to reflect improved performance (Heslin, 2003; Heslin, VandeWalle, & Latham, 2006).

Teacher ability. Three aspects of teacher ability possibly influencing principal ratings of teachers emerged from recent research. First, principals are able to most accurately rate the most and least effective teachers, but have more trouble accurately rating the 60-80% of teachers in the middle (Jacob & Lefgren, 2008). Second, principals may allow perceptions of overall teacher ability to influence ratings on all indicators (Harris & Sass, 2009). Third, principals may perceive veteran teachers as having more ability than novice teachers (Holtzapple, 2005: Sartain et al., 2011).

Principals are not consistently competent in evaluating teacher ability. Principals can more easily identify the top and bottom 10-20% of teachers but are far less able to discern differences in the effectiveness of the 60-80% of teachers in the middle. Principals based their judgments of teachers on three types of data: (a) data collected from formal and informal observations of teacher working with students and colleagues; (b) feedback from parents; and (c) student test scores. Elementary principals and teachers in a midsize district were participants in this study ($n_{principals} = 13$, $n_{teachers} = 201$). When asked to identify the most effective teachers based on their abilities to raise student achievement scores, principals were able to identify their most effective reading teachers 55% (SE = .18, p = .02) of the time and their most effective math teachers 70% (SE = .13, p = .01) of the time. They were able to identify their least effective reading teachers 38% (SE = .22, p = .19) of the time and their least effective math teachers 61% (SE = .14, p = .01) of the time. Principals were able to identify reading teachers who were above

the median according to value-added measures 62% (SE = .12, p = .01) of the time and math teachers who were above the median according to value-added measures 59% (SE = .14, p = .02) of the time. Principals were able to identify reading teachers who were below the median according to value-added measures 51% (SE = .11, p = .16) of the time and math teachers who were below the median according to value-added measures 53% (SE = .13, p = .03) of the time (Jacob & Lefgren, 2008).

Teachers perceived by the principal as effective may be rated higher on all indicators. In a study described in the section, "Teacher Content Area and Grade Level," positive and significant pairwise correlations were found among all teacher characteristics ($p < .05$). Correlation coefficients among principal ratings of teachers on overall ability, teacher ability to raise test scores, interpersonal skills, motivation/enthusiasm, works well with others, and knowledge/teaching skills/intelligence ranged from .550 to .856. Harris and Sass (2009) suggested a possible "halo effect" for teachers considered to be effective by the principal:

> It is also noteworthy that the teacher-characteristics factors are all positively correlated with one another and are often highly correlated. It is not obvious that this should be the case—for example, that teachers who are more knowledgeable would also tend to have better interpersonal skills. It is possible there is a "halo effect" whereby teachers who are rated by the principal overall are automatically given high marks on all the individual characteristics. . . (Harris & Sass, p. 18)

Finally, principals possibly perceive the ability of novice and veteran teachers differently or approach the rating of these groups of teachers differently. When Cincinnati Public Schools implemented a new evaluation process, the mean and median evaluation scores were lower for 2001-2002 than for 2000-2001 with $n = 80$ and $n = 166$. The percentage of teachers receiving a distinguished rating fell from 31 to 12 % resulting in an increase in the number of basic and proficient ratings. Some of these differences

could be attributed to an increase in the number for novice teachers, from 14% in 2000-2001 to 34% in 2001-2002 (Holtzapple, 2005). Two studies using mixed methods found some principals approached the rating of novice and veteran teachers differently. A study using mixed methods including interviews of principals (n = 7) as a follow-up to surveys (n = 98) and interviews with teachers (n = 21) found three administrators specifically referred to differences in implementing the evaluation process with veteran and novice teachers (Rowe, 2004). A second study discussed further in the subsection "Leniency in ratings" concluded principals are not likely to rate veteran teachers lower than they were rated in previous evaluations citing possible damage to principal-teacher relationships as the reason (Sartain et al., 2011). These findings are not generalizable but suggest possible factors to be considered in future studies.

Teacher Content Area and Grade Level

Valid evaluation instruments are, "those that include all criteria considered essential for effective performance, exclude criteria considered extraneous to effective performance, and weight relevant criteria in proportion to their importance" (Glickman, Gordon, & Gordon, 2010, p. 275). If effective performance is to be measured by student achievement, what other criteria should be considered as influencing the strength of correlation between teacher evaluation ratings and teacher value-added effects? Case studies of evaluation systems in Cincinnati, Ohio; Washoe County, Nevada; a Los Angeles, California, Charter School; and mid-size school districts in Florida suggest grade level and content area should be considered as variables influencing principal ratings of teachers.

Cincinnati Public Schools. Three studies of the Cincinnati Public Schools evaluation system published since 2003 identified different strengths of correlation between principal ratings and teacher value-added effect based on content area and grade level. In the first study, composite evaluation ratings by principals correlated to mean classroom gains in the four core content areas over two years, 2000-2001 and 2001-2002. The strength of the correlation varied based on content area. The highest correlation was between principal ratings and math classroom gains (r_τ = .383 and r_τ = .379, p = .001). The weakest correlation was between principal ratings and science classroom gains (r_τ = .271 and r_τ = .260, p = .008 and p = .003). Only teachers of basic content areas in Grades 3 through 8 (n = 80 teachers in Year 1, n = 166 teachers in Year 2) were included (Holtzapple, 2005).

A second study of Cincinnati Public Schools controlled for student factors possibly influencing test scores and did not find a significant relationship between teacher evaluation ratings and student achievement at every grade level. Teachers and students in Grades 3 to 8 were included in this study. A two-level hierarchical linear model was used to control for the effect of factors possibly influencing test scores: students' prior achievement, gender, race, poverty level, and days enrolled in school. Although Cincinnati Public Schools had 3,000 teachers, only 212 teachers taught third through eighth grade, had the necessary evaluation scores, taught tested subjects, and had the necessary test scores. For this study, the measure of average student performance was calculated using Empirical Bayes Intercept Residuals. Some findings were not significant. This lack of significance could be caused by the small number of teachers who could be included (i.e., for eighth grade math n = 16.). Correlations between teacher

evaluation ratings and mean student performance ranged from .03 in fourth grade reading to .45 in third grade reading, from .20 in seventh grade math to .57 in seventh grade math, and from -.01 in seventh grade science to .43 in sixth grade science. Milanowski (2004) combined grades and found moderate correlations with a moderate to large effect size between Empirical Bayes Intercept Residuals measure of student achievement and total teacher evaluation system scores to be .32 for reading (d = .68, p < .001), .43 for math (d = .95, p < .001), and .27 for science (d = .56, p = .003.)

In a third study of Cincinnati Public Schools' teacher evaluation system, there was a correlation between nonprincipal evaluator ratings and student achievement but the strength of the correlation varied from reading to math. Using evaluation data for Cincinnati from 2000-2001 to 2008-2009, researchers used evaluator rankings on each specific element to calculate a score for the eight teacher effectiveness standards. Mean scores were then calculated for each teacher's overall classroom practices, classroom management versus instructional practice, and questions/discussion versus standard/content. In this sample of 365 reading and 200 math teachers, a one point increase in Overall Classroom Practices score created an increase of 1/7 of a standard deviation in student reading achievement and 1/10 of a standard deviation in student math achievement. In this study, external observers were used, trained, and assessed in their ability to score (Kane et al., 2011).

Washoe County School District. The next study found differences in the strength of the correlation between teacher evaluation rating and student achievement data based on grade and content area taught. The Washoe County School District implemented a standards-based evaluation system designed to promote discussion,

feedback, and reflection and to serve administrative purposes including contract renewal and tenure decisions. Kimball et al. (2004) used multilevel statistical modeling to estimate random effects of classrooms on student achievement and hierarchical linear modeling (HLM) to predict student posttest scores based on demographics and pretest scores and to estimate the effect of teachers with higher or lower evaluation scores. The results were mixed. In third grade (n = 123), the relationships between teacher evaluation scores and student achievement in reading, math, and reading and math were not significant with p = .287 in reading, p = .254 in math, and p = .289 in combined reading and math. In fourth grade (n = 87), only the correlation between teacher evaluation scores and fourth grade reading achievement was significant with a 5.41 increase in reading achievement for every one-point increase in evaluation score (p = .01). There were statistically significant relationships between teacher evaluation scores and reading, math, and combined student achievement in fifth grade (n = 118). For every one point increase in teacher evaluation scores, student performance increased 12.7 points in reading, 20.1 points in math, and 16.3 points for combined reading and math (p = .001, p < .001, and p < .001). Kimball et al. (2004) suggested the following reasons for the weakness of some relationships: only 7 of 23 components, the performance composite, of the evaluation system were used; a potential lack of alignment between curriculum, assessments, and evaluation standards; the second grade test used to measure third grade achievement had a nonnormal distribution; and the evaluation system was a low stakes system.

Vaughn Elementary School. Vaughn Elementary School, now Vaughn Next Century Learning Center, a charter school in the Los Angeles Unified School District,

implemented an innovative teacher evaluation system with subject-specific standards in 1998-1999 at the same time they implemented a pay for knowledge and skills plan. There were statistically significant classroom effects in each content area with the following effect sizes: reading (ρ = .18), math (ρ = .23), language arts (ρ = .22), and composite (ρ = .27) (Gallagher, 2004). Although the classroom effect in reading was smaller, the correlation between classroom effect in reading and teacher evaluation score was .50 (p = .01, two-tailed.) The correlation between composite classroom effect and teacher evaluation score was .36 (p = .05, two-tailed). The relationships between classroom effect/math and classroom effect/language arts were not significant. Gallagher suggested the reading standards might be better aligned to reading achievement expectations than the math and language arts standards. Gallagher used hierarchical linear modeling to predict the relationship between teacher evaluation scores and value-added data and found teacher composite evaluation scores explained 13% of classroom effect on student composite achievement (t = 2.19, p < .05). Teacher evaluation scores in literacy explained 34% of the classroom effect on student achievement (t = 3.71, p < .001). Relationships for math and language arts achievement were not significant.

Florida school district. In a midsize Florida school district, principal ratings of elementary teachers were significantly and positively correlated to teacher value-added effects on student achievement in math and reading, but a one-point difference in principal rating was correlated with two times the value-added effect in math as in reading. In middle schools and high schools, there was no significant relationship between principal ratings and teacher value-added effects on student achievement. Principal experience in administration and in working with a particular group of teachers

was significantly related to the principal's ability to accurately rate teachers. The sample size of principals was relatively small with $n = 25$ for overall rating and $n = 23$ for ability to raise test scores (Harris & Sass, 2009).

There were correlations between principal ratings of teachers and teachers' value-added data. A teacher with an overall rating one point higher raised student math test scores 2.374 scale score points per year more than the average teacher ($t = 4.50$; $p < .01$, two-tailed test; $R^2 = .34$) and raised student reading scores 0.858 scale score points ($t = 3.26$; $p < .01$, two-tailed test; $R^2 = .06$). A teacher with a rating for *ability to raise test scores* one point higher raised student math tests scores 2.199 scale score points per year more than the average teacher and raised student reading test scores 0.845 points per year more than the average teacher ($t = 3.83$ and $t = .845$; $p < .01$, two-tailed test; $R^2 = .13$ and $R^2 = .06$). There was a stronger correlation between overall ratings and teacher *ability to raise test scores* than principal ratings of teacher *ability to improve test scores* and their actual ability to raise test scores with a stronger effect size for overall rating than ability to raise test scores rating. Differences in significance of correlation and strength of correlation were found across grade spans and content areas. Excluding elementary teachers, the ordinary least squares were greater for *teacher ability to raise test scores* rating. For elementary teachers who were rated one point higher overall, the ordinary least squares estimate of the determinants of teacher effects was 2.956 for math and 1.072 for reading ($t = 4.91$ and $t = 3.53$; $p < .01$, two-tailed test; $R^2 = .13$ and $R^2 = .07$). For elementary teachers who were rated one point higher for "ability to raise test scores" the ordinary least squares estimate of the determinants of teacher effects is 2.967 for math and 1.21 for reading ($t = 4.51$ and $t = 3.52$, $p < .01$, two-tailed test, $R^2 = .15$ and $R^2 = .08$).

For middle and high school teachers none of the *t* tests were significant. There was a high correlation (.70) between principal ratings of teachers and principal ratings of teacher ability to improve student test scores (Harris & Sass, 2009).

Summary

Principal ratings of teachers are influenced by characteristics of the teacher and characteristics associated with the teacher because of the students they teach. Value-added models are being incorporated into teacher evaluation instruments based on the belief that coupling student gains with the evaluation process will create teachers who are more effective. Although there are statistical concerns with value-added models, prevailing policy makers believe the data yielded by value-added models outweigh these statistical concerns. One concern that has been explicitly addressed in evaluation models is the temporal variability in teacher effectiveness. McCaffrey et al. (2009) found that using three years' data versus one years' data greatly increased the stability coefficient from as low as .157 to as high as .66. This variability in ratings coupled with a tendency for principals to not assign lower ratings to veteran teachers than they had previously received (Holtzapple, 2005; Rowe, 2004; Sartain et al., 2011) would result in higher mean ratings of teachers. Because this study used the principal as the unit of study, the percent of teacher with less than three years' experience and the school teach turnover rate were included as exploratory variables.

Principals are able to more easily identify the top and bottom 10-20% of their teachers. They are less able to distinguish among the middle 60-80% of teachers (Jacob & Lefgren, 2008). When they identify a teacher as a "good" teacher, they are more likely to assign these teachers higher ratings on all standards (Harris & Sass, 2009). Choosing

to rate "good" teachers higher on all standards results in a higher mean rating and could result in less variability in ratings.

Finally, the grade level and content area of the teacher influence principal ratings of teachers with different correlations between teacher evaluation ratings and student gains in math, reading, and science across third through eighth grades. Although most studies found statistically significant relationships across grade levels and content areas, R^2 statistics ranged from -.01 to .43. Although this study used the principal as the unit of study and did not address grade level and subject area, future studies should include nested models to examine this phenomenon. Because this study used the principal as the unit of study, only grade span of the school was included.

Student Factors Influencing Evaluation Ratings

Student characteristics cause statistical concerns in measuring teacher effectiveness. As described earlier in this literature, the nonrandom distribution of students across schools and districts creates difficulties in developing models able to account for schooling inputs, nonschooling inputs, and past schooling inputs (Ballou, 2008a; Ballou, 2008b). This section of the literature review examines research surrounding the influence of student demographics and prior achievement on principal ratings of teachers.

Kimball et al. (2004) found mixed results in Washoe County School District when controlling for student demographics and pretest scores among third, fourth, and fifth grade teachers (n = 123, 87, and 118). In this study the relationships between teacher evaluation scores and student achievement were not significant for any third grade test, but were significant for fifth grade reading, math, and combined reading and math (p =

.001, $p < .001$, and $p < .001$). In fourth grade, the only significant relationship was between teacher evaluation score and reading achievement ($p = .01$). In Harris and Sass's (2009) study discussed in the previous section, the correlation between principal ratings of teachers and teachers' value-added effect was significant, but for middle and high school the relationships were not significant. The results of these studies suggest student demographics and pretest scores or prior achievement moderate the correlations between principal evaluation ratings and teacher effect on student achievement. These two studies were also addressed under "Teacher Factors Influencing Evaluation Ratings."

Studies suggest principals can interpret high student achievement levels as teacher effectiveness. Because there is not a normal distribution of students among teachers or among schools, teachers with higher concentrations of lower achieving and economically disadvantaged students can be rated as less effective than teachers with lower concentrations of lower achieving and economically disadvantaged students with the same value-added effect (Borman & Kimball, 2005; Jacob & Lefgren, 2008). Principal ratings can be more closely correlated to student achievement levels than to teacher value-added effect (Jacob & Lefgren). When principals in a mid-size school district in the western United States were asked to rate teachers' abilities to raise student achievement, there was a higher correlation between their ratings and student achievement levels than between their ratings and value-added measures. After adjusting for estimation errors, the correlations of principal ratings of teachers' effectiveness at raising student achievement to value-added data were .29 for reading and .32 for math ($p < .05$). The correlation of principal ratings to average student achievement after adjusting for estimation errors was .55 for reading and .37 for math ($p < .05$) (Jacob & Lefgren).

This study also found differences in principal ability to discern the effectiveness of teachers in the middle 60-80% as described in the section, "Teacher Ability."

In a study comparing 400 teacher evaluation ratings to the concentration of students of poverty, minority students, and students with lower pretest scores, teachers with lower ratings were assigned higher concentrations of students of poverty, minority students, and students with lower pretest scores. Classrooms were separated into classes with concentrations of students of poverty above and below the sample mean as well as into classes with concentrations of minority students above and below the sample mean. Evaluation ratings of teachers with low concentrations of children of poverty (n = 214, M = 2.71, SD = .39) were higher than ratings of teachers with high concentrations of children of poverty (n = 158, M = 2.48, SD = .48) at a significant level, $t_{370} = 5.07$, $p <$ .001. Evaluation ratings of teachers with low concentrations of minority children (n = 206, M = 2.74, SD = .37) were higher than teachers with high concentrations of minority children (n = 166, M = 2.47, SD = .49) at a significant level ($t_{370} = 6.13$, $p < .001$.) Evaluation ratings of teachers with higher concentrations of students with low reading or math pretest scores (n = 178, Reading M = 2.48, Reading SD = .48; Math M = 2.49, Math SD = .49) were lower than teachers of students with higher concentrations of students with high pretest scores (N = 191, Reading M = 2.74, Reading SD = .37; Math M = 2.74, Math SD = .37) at a significant level (Reading $t_{367} = -5.78$, Math $t_{367} = -5.46$, $p < .001$.) Teachers with higher ratings were assigned classes with higher achieving students, but there was no evidence teachers with higher evaluation ratings closed achievement gaps (Borman & Kimball, 2005). Borman and Kimball suggested possible explanations could be the influence of student attributes in the classroom on evaluator perceptions, the

limited diversity of students within classes, and school attributes such as organizational capacity and professional cultures.

Principals are not consistently integrating situational factors influencing teacher and student actions in the classroom into their ratings. When controlling for student demographics and student prior achievement, the relationships between teacher value-added effect and principal ratings of teacher are mixed (Kimball, 2004). Teachers with the same value-added effect but higher concentrations of low achieving or economically disadvantaged students are rated lower than teachers with lower concentrations of low achieving or economically disadvantaged students (Borman & Kimball, 2005; Jacob & Lefgren, 2008). Implicit person theory research described later in this literature review suggest principals holding an entity theory will be less likely to consider situational factors in rating teachers (Molden, Plaks, & Dweck, 2006). When operating under cognitive load, entity theorists will be even less likely to consider the situational factors (Molden et al., 2006).

Evaluation Process Factors Influencing Evaluation Ratings

Factors associated with the evaluation process can influence the correlation of principal ratings of teachers and teacher value-added effects. A phenomenon called the *widget effect* was termed after studying the ineffectiveness of evaluation processes in distinguishing between effective and ineffective teachers (Weisberg et al., 2009). This section of the literature review examines three factors associated with evaluation processes potentially influencing principal ratings of teachers: (a) the purpose of the evaluation; (b) the standards and instrument; and (c) the available resources.

Purpose

Data collected through the evaluation process are used for different purposes. Some theoretical literature attributes two purposes to teacher evaluation processes: accountability and professional development; others describe these purposes as summative and formative (Duke & Stiggins, 1990; Wise et al., 1985); and other literature adds role definition as a third function (Peterson & Peterson, 2006; Youngcourt, Leiva, & Jones, 2007). This section of the literature will examine the expansion of evaluations to include formative purposes, empirical literature examining the influence of summative and formative purposes on evaluation, and the influence of competing purposes on evaluations.

Historical shift to include formative purposes in theoretical literature. The earliest teacher evaluations in the United States were simple checklists completed by community members with explicit criteria such as if the school had a building with a teacher and desks (St. Maurice & Cook, 2005). By the mid-19th Century, supervision began to focus on instruction as illustrated by the following statement in the *Annual Report of the Superintendent of Common Schools of the State of New York (1845),* "Instruction is the primary object of visitation" (as cited in Marzano, Waters, & McNulty, 2005, The early days of supervision section, para. 5). These early evaluation systems focused on summative purposes. Cubberley (1916) described a differentiated, formative supervisory model with a supervisor who builds positive relationships with teachers and provides constructive feedback:

> If the schools in any city are to render good service, there must be plenty of close, personal, and helpful supervision of the instruction corps. . . Even the reasonably well-trained normal-school graduate requires much help at first to adjust herself properly to the work of a city school system, and to enable her to, in the course of

> four or five years, to reach a maximum of efficiency with a minimum of mistakes and struggles. (Cubberley, 1916, p. 240)

Despite Cubberley's recommendations that supervision include a formative purpose particularly for novice teachers, the checklists and ratings scales that dominated evaluation processes for half a century, according to the literature review completed by Morsh and Wilder (1954), were not successful in meeting formative purposes.

A movement toward using clinical supervision models in teacher evaluation started in the 1950s culminating in a five-phase process developed by Goldhammer (1969) and a set of objective criteria and seven-step lesson plan developed by Hunter (1984). The purpose of the clinical supervision model was for supervisors to visit classrooms and provide feedback to help teachers improve their instruction (Sergiovanni, 2002). Hunter provided a description of expected outcomes to supervisory conferences with a focus on formative purposes: apply research to observed instructional behaviors, encourage teachers to consider new approaches, help teachers identify ineffective or less effective aspects of teaching, base evaluations on evidence not opinion, and promote teacher growth (Hunter, 1984; Marzano et al., 2005). The reforms of the 1980s and 1990s focused on teacher professionalism with more complex developmental or reflective models. These reformed instruments called for differentiation within the evaluation systems to promote teacher growth (Darling-Hammond, 1990; Marzano et al., 2005; Milanowski, 2004).

The economic crisis beginning in 2008 allowed the federal government to step in and encourage additional reforms to education systems through the investment of $787 billion. Race to the Top (2009) called for teacher evaluation systems incorporating multiple rating categories and taking student growth data into account, and flexibility

requests related to No Child Left Behind Act's (2001) all or nothing requirements required evaluation systems that supported effective teachers and leaders. Evaluation systems were required to integrate formative purposes including promoting continual improvement of instruction and providing clear, timely, and useful feedback guiding professional development.

In the policy brief, *Generating Teaching Effectiveness: The Role of Job Embedded Professional Learning in Teacher Evaluation,* the view of teacher evaluation as a professional development tool is described,

> Although the federal and state policies and associated guidance continue to refer to teacher learning as "professional development" rather than "job-embedded professional learning," the spirit of the policies is clearly directed toward harnessing teacher evaluation for the continuous improvement of teaching effectiveness through the provision of evidence-based feedback to teachers. (Coggshall et al., 2012, p. 2)

Empirical research. The historical shift to use evaluation for formative purposes and the necessity for evaluation systems to continue to fulfill summative purposes sometimes come into conflict. This section of the literature review will examine empirical research related to summative purposes, formative purposes, the dissonance between the two purposes, and the potential for formative purposes to influence principal ratings of teachers.

Summative purposes. Data collected on summative purposes of evaluation instruments have focused on descriptive statistics related to the distribution of ratings, comparative studies of different evaluation systems, and frequency of use of instruments for varying purposes. Although teachers are seldom dismissed for poor performance, evaluations are at times used for this administrative function. Based on school administrators' responses to the Schools and Staffing Survey, by 2003-2004, public

school districts dismissed a mean of 1.2 teachers with less than three years' experience and 1.9 teachers with more than three years' experience (National Center for Education Statistics, 2004). In 2007-2008, data was reported by tenured versus nontenured, and the mean number of non-tenured teachers dismissed based on poor performance was 1.4, but the mean number of tenured teachers dismissed based on poor performance was 3.0. In North Carolina, the mean number of non-tenured teachers dismissed was 1.5 and the mean number of tenured teachers dismissed was 3.0 (National Center for Education Statistics, 2009). In 2011-2012, a mean of 1.0 tenured teaches and a mean of 0.2 nontenured teachers were dismissed (National Center for Education Statistics, 2012). Despite concerns with teacher quality, the number of teachers actually dismissed per district has been a relatively small percent of the mean number of teachers per district (*M* = 187, *SD* not reported) (NCES, 2012).

Among 12 districts across four states, all districts used teacher performance for remediation and dismissal. Only 3 of the 12 districts used teacher performance in making other decisions: Denver, Colorado, used performance in compensation; Cincinnati, Ohio, used performance in hiring and placement; and Toledo, Ohio, used performance for granting tenure (Weisberg et al., 2009). This study was based on surveys of 15,000 teachers and 1,300 administrators and termed the indifference to teacher effectiveness the "widget effect" and suggested the following concerns with evaluation systems: (a) 99% of teachers received satisfactory ratings; (b) Effective teachers were not recognized and rewarded; (c) There was only a loose coupling between evaluations and professional development with 73% of teachers responding no areas of improvement were identified

in their evaluations; (d) Instruments do not adequately capture novice teacher performance; and (e) Poor performing teachers are not dismissed (Weisberg et al., 2009).

Formative purposes. The idea of using evaluations to improve instruction is not new. Literature in the mid-19th century described classroom visits by supervisors with instruction as the primary purpose of the visit (Marzano et al., 2005). Cubberly (1916) described the need for a differentiated model of supervision that recognized the need for teacher growth among teachers in their first four to five years of teaching. Whitehead (1952) examined teacher perceptions of evaluation purposes through surveys of 115 teachers at African American schools representing 69 of the 100 counties in North Carolina regarding their perceptions of the evaluation process. Among the teachers surveyed, 80% of the teachers believed the primary purpose of administrator evaluations was to improve instruction, and 20% believed the primary purpose for administrator evaluations was to rate teachers and inspect the physical environment. No tests were used to determine the significance of the findings; no effort was made to address representative sampling. These findings are included due to historical significance. Most recent studies have examined the roles of formative and summative purposes in teacher evaluation.

Formative versus summative purposes. Most empirical research related to formative and summative purposes of teacher evaluations has focused on the influence of formative versus summative nature of the process on the teacher or principals. The section will look at empirical literature related to formative and summative purposes and different models, lack of alignment of purposes, and dual purposes in using evaluation instruments.

Formative versus summative models. Teachers perceive some models of standards more appropriate for formative evaluations and others as more appropriate for summative evaluations (Kyriakides, Demetriou, & Charalambous, 2006). Forty-two standards for teacher evaluations were gleaned from teacher effectiveness research and combined into seven models for evaluating teachers. These seven models were implemented and studied in Cyprus based on teacher perceptions of each model's primary purpose (Kyriakides et al., 2006): (a) the Goals and Tasks Model; (b) the Resource Utilization Model; (c) the Working Process Model; (d) School Constituencies Satisfaction Model; (e) the Accountability Model; (f) the Absence of Problems Model; (g) the Continuous Learning Model; and (h) the School Constituencies Model. Teachers ($N = 237$) were asked to rate the appropriateness of each model for formative and summative purposes. The paired *t*-test was used to determine if the Cypriot teachers did or did not consider the models equally important for summative and formative purposes. Teachers ranked four models as significantly more important for formative evaluations than for summative purposes: (a) Working process model ($t = 7.1$, $df = 231$, $p < .001$); (b) Goal and task model ($t = 6.32$, $df = 231$, $p < .001$); (c) Partners' and employers' satisfaction model ($t = 4.60$, $df = 231$, $p < .001$); and (d) Resource utilization model ($t = 3.98$, $df = 231$, $p < .001$). For the remaining models, there was not a statistically significant difference in teacher perceptions of appropriateness for summative or formative purposes. This study supports the ability of teachers to distinguish the appropriateness of various models for formative and summative purposes and suggests teachers might rank evaluation processes with which they are comfortable higher (Kyriakides et al., 2006). Although this study does

not examine principal perceptions of different models, the findings suggest characteristics of the evaluation instrument might encourage the use of one purpose over the other.

Negative impact of lack of alignment. Lack of alignment among purposes for evaluations can impede the effectiveness of evaluation processes (Bosetti, 1994). Bosetti examined espoused theory versus theory in action in Alberta, Canada, and found the evaluation policy tried to meet too many political, bureaucratic, and administrative goals. Data were collected from surveys of 29 teachers from two schools. Although the stated purposes of the evaluation system were formative, professional growth for teachers and improved student performance, the evaluation system was used primarily for summative tasks: teacher certification, personnel decision-making, and contractual decisions. There was little connection between the professional growth goals intended by the policy and the actual goals of the evaluation system as implemented except among new teachers (Bosetti).

Dual purposes of instruments. Although teachers have expressed concern with assigning dual purposes to single instruments (Bradshaw, 2002; Kimball, 2002), Piggot-Irvine (2002) recommended an approach integrating formative and summative purposes. Using the same instrument for formative and summative purposes might impede the effectiveness of evaluations for formative purposes. Bradshaw (2002) conducted a quantitative study of 21 northeastern North Carolina school districts regarding the implementation of the Teacher Performance Appraisal System. Surveys based on Stiggins and Duke's (1988) teacher evaluation profile were returned by 4,092 teachers and 177 evaluators, a responses rate of approximately 60%. Most respondents believed the evaluation process positively impacted school improvement goals, school climates,

teacher quality, and student achievement, but few teachers believed the evaluation process had any impact on their own teaching. Concerns with the process included the use of one instrument for formative and summative purposes (Bradshaw, 2002).

Administrators use evaluation processes to provide feedback to teachers on performance (Kimball, 2002). Kimball studied the implementation of standards-based evaluation systems in three districts through interviews of six evaluators in each district, 19 teachers in one district, and 18 teachers in two other districts. Two of the districts attempted to separate formative and summative purposes in their evaluation systems, but these functions could overlap. For example, teacher portfolios used in their self-appraisal process could also be used as evidence in their evaluations. In interviews, teachers in this district expressed concerns about the different purposes of the evaluation. The third district did not attempt to separate the processes but created a dual purpose system, but teacher responses reflected an acceptance of the dual purposes, "I felt it was for growth, I guess for accountability too, but more for growth…" (Kimball, p. 258). Teachers recognized formative purposes in the instruments and received feedback on instruction. Teachers perceived the feedback as generally affirmative in nature and lacking specific suggestions to improve instruction. Administrators were able to apply provide feedback on teacher practices including pointing out suggestions regarding classroom management, student grouping, student engagement, and record keeping. According to teachers, administrators evaluating teachers in unfamiliar content areas can limit the depth and quality of feedback.

Effective evaluation systems integrate formative and summative approaches (Piggot-Irvine, 2003). In a synthesis of data from three previous parallel studies, Piggot-

Irvine (2003) concluded the following characteristics related to purpose were necessary for effective evaluation systems: an integrated formative and summative approach highlighting areas for future development as well as strengths and weaknesses; and separation of disciplinary and formative evaluation processes. The three studies included a four-year study involving an average of 70 appraisers each year to monitor the effects of requirements and training for evaluation processes in New Zealand; an analysis of the training components for 219 evaluators from 25 schools; and an evaluation of eight meetings regarding action research.

This section focused on formative and summative purposes of the evaluation instrument. Literature examining the influence of principal use of these purposes is examined further under "Principal Characteristics Influencing Principal Ratings of Teachers."

Standards and Instruments

In her book *Teacher Evaluation to Enhance Professional Practice,* Danielson (2000) wrote that standards should (a) provide clear criteria based on effective teaching research; (b) define inputs or acceptable teacher tasks; and (c) define outputs or teacher effects particularly on student learning. Limited empirical research has been completed in the last 15 years addressing Danielson's premise. This gap could be due to an assumption in teacher evaluation literature that criteria, standards, or expectation are clear and relevant (Kyriakides, et al., 2006; Piggot-Irvine, 2003; Weisberg et al., 2009; Ovando & Ramirez, 2007; Kimball, 2002). The validity and reliability of the evaluation standards influence the validity and reliability of the evaluation process. Glickman et al. (2010) described reliable instruments as instruments using low-inference indicators such as "uses

examples when explaining" (p. 275), not high-inference indicators such as "teacher clarity" (p. 275). Indicators can be more or less valid depending on the level of inference required by the evaluation (Glickman et al.)

The content of standards affect principal ratings of teachers, but sometimes principals feel the standards do not accurately capture what makes effective teachers (Glickman et al., 2010; Sangora, 2012; Weisberg et al., 2009). Standards can fail to capture the soft skills principals agree are necessary for effective teachers or fail to capture skills needed for particular grade levels or subjects. In interviews of six middle school principals, the principals all agreed evaluation instruments did not adequately capture student-teacher relationships. Their comments included the following: "Being an effective teacher is some amalgamation of building a personal relationship with kids that centers on their learning and mastery of whatever it is that they're assigned to master" (Sangora, p. 93), and "You have to love kids first to be a great teacher" (Sangora, p. 93). Principals described the difficulties in quantifying relationships, attitude, and with-it-ness (Sangora).

Not only do standards fail to capture soft skills, standards can also fail to capture the skills needed across all grade levels and content areas. In studies examined in other sections of this literature review, standards were perceived by teachers and administrators as more valid for certain grade levels or content areas and showed stronger correlation with student achievement in certain grade levels or content areas (Gallagher, 2004; Kimball et al., 2004; McCaffrey et al., 2009; Medley & Coker, 1987). Standards also affect evaluators' attitudes. In a study of supervisors from numerous fields including academic, business, and industrial, supervisor opinions toward the evaluation instrument

affected the variability of their ratings across ratees and across dimensions suggesting principals' attitudes toward the standards might affect their ratings of teachers (Tziner, et al., 2002). This study is addressed further under "Principal Factors Influencing Evaluation Ratings."

Available Resources

Organizations need resources to survive and grow. Morgan's (2006) image of organizations as organisms stressed the importance of the relationship between an organism and its environment. This relationship includes recognizing the needs of the organization, acquiring the resources to meet these needs, and adapting to changes in the environment. Evaluation processes within an organization need resources to be effective. The Personnel Evaluation Standards address the following standards related to resources and teacher evaluations: appropriate policies and procedures; clear, timely, accurate, and germane reports; professional development for users with the ultimate goal of meeting the needs of students; efficient use of time and resources; and adequate funding (Joint Committee on Standards for Educational Evaluation, 2014).

Most empirical studies related to available resources in implementing an evaluation system have focused on the principals' perspective regarding the lack of available resources necessary to successfully implement the process. Studies previously addressed in the literature review have identified time and training as resources needed to implement new evaluation processes (Kimball, 2002; Piggot-Irvine, 2003; Weisberg et al., 2009; Wise et al., 1985). In Kimball's case study of three districts, the new evaluation system required more observations and evaluation conferences than had previously been required. Piggot-Irvine, Kimball, and Weisberg et al. concluded

sufficient time for training should be provided in implementing a new evaluation system. Wise et al. concluded for evaluation systems to be successful sufficient resources including time and training must be provided (Wise et al.).

Barriers to effective evaluation processes include the lack of the following resources: time, training, and teacher support and readiness (Painter, 2000; Rowe, 2004). Painter (2000) surveyed elementary and middle school principals ($N = 781$) and identified two barriers to the successful implementation of a teacher evaluation system clearly related to lack of available resources: 56.8% of principals identified lack of time, and 27.0% of principals identified lack of sufficient training and resources. Two other barriers had implications for principal ratings of teachers: staff morale (18.9%) and emotional strain and principal isolation (18.9 %). These barriers are addressed further under the section, "Principal Factors Influencing Evaluation Ratings of Teachers."

Rowe (2004) conducted a four-part study by collecting data from 90 teacher surveys, 21 teacher interviews, 7 principal interviews framed by the themes of the teacher interviews, and 4 school vignettes. All individuals worked in the same Northwest School District in Canada. The results of this study identified the following possible barriers to implementing effective teacher evaluation systems: lack of participant readiness, limited time and opportunities for professional development, lack of administrator accountability, other administrative demands, individual differences such as teacher comfort levels with being evaluated, teacher ability to understand the process itself, and teacher trust and rapport with the evaluator (Rowe). This study included similar barriers to the Painter (2000) study: time and training. Differences in other top barriers could be due to a difference in teacher perspective (Rowe) and principal perspective (Painter) as well as the

focus in the Painter study on using the evaluation process specifically for low performing teachers.

Another study surveyed principals in Wyoming ($N = 143$). One part of the instrument asked principals to identify their frustrations with the evaluation process in Wyoming. The researchers applied quantitative methods and coded the responses for themes. The following frustrations emerged: principals in Wyoming were frustrated with lack of time, outdated instruments, and teacher willingness to change (Range, Sherz, Holt, & Young, 2011). Although no empirical research was found directly addressing available resources and principal ratings of teachers, studies on interrater reliability suggest training increases the reliability of principal ratings of teachers (Sartain et al., 2011). Studies on evaluation system implementation suggest available resources affect principals' attitudes toward the evaluation system, and principal attitude affects principal ratings (Painter, 2000; Range et al., 2011; Rowe, 2004). This relationship is addressed further under the subheading "Training" in the following section on "Principal Factors Influencing Evaluation Ratings."

Summary

Evaluation instrument and system characteristics influencing the rating process include competing formative and summative purposes, perceptions that some instruments are more relevant for one purpose over the other, and standards considered by some participants to better reflect effective teaching than others (Glickman et al., 2010; Sangora, 2012; Weisberg et al., 2009). Although this study only examines one evaluation system, the North Carolina teacher evaluation process, the instrument and principal perceptions of the instrument could influence principal ratings. The next

section “Principal Factors Influencing Evaluation Ratings” addresses the potential for characteristics of the evaluation system to influence principals in rating teachers.

Principal Factors Influencing Evaluation Ratings

The individuals involved in the evaluation process influence the effectiveness of the process. “Although considerable attention has been paid to the need to evaluate teachers to determine their effectiveness, far less attention has been paid to ensuring principals are prepared to conduct the evaluations” (National Governors Association, 2011, p. 1). Principal effect on the evaluation process can be influenced by principal leadership decisions and behaviors, previous experiences and competence, leadership frame, and perceptions of others abilities (Bryant & Currin, 1995; Chow et al., 2002; Kerrins & Cushing, 2000; Kimball, 2002; Wise et al., 1985).

Leadership Behaviors

Although numerous models for instructional leadership exist including those models developed by Hallinger (1983), Leithwood (1992), and Marzano, Walters, and McNulty (2005), the core of instructional leadership is making decisions focused on improving teaching and learning. These decisions influence the effectiveness of teacher evaluation processes (Leithwood, Louis, Anderson, & Wahlstrom, 2004). Although no empirical studies were found addressing leadership style and principal ratings of teachers, studies have addressed the influence of principal leadership style on their behaviors related to teacher evaluations. This section of the literature review focuses on specific decisions made by leaders in implementing evaluation processes that influence the outcomes of the process and the professional relationships established by the principal with teachers. Principals decide if they will implement the evaluation process with

fidelity, if they will rate more or less leniently, or if they will use evaluation processes primarily for summative or formative purpose.

Implement with fidelity. The beliefs, attitudes, and commitment of evaluators from a variety of settings influenced their rating levels, discrimination among ratees, and discrimination among dimensions (Tziner, et al., 2002). This study included the following samples: 70 military cadets, United States; 40 utility workers, United States; 121 academics, Canada; 36 managers, Canada; 30 manufacturing evaluators, Canada; 40 managers, Israel; and 40 bankers, Israel. Three dimensions of the ratings were examined: rating levels, discrimination among ratees, and discrimination among dimensions. Rating level was the mean rating for each rater; discrimination among ratees was the standard deviation of ratings for each rater; and discrimination among dimensions was the variance in the standard deviation on each rater's mean rating for each dimension. Structural equation modeling was used to create models for rating level, discrimination among ratees, and discrimination among dimensions using the following variables as the coefficients: attitudes toward the organization including organizational commitment; beliefs about the appraisal system including purpose and freedom from political influence; and orientation to the appraisal process including comfort level and self-efficacy. The first model found positive relationships among rating level, attitudes toward the organization, and beliefs about the appraisal process but found a negative relationship between rating level and orientation to the appraisal process. The second model found negative relationships among discrimination among ratees, attitudes toward the organization, and orientation to the evaluation process, but found a positive

Table 3

Structural Coefficients Linking Attitudes, Beliefs, and Orientation to Ratings

	Rating Level	Discrimination Among Ratees	Discrimination Among Dimensions
Attitudes toward the organization	.16	-.31	.15
Beliefs about the appraisal process	.20	.46	-.40
Orientation to the appraisal process	-.27	-.21	.50
R^2 for the structural model	.07	.12	.22

Note: From "Relationships between Attitudes toward Organizations and Performance Appraisal Systems and Rating Behavior" by Tziner, et al., 2002, p. 235.

relationship between discrimination among ratees and beliefs about the appraisal process. See Table 3 above for structural equation coefficients and R^2 of each model.

Leniency in ratings. The concern over discrimination among ratees and discrimination among dimensions has been addressed in limited studies related to educational leadership. Sartain et al. (2011) described the tendency of principals to rate all teachers higher or lower. Some principals rated all teachers within their building higher or lower, but their ratings agreed with external observers on who were the best and who were the worst teachers.

Principals can rate leniently to preserve or build relationships with teachers. Sartain et al. (2011) interviewed principals involved in implementing the new evaluation system in Chicago Public Schools. The study included observations to address reliability ($n_{\text{teachers}} = 257$) and validity ($n_{\text{teachers}} = 501$); relationship between evaluation ratings and value-added models in math ($n_{\text{teachers}} = 340$); a pilot group and control group each with $n = 37$; interviews of principals ($n = 39$) and teachers ($n = 26$); principal focus groups ($n = 23$); and case studies of eight schools that included 8 principal interviews, 8 assistant principal interviews, and focus groups involving 18 teachers. In this mixed methods

study, 52% of teachers received at least one distinguished rating, the highest rating, from the principal. In interviews, principals recognized they rated teachers who had previously received distinguished ratings as distinguished to "preserve relationships" (Sartain et al., p. 15). One principal stated, "I am not going to get in a big fight between these two things (Proficient versus Distinguished) because what good does it do? You just ruin your relationship with the teacher. It is much better to coach them. . ." (Sartain et al., p. 14-15).

In the dissertation, *An Examination of Principal Consistency in Evaluating Teachers* (Sagona, 2012), interviews of six middle school principals found concerns similar to Sartain et al. (2011) regarding the influence of principal-teacher relationships on teacher ratings. Some principals struggled with discussing unsatisfactory ratings with teachers and separating their personal feelings from their professional responsibilities. Principals agreed teachers did not view unsatisfactory ratings as an opportunity for professional growth but viewed low ratings as failure (Sagona). Principal purpose in evaluating teachers can influence leniency in ratings. Principals who wish to build relationships in order to coach teachers tend to be more lenient, but these lenient ratings tend to have weaker correlations to student achievement data (Kluger & DeNisi, 1996).

Formative or summative purposes. Principals' decisions to use evaluation processes for formative versus summative purposes, to distinguish among teachers, and as a multiyear growth tool can influence principal ratings of teachers. Cleveland, Murphy, and Williams (1989) described some principals' (N = 106) use of ratings systems as a tool to identify individual teacher strengths and weaknesses instead of identifying the best and worst teachers. Some principals used evaluation processes to

identify which standards are the strengths and weaknesses of individual teachers. In this case higher and lower ratings were relative to the teacher and not comparable to other staff members. Principals varied in the degree they used evaluations to distinguish between teachers versus the degree they use evaluations to identify strengths and weaknesses of each teacher (Cleveland, et al.).

Building grade span can influence principal instructional leadership decisions related to using evaluations for formative or summative purposes. A multiple case study approach was used to identify common principal instructional leadership practices in implementing a teacher evaluation system in schools with ratings of recognized and exemplary by state accountability measures. Different themes were found at different grade levels ($N = 6$). Elementary and middle school administrators adopted multiyear appraisal processes. Middle and high school administrators provided assistance to struggling teachers and used the evaluation system as a summative tool. Elementary principals tended to focus more on using the evaluation system as an on-going process and focused on instructional strategies aligned with the instrument. Middle school administrators developed instructional plans based on data and established professional development goals that supported classroom instruction; again these decisions were based on data. High school administrators selected specific instructional strategies and used the evaluation system as a summative tool (Ovando & Ramirez, 2007). Although the study involved only six administrators from a single district, the findings from this study suggest that instructional leadership practices involved in the implementation of teacher evaluation systems can differ from grade level to grade level.

The purpose as perceived by administrators affects how the administrator implements and conducts the evaluation:

> How evaluators actually conduct teacher evaluations is contingent upon their values, attitudes and beliefs with regard to teaching and education, and how they define their role as evaluator and educational leader. Administrators who view teacher evaluation as an administrative task that fulfills a provincial mandate and discharges legal responsibilities will probably treat the process in an instrumental fashion. However, administrators who view teacher evaluation as an integral part of a plan for school improvement, or as an ongoing process in the maintenance of an effective school, are more likely to incorporate the evaluation process into a continuing professional development plan for teachers to enable them to be more effective in the classroom. Teacher evaluation is then aimed at a specific goal--instructional effectiveness to enhance student learning. In schools where teachers and administrators work collaboratively it would be a natural part of an ongoing review of practice by teachers and administrators to assess the attainment of school improvement goals, to ensure that instructional practices continue to be effective, and to provide input into the refinement of policies and practices in the school. (Bosetti, 1994, p. 54-55)

Expertise

The previous section described principal decisions and leadership behaviors related to teacher ratings. The following section summarizes literature related to principal characteristics possibly connected to principal decisions and leadership behaviors regarding evaluation processes. These characteristics include expertise, leadership frame, and perceptions of others' abilities. Theoretical literature exists espousing the importance of evaluator training and expertise in implementing a teacher evaluation system, but limited attention has been devoted to this relationship in empirical literature. Danielson and McGreal (2000) wrote,

> Many teachers are more expert regarding their work than the administrators who "supervise" them—more knowledgeable about their discipline, current pedagogical approaches, or the developmental characteristics of the students they teach. (p. 6)

This section examines three factors related to principal expertise: training, competence, and experience.

Training. In studies addressed previously in this literature review, Kimball (2002) and Weisberg et al. (2009) concluded sufficient time for training should be provided in implementing a new evaluation instrument. Kimball's conclusion was based on qualitative teacher responses in a multiple district case study. Weisberg et al.'s conclusion was based on background data indicating most principals received only an initial training on teacher evaluation and survey data indicating only 51% of principals felt they received extensive or very extensive training on the evaluation system. Other empirical studies have also found the lack of sufficient training to be a barrier to successful evaluations (Piggot-Irvine, 2003; Sartain et al., 2011; Wise et al, 1985) and found specific needs for evaluator training regarding interactions with teacher (Piggot-Irvine, 2003) and interrater reliability (Sartain et al, 2011).

As part of a larger mixed methods study of 32 school districts, a case study of four districts examined the evaluation process through multiple lenses. The following were three of the most frequently cited problems related to principal implementation of teacher evaluation systems: (a) principal lack of resolve and competence to evaluate accurately; (b) lack of uniformity and consistency; and (c) inadequate training for evaluators (Wise et al., 1985). Each of these barriers is related to the training, experience, and competence of the principal. Wise et al. (1985) found districts varied in the frequency and intensity of training provided to evaluators when implementing a new evaluation system from no training to frequent and intense training, but concluded inadequate training for the evaluator and difficulty in evaluating highly specialized teachers inhibited the effectiveness of evaluation systems. This finding is in agreement with other literature

that states not providing adequate training leads to ineffective evaluation systems and cultures of distrust (Danielson, 2000; Hassan, 2007).

A series of three studies were conducted in New Zealand regarding training and implementation of a teacher appraisal system. The first study included data from an average of 70 appraisers each year from 1996-1999, the second study included data from 45 appraisers, and the final study was an action research study. From these studies, Piggot-Irvine (2003) concluded training should include self-evaluations, nondefensive reactions to problem solving, models for promoting educative interactions, and movement from espoused theory to theory in action. The training should be intensive and ongoing.

In the study of Chicago Public Schools previously addressed in this literature review (Sartain et al., 2011), the influence of training on interrater reliability was examined. Principals and experts were trained in using the Danielson Framework to evaluate teachers in Chicago Public Schools; those participating in the training applied the rating scale consistently. This study of initial implementation, 2008-2009, of the Excellence in Teaching Project included quantitative data from 277 matched observations and qualitative data from interviews of 39 principals and 25 teachers. Principals received 50 hours of training before rating teachers. The principals and external evaluators were consistent in who received higher and lower marks, but some principals (30%) gave lower ratings to all of their teachers and some principals (16%) gave higher marks to all of their teachers (Sartain et al., 2011).

Differences in observers, students, and lessons taught influence ratings, but multiple evaluations by trained observers can reduce the volatility of ratings (Kane &

Staiger, 2012). The Measures of Effective Teaching (MET) Project (Kane & Staiger) found the proportion of variance due to differences in teacher practice as opposed to variations attributable to differences in observers, students, or the lesson taught ranged from .14 to .37 for single observations by trained observers. When four observations by different trained observers were used, the proportion of variance due to differences in teacher practice ranged from .55 to.67 in math classes and from .51 to .54 in English language arts classes. The findings were based on ratings of 7,491 lesson videos created by 1,333 teachers from fourth to eighth grade using two cross-subject instruments and two subject-specific instruments for math classes and one subject-specific instrument for English language arts classes. Evaluators were trained outsiders with no relationship to the teachers.

Competence. Evaluators must be able to recognize, correctly interpret, and link performance to prescribed levels of performance when a teacher meets established criteria through teacher observation, analysis of data, or examination of artifacts (Danielson, 2000). Teachers judge a principal's competence in conducting evaluations based on the principal's ability to recognize effective teacher practices and make appropriate recommendations. Competence influences principals' ability to correctly rate teachers and provide meaningful feedback (Danielson, 2000; Wise et al., 1985), and teacher perceptions of principal competence influences their reactions to the feedback and their relationship to the observer (Ing, 2010; Wise et al.).

Although these studies support the importance of evaluator competence, and Wise et al. (1985) identified competence as one of the four factors distinguishing meaningful evaluations from rituals, a 32-district survey found at least one teacher in every district

believed their principals lacked the competence to recognize good teaching and recommend appropriate strategies (Wise et al.). These findings were based on a RAND study prepared for the National Institute of Education examining teacher evaluation operations in school including the instruments, procedures, and school context (Wise et al.). In a later study on informal observations (N = 15,818), only 42% of teachers responded positively to the statement their principal had the expertise to work with teachers on instruction (Ing, 2010).

Experience. Principal experience measured by their number of years of prior experience as a supervisor and their background before becoming an educational supervisor influenced the evaluation process and the nature of feedback given through the evaluation process (Bryant & Currin, 1995; Chow et al., 2002; Kerrins & Cushing, 2000; Kimball, 2002). Experience affects the focus of evaluations, the credibility of feedback given, and the relationship between the evaluator with the teacher (Bryant & Currin, 1995; Kerrins & Cushing, 2000). Expert (n = 6) and novice (n = 6) evaluators observed a teacher performance and interpreted the data collected. The expert and novice evaluators differed in their focus of attention, recording of data, suspension of judgment, and definitions of teacher-evaluator relationships. Experts focused more on the teacher's actions by recording more data related to teacher practices, delayed judgments until the postconference, and defined their relationships more as partners than as monitors (Bryant & Currin, 1995).

In a similar study, expert (n = 5) and novice (n = 6) evaluators were asked to watch a 45-minute video of a seventh grade math lesson twice and respond to questions after the first viewing and provide comments regarding their impressions during the

second viewing. Expert evaluators were principals with five years or more experience and novice evaluators were principal candidates or principals with less than one year of experience. An analysis of their comments yielded the following results: (a) Expert evaluators provided more interpretive comments (28%), evaluations (16%), and qualifiers (8%) about the teacher's behavior than novice evaluators (24%, 10% and 3%); (b) Novice evaluators provided more descriptive comments (55%) about the teacher's behavior than expert evaluators (39%); (c) Expert evaluators commented on the lesson as a whole, but novice evaluators failed to question the sequencing and coherence of the lesson as a whole; (d) Expert evaluators were more likely to provide qualifiers and supporting statements to their comments; and (e) Expert evaluators provided recommendations (9% of comments) to a question regarding classroom management, but novice evaluators provided no recommendations. This study was a qualitative study with a small nonrepresentative sample of participants conducted in a laboratory setting, but the results suggest that expert evaluators and novice evaluators might implement evaluation processes differently (Kerrins & Cushing, 2000).

Limited research has examined the influence of type of previous experience on principal evaluations of teachers. Wise et al. (1984) identified the issue of generalist evaluators, principals, evaluating specialist teachers particularly at the secondary level. For example, high school principals with no experience in the content area must evaluate physics, calculus, art, and foreign language teachers. This study consisted of preliminary interviews and site visits of 32 districts followed by more in-depth interviews and study of four districts. In a study that included interviews of teachers from three districts (n = 19, 18, and 18), some teachers noted administrators from a nonteaching background such

as school counselor or from another content background, particularly at the high school level, lacked specificity in feedback (Kimball, 2002). Although this study did not address principal ratings of teachers, feedback from the principal to the teacher should support ratings given.

Leadership Frames

In their book, *Reframing Organizations: Artistry, Choice, and Leadership,* Bolman and Deal (2003) presented a theoretical framework capturing four frames used by leaders and managers to make meaning in organizations: the human resource, structural, political, and symbolic frames. They identified how leaders can use each of the four frames in addressing organizational processes. One of the processes addressed is evaluating. In the structural frame, evaluating is a process for distributing rewards and sanctions and controlling performance; in the human resource frame, evaluating is a process for providing feedback to individuals and helping individuals grow; in the political frame, evaluating is a process for exercising power; and in the symbolic frame, evaluating is a process of individuals playing roles in shared rituals.

After addressing each of the four frames, Bolman and Deal examined the multiframe approach to leadership. Bolman and Deal wrote successful leaders must be able to reframe, break their frames and use another frame, and use multiple frames. Fluid frames result in rapid cognition (Gladwell, 2005) and provide leaders a sense of confidence regarding their understanding of situations, but using a single frame can provide leaders a false sense of confidence regarding their understanding and actions (Bolman & Deal).

Empirical research related to Bolman and Deal's (2003) leadership frames and educational leadership have focused on multiframe approaches. In a mixed methods study of principals in Florida and Singapore, principals were asked to type a description of a challenging leadership incident with which they had been involved. Most responses were one to two-pages long with some as short as a paragraph and others several pages long. The responses were analyzed for the four leadership frames and the number of frames used in the case. This study found principals typically used two frames when dealing with challenging situations with 58% of the principals in Florida and 55% of the principals in Singapore describing their cases using two frames. The frames referenced most frequently by the Florida school administrators were human resource frame by 86%, structural by 58%, and political by 50%. The frames referenced most frequently by Singapore principals were human resources 98%, structural 62%, and political 21% (Bolman & Deal, 1992).

A quantitative study (Bolman & Deal, 1992) demonstrated that the leadership frames were able to capture actions of school administrators in the United States and Singapore, and found a stronger correlation between principals who were dominant in the use of multiple frames and both effectiveness as a leader and manager in the United States and Singapore with R^2 values ranging from .54 to .68. For dominance in individual frames, R^2 values ranged from .15 to .58. The strongest correlations were between symbolic frame and effectiveness as a leader in Singapore (R^2 = .58) and structural frame and effectiveness as a manager in the United States (R^2 = .47). A similar study (Bolman & Deal, 1991) found correlations among United States and Singapore administrators' ratings on individual frames and perceptions of effectiveness as a

Table 4

The Four Frames' Relationship to Effectiveness as Leader and Manager

Structural Frame	Corporate Middle Managers	Higher Education Administrators	U.S. Schools Administrators	Singapore Schools Administrators
Effectiveness as Manager				
Structural Frame	.17	.50***	.40**	.26*
Human Resource Frame	.30**	.19***	.05	-.13
Political Frame	.40***	.30***	.14	.15
Symbolic Frame	.12	.04	.32**	.64***
Multiple Frames	.77***	.69***	.71***	.59***
N	90	187	205	274
Effectiveness as a Leader				
Structural Frame	-.28*	.12*	.01	.08
Human Resource Frame	.31**	.18**	.10	-.06
Political Frame	.36**	.28**	.24*	.50**
Symbolic Frame	.73***	.44***	.53*	.30*
Multiple Frames	.87***	.73***	.72***	.62***
N	90	187	205	274

Note: * $p < .05$; ** $p < .01$; *** $p < .001$. From "Leadership and Management Effectiveness: A Multi-Frame, Multi-Sector Analysis" by Bolman and Deal (1991, p. 583).

manager and leader ranging from -.13 to .64. The strongest relationships were found among symbolic frame rating and effectiveness as a manager rating among Singapore administrators ($r = .64$) and symbolic rating and effectiveness as a leader rating among United States administrators ($r = .53$). The relationships among multiple frames and perceptions of effectiveness as a manager and leader for school administrators ranged from .59 to .72. A summary of results of this study including data on education leaders and non-education leaders can be found in Table 4 above.

Numerous doctoral dissertations have been based on Bolman and Deal's leadership frames. Among the 52 dissertations posted by Bolman (2013), there are 21 studies related to higher education, 4 studies related to nursing, 2 studies related to extension workers, and 1 study related to each of the following: clergy, military, and journalism. Studies related to K-12 educational leadership include eight studies related to

principals, three studies related to superintendents, two studies related to school board chairpersons, and two studies related to special education administrators. The eight studies related to principal leadership style include studies related to principal leadership in alternative school settings (Coles, 2005), elementary principal leadership frames (Martinez, 1996), secondary principal leadership frames based on participation in the California School Leadership Academy (Meade, 1992), a comparison of leadership frames for Asian-American principals in California compared to other principals (Suzuki, 1994), leadership orientations of national recognized administrators (Durocher, 1995), and leadership orientations and teacher empowerment (Eckley, 1997).

Among the eight studies related to principal leadership, most studies focused on the use of the four frames but did not specify function (Durocher, 1995; Meade, 1992; Suzuki, 1994). Meade compared the leadership frames of principals who had participated in the California School Leadership Academy (CSLA) (n = 156) to principals who had not (n = 190). For principals who had completed the CSLA, 56% used a single orientation, 24% used two orientations, and 20% used three or more orientations. For principals who had not completed the CSLA, 60% used a single orientation, 19% used two orientations, and 21% used three or more orientations. The differences between the two groups were not statistically significant. In another dissertation, Suzuki found gender to be a significant factor with female principals using the human resource frame more than males (F (1) = 5.31, $p < .05$, N = 124, effect size not given). Suzuki also found nationality to be significant with principals born outside the United States using the structural frame more than principals born in the United States (N = 124). Suzuki

compared the use of multiple frames by Asian-American principals to other principals in previous studies by Meade (1992).

Durocher (1995) examined the use of the four frames among nationally recognized principals ($N = 70$). Fewer principals operated in a single frame than in Suzuki's (1994) study. In this study, 20% of principals operated in one frame *often* or *always*, 21.4% operated in two frames *often* or *always*, 21.4% operated in three frames *often* or *always*, 24.3% operated in four frames *often* or *always*. The human resource frame was the frame most frequently identified as dominant based on principal responses. Among principals using one frame, 71.4% operated in the human resource frame *often* or *always*; among principals using two frames, 80.0% operated in the human resource frame *often* or *always*; and among principals using three frames, 100% operated in the human resource frame *often* or *always*. There were significant relationships between self-ratings on effectiveness as a manager and structural frame orientation ($r = .42$, $p < .05$) and self-ratings on effectiveness as a manager and political frame orientation ($r = .29$, $p < .05$). Female administrators demonstrated a higher use of the structural and human resource frame than male administrators using a two-tailed *t*-test ($t = -2.42$ and -2.67, $p < .05$). The data from this study are compared to findings from eight previous studies, but only descriptive statistics were reported in these comparisons (Durocher).

Martinez (1996) focused on comparing self-perceptions to administrative designee perceptions and did not address principal behavior in rating teachers. This study compared the leadership orientation ratings of principals who had participated in California School Leadership Academy to principals who had not participated. Eckley (1997) focused on teacher empowerment. Teachers ($N = 525$) who perceived principals

to be operating in the human resource frame felt more empowered than teachers who did not perceive principals to be operating in the human resource frame at a statistically significant level ($t = 6.72$, $p < .001$). They also felt more empowered when they perceived the principal to be operating in the structural frame ($t = 2.17$, $p = .04$). This study used hierarchical linear modeling which showed only the human resource frame to be significant at the teacher level with a coefficient of 1.38 ($p < .01$).

The only study to explicitly address teacher evaluations examined three first year principals at an alternative school. This study examined the challenges faced by first year principals associated with each frame (Coles, 2005). These three principals focused on using evaluations to hold staff accountable, terminate the worst staff, and build instructional capacity among the rest of the staff. The only references to evaluation processes were among the human resource and structural frame suggesting principal ratings on human resource and structural frames might be more important than ratings on symbolic and political frames in predicting principal mean ratings of teachers.

Implicit Person Theories

Implicit person theory research has examined how individuals' views of their own and others' intelligence, personal characteristics, abilities, and morality influence their judgments of others, selection of goals, and responses to failure (Chiu et al., 1997; Dweck, et al., 1995). Earlier implicit theory research reviewed by Chiu et al (1997) focused on how individuals' self-theories related to intelligence influenced their responses to complex problems, failure, and meeting or not meeting goals. Entity theorists focused on how much intellectual ability they believed they possessed and attributed failure to their intellectual ability. Incremental theorists focused on developing

their ability and improving performance through exerting effort and applying problem-solving strategies (Chiu et al.). This lens was expanded to examine the influence of implicit person theories of others on judgments, stereotypes, and other reactions to others' behaviors and traits. This section of the literature will focus on how theories of others influence judgments. The first section will present research related to the instrument with subsequent sections examining literature related to judgments and evaluations of others.

Implicit person theory instruments. Dweck, Chiu, and Hong (1995) published the results of six studies (n = 69, 184, 139, 121, 93, and 32) measuring the validity and reliability of their measures of implicit theories. These studies examined three instruments measuring implicit theories about kind of person, morality, and intelligence. The first study included a retest (N = 62). Internal reliability of these studies ranged from .90 to .98 with means ranging from 3.11 to 3.96 and standard deviations ranging from 0.95 to 1.49. Test-retest reliability over a two-week period ranged from .80 to .82 in Study 1. Levy and Dweck (1996) also compared the use of only the expanded incremental items to the use of both incremental and entity items. Participants (N = 101) completed the survey using both types of items and using only the expanded incremental items with a delay of a week or less between the first and second surveys. There was a correlation of classifying individuals as incremental or entity theorists of .83.

Further tests examined the relationship between implicit person theory and the following factors: age, sex, self-presentation concerns, cognitive abilities, confidence and optimism, and political attitudes (Dweck et al., 1995). Implicit person theory was a distinct factor. None of the correlations were significant; r values ranged from -.16 to .15

with the highest absolute values for *r* found using the Referent Scale (Kerlinger, 1984), *r* = .16; Right-Wing Authoritarianism (Altemeyer, 1981), *r* = -.16; Social Desirability Scale (Paulhus, 1984), *r* = .15; and Social Attitude Scale (Kerlinger, 1984), *r* = -.15. This study used implicit person theory rating as a continuous variable instead of classifying individuals as holding an incremental theory, entity theory, or undetermined. Factor analysis was also used to investigate if responses could be the results of acquiescence. Participants in the first five studies (*n* = 69, 184, 139, 121, and 93) responded to items related to views of the world, intelligence, and morality. Factor analysis confirmed that the three views were distinct factors supporting the premise that ratings were not the result of acquiescence.

Although individuals can theoretically hold both incremental and entity beliefs, one belief is generally more dominant than the other (Dweck et al., 1995). Dweck et al. wrote in a peer-reviewed journal article discussing implicit person theory and previous research that after excluding 15% of individuals who typically do not fall clearly into either category, participants were evenly distributed across entity and incremental theories unrelated to level of education, cognitive ability, self-esteem, or optimism. Dweck et al. did not provide further data in this particular article. To verify the distribution of incremental and entity theories, distributions of individuals holding incremental and entity theories for articles reviewed in this literature review were collected, the number of individuals holding each theory were used to calculate the overall percent of individuals holding each theory. The data are displayed in Table 5. An analysis of the participants in these studies found 44% of participants held an incremental theory, 40% held an entity theory, and 15% were undetermined.

Table 5.

Distribution of Incremental and Entity Theorists

Author	Year	Participants	*N*	Incremental	Entity	Undetermined
Levy, Stroessner, & Dweck	1988	Study 1—Students	78	53%	38%	10%
		Study 2—Students	114	30%	50%	20%
		Study 3—Students	121	49%	51%	0%
Hong et al.	1997	Students	124	37%	44%	19%
Chiu, Hong, & Dweck	1997	Study 1—Students	40	38%	38%	25%
		Study 2—Students	50	22%	42%	36%
		Study 3—Students	37	54%	27%	19%
Chiu, Hong, & Dweck	1999	General	37	54%	27%	10%
Levy & Dweck	1999	Study 1—Elementary Students	78	36%	38%	26%
		Study 2—Elementary Students	44	39%	43%	18%
Heslin	2003	Study 1—Managers	84	29%	39%	32%
		Study 2—Managers	43	30%	53%	7%
		Study 3—Managers	115	46%	54%	0%
Gutshall	2013	Teachers	238	62%	26%	12%
Overall			**1,203**	**44%**	**40%**	**15%**

Gutshall (2013) found 61.8% of teachers held an incremental or growth mindset, 26.1% held an entity or fixed mindset, and 12.2% did not clearly hold either mindset. Gender, years' experience, special education versus regular education, and school type did not significantly influence teacher theory ($p > .05$): gender, $\chi^2(1, n = 134) = .274$; years of experience, $\chi^2(1, n = 218) = .530$; special education/regular education, $\chi^2(1, n =$

206) = .260; and school type, $\chi^2(1, n = 238) = .494$. This study also referenced two unpublished dissertations examining teacher implicit person theory with the first dissertation ($N = 142$) finding 73.6% of teachers held an incremental theory, 24.6% of teachers held an entity theory, and 9.2% of teachers were undetermined (Strosher, 2003, as cited in Gutshall, 2013) . The second dissertation ($N = 63$) found 50% of teachers held an incremental theory, 18% of teachers held an entity theory, and 32% of teachers were undetermined (Subert, 2006, as cited in Gutshall, 2013).

Behaviors and traits of others. According to previous literature reviews and theoretical literature, individuals who hold an entity theory stereotype others more frequently, agree more strongly with stereotypes, more closely associate groups with stereotypes, attribute traits to members of groups based on group identity, use stereotypes to reflect group differences, group individuals based on stereotypes, and use more extreme qualifiers to describe traits of groups. This tendency to stereotype and use traits as the primary measure in evaluating others is attributed to entity theorists' beliefs that traits are inherent. Entity theorists more quickly make judgments of groups because they believe a person demonstrating trait-related behavior in one situation meant the person was more likely to demonstrate trait-related behavior in subsequent situations (Levy et al., 1998; Levy et al., 2001; Chiu et al., 1997). This section addresses empirical literature related to the influence of implicit person theories on judgments of others.

College students participated in a series of studies that analyzed the influence of their implicit person theory on their evaluations of others based on trait-related behaviors. A series of five studies ($n = 40, 50, 37, 310$, and 46) examined the influence of implicit person theory and trait-related behaviors on student judgments of others (Chiu, et al.,

1997). In Study 1 individuals were given a situation and asked to indicate on a probability scale of 0 to 1 the likelihood that if they met one of the people described in the situation the person would act differently. Individual ratings on entity theory items were treated as a continuous variable. Entity theory and ratings of future behavior correlated significantly in the social domain ($r = .61$, $p < .001$) and the ability domain ($r = .45$, $p = .01$). Entity theorists made stronger predictions ($M = .57$) about future behavior in the social domain than incremental theorists ($M = 42$), $t\ (28) = 4.03$, $p < .001$. They also made stronger predictions ($M = .55$) about future behavior in the social domain than incremental theorists ($M = 44$), $t\ (28) = 2.70$, $p = .01$. In Study 2, participants were given 10-items describing a trait of a person and asked to predict the person's behavior in a future situation using a probability scale. The items included positive and negative traits. There was a positive correlation between agreement with entity theory items and participant predicted probability that the individual would display behavior associated with that trait in the future ($r = .38$, $p < .05$).

In Study 3, participants were given descriptions of trait-related behaviors and asked to rate on a scale of 1 to 5 the extent to which the single behavior revealed a person's character. Entity theorists were more likely than incremental theorists to believe a single behavior could reveal a person's character ($M = 3.62$, $M = 2.96$, $t\ (28) = 3.09$, $p < .005$). Study 4 compared the extent to which students in Hong Kong and the United States believed behaviors indicated a person's character. For both groups of students, entity theory influenced judgments (United States, $r = .24$; Hong Kong, $r = .22$; $p < .01$). Study 5 examined the influence of reading materials supporting implicit person theory worldview versus reading materials supporting entity theory worldview ($n = 46$).

Students assigned to read entity materials allowed traits to more strongly influence judgments ($M = 3.72$ versus $M = 3.27$) at a significant level, $F(1, 44) = 4.27$, $p < .05$ (Chiu et al., 1997).

A second group of studies examined the influence of student implicit person theories on their judgments of schools; their willingness to interact with others; and their attributing behaviors to traits, psychological processes, or external factors. Two sets of sixth grade students were asked to rate behaviors of students on a 9-point scale with -4 being *very, very* mean and 4 being *very, very nice.* For the first part of the study, researchers asked sixth grade students ($N = 78$) to complete a survey related to their beliefs about the malleability of others' personality. Students were classified as entity theorists ($n = 30$), incremental theorists ($n = 28$), and undetermined ($n = 20$). The 20 students who were not clearly entity or incremental theorists were not included in the remaining activities in the study because the goal of the study was to compare judgments of students holding an entity theory to judgments of students holding an incremental theory. The remaining students ($n = 58$) were asked to read negative and neutral behaviors of nine students in a simulated school and rate the behaviors on a 9-point scale. The researchers hypothesized that entity theorists would assign more extreme ratings. Students were then asked to rate the school based on the behaviors of the nine students. Students holding an entity theory rated the school lower ($M = -2.03$) than students holding an incremental theory ($M = -1.21$) at a significant level ($F(1, 57) = 5.95$, $p < .05$). Students holding an entity theory made a greater number of extreme responses ($M = 2.63$) than students holding an incremental theory ($M = 1.18$) at a significant level ($F(1, 57) = 4.50$, $p < .05$). Students were also asked to rate their willingness to interact with

the students from this school associated with students with negative and neutral behaviors. A 5-point scale with 1 being *not at all* to 5 being *very, very much* was used. Entity theorists wanted to interact less ($M = 1.51$) with these students than incremental theorists ($M = 1.96$) at a significant level ($F(1, 57) = 5.39$, $p < .05$). Effect size was not reported (Levy & Dweck, 1999).

Students were asked to respond to an open-ended question describing the behaviors of the nine students, "Why do you think the students at the school acted the way they did?" (Levy & Dweck, 1999, p. 1168). These responses were then coded as attributing behavior to traits such as "they are mean" or "they are dishonest," (Levy & Dweck, p. 1170), psychological processes such as "to get attention" or "they wanted to be mean," (Levy & Dweck, p. 1170), and external factors such as "others were acting that way," or "the teacher doesn't teach them right or their parents don't teach them right" (Levy & Dweck, p. 1170). Among students holding an entity theory, 64% attributed behavior to traits, 25% attributed behavior to external factors, and 11% attributed behavior to psychological processes. Among students holding an incremental theory, 68% attributed behavior to external factors, 16% attributed behavior to psychological processes, and 16% attributed behavior to traits. Students holding an entity theory and students holding an incremental theory differed significantly in their explanations (χ^2 (2, $N = 53$) = 19.22, $p < .05$) when comparing the number of children expressing either of the three explanations with the explanation acting as the dependent variable and their theory acting as the independent variable (Levy & Dweck).

This study also asked a group of sixth grade students ($n = 44$) to read descriptions of nine students from a fictitious school with negative and neutral behaviors and nine

description of students from another fictitious school with positive and neutral behaviors. Entity theorists were more likely to rate the school of students with positive behaviors higher (M = 3.19) than incremental theorists (M = 2.76) at a significant level ($F(1, 35)$ = 4.13, p = .05). They were also more likely to rate the school of students with negative behaviors lower (M = -2.43) than incremental theorists (M = -1.46) at a significant level, F (1, 34) = 7.90, p < .01 (Levy & Dweck).

Situational differences. University students (N = 238) were asked to watch a video of a woman demonstrating anxious behavior. Some students were described a *stressful* situational context and other students were provided a *mundane* situation context. Some students were also given a *cognitive load* and asked to press the spacebar each time the number 16 appeared on the screen. The number was spliced into the video at random times. Students were asked to rate the anxiousness of the woman by answering the following items on a scale of 1 to 9 with 1 representing *not at all* and 9 representing *extremely*: (a) "How anxious a person is the woman in the video?", (b) "How anxiously did the woman appear to be acting in the video?", and (c) How anxious do you think the woman would be if she were asked to give an impromptu presentation in a seminar?" (Molden, et al., 2006). Of the responses, 220 were analyzed. Other responses did not correctly report the topic, did not complete the task, or did not follow directions. Responses were centered and hierarchical regression was conducted with the following levels: (a) implicit person theory, (b) stressful or mundane situation, and (c) all interactions. Implicit person theory did not influence their rating of behaviors as anxious (M = 6.0, SD not reported). The topic (*stressful* or *mundane)* had a significant effect ($\beta = -.24$, t (215) = 3.70, p < .001) on students' ratings of the woman's behavior. The

topic × cognitive load interaction (β = .13, t (212) = 1.98, p < .05) and the predicted theory × topic × load interaction (β = −.44, t (211) = 1.95, p = .05) were also significant. Simple slope analyses indicated that entity theorists decreased their ratings of the woman's behavior when they were not experiencing cognitive load (β = −.46, t (211) = 3.7, p < .001) as compared to entity theorists who were experiencing cognitive load (β = .06, t (211) = .46, p = .65). For incremental theorists information about the situation influenced their ratings both when they were experiencing cognitive load (β = −.26, t (211) = 1.97, p = .05) and when they were not experiencing cognitive load, β = −.26, t (211) = 1.98, p < .05 (Molden et al., 2006).

In Study 2, participants with valid responses (n = 214) were given information about the woman's calm or anxious disposition. Members of the group given information categorizing the woman as a calm person categorized her actions as less anxious (M = 5.5) than members of the group given information about the woman's anxious disposition (M = 6.6) at a significant level, (β = .32, t (210) = 4.9, p < .001). Implicit person theory rating was treated as a continuous variable. There was a significant effect of disposition and implicit person theory. For individuals with higher ratings (incremental theorists), information about the woman's disposition affected their ratings of her anxiety when they were not experiencing cognitive load (β = -.23, t (205) = 2.29, p < .05) but did not influence their ratings of her anxiety when they were experiencing cognitive load (β = .11, t (205) =1.0, p = .33). For individuals with lower ratings (entity theorists), information about the woman's disposition influenced ratings of anxiety when experiencing or not experiencing cognitive load, β = -.19, t (205) = 1.96, p = .05, and, β =.10, t (205) = .92, p = .36 (Molden et al., 2006).

Study 3 addressed the manipulation of a person's implicit person theory. Participants (N = 174) received either an incremental or entity theory induction. Only 157 of responses were usable in analysis due to protocols in place. For example, 2% of participants guessed that the articles used in the induction were not real, so they were excluded from analysis. Participants in the incremental induction group were influenced by the situation in rating the woman's anxiety level, F (1, 68) = 10.9, p < .01. Participants in the entity induction group were not significantly influenced by the situation in rating the woman, F (1, 68) = .11, p = .83 (Molden et al., 2006).

Evaluating performance. Heslin (2003) presented the results of three studies in the dissertation, *The Effects of Prior Judgment and Implicit Person Theories on Performance Appraisals*. These three studies included 82 managers, 42 managers, and 62 individuals holding entity theories. In Study 1, hierarchical regression tests were used to analyze the influence of implicit person theory on the variability in two ratings of individual performance over time. Study 1 used a first video of poor performance followed by a second video of good performance. For Study 2, participants rated a first video of good performance followed by a second video of poor performance. In the first study (measuring poor to good performance), there was a significant relationship between Rating 1 and Rating 2 (R^2 = .10, p < .01), but implicit person theory also significantly influenced the mean ratings on the second evaluation after controlling for initial ratings (adj. R^2 = .32, B = .35, p < .001). In the second study using an initial video of good performance followed by a video of poor performance, there was not a statistically significant difference between Rating 1 and Rating 2 (R^2 = .08, p > .05). There was a

statistically significant difference in Rating 2 after controlling for Rating 1, (adj. $R^2 = .16$, $B = -.19$, $p < .01$).

In the third study, only individuals holding an entity theory were chosen to participate in the study ($n = 62$). One group of entity theorists participated in an incremental workshop condition and one group participated in a placebo condition. On the first rating of poor performance, there was not a significant difference in the means of the treatment and control groups ($M = 1.72$ and $M = 1.75$), but there was a significant difference in the mean ratings for the second groups with the treatment group rating performance higher ($M = 4.32$) than the placebo group ($M = 3.47$) at a significant level ($F(58) = 21.54$, $p < .001$, partial $\eta^2 = .12$). Not only did the treatment influence mean ratings, the treatment also affected willingness to coach, number of suggestions given, and quality of suggestions. Using a Likert scale of 1 to 5, the treatment group ($M = 4.04$, $SD = 0.78$) rated their willingness to coach higher than the control group ($M = 3.33$, $SD = 1.15$) at a significant level, $F(1, 58) = 7.59$, $p < .01$, partial $\eta^2 = .12$. The treatment group ($M = 3.38$, $SD = 1.74$) gave more suggestions than the control group ($M = 2.19$, $SD = 1.67$) at a significant level ($F(1, 58) = 7.55$, $p < .01$). The quality of coaching was rated by two independent raters on a scale of 1 to 5. The quality of coaching of the treatment group ($M = 3.10$, $SD = 1.39$) was rated higher than the control group ($M = 1.88$, $SD = 1.33$) at a significant level ($F(1, 58) = 12.80$, $p < .01$, partial $\eta^2 = .19$) (Heslin, 2003).

Another series of three studies following similar procedures to Heslin's (2003) procedures examined the influence of implicit person theory on manager evaluations of and feedback to employees. In Study 1, employees were asked to rate the coaching

behaviors of their supervisors. After controlling for manager years' experience and supervisor age, supervisor implicit person theory significantly explained 14% of the variability in employees' ratings of supervisors' coaching behaviors ($N = 45$). The hierarchical regression model was statistically significant (F (3, 42) = 3.06, $\beta = .38$, $p <$ .01). Study 2 used a larger sample ($N = 92$). After controlling for manager years' experience and supervisor age, supervisor implicit person theory significantly explained 21% of the variability in employees' ratings of supervisors' coaching behaviors. The hierarchical regression model was statistically significant (F (3, 88) = 8.93, $\beta = .48$, $p <$ .01. The third study used incremental theory induction and a control group to manipulate the beliefs of entity theorists ($n = 62$). The incremental theory induction had a moderate effect on supervisors' willingness to coach ($M = 4.04$, $SD = .78$) when compared to the control group ($M = 3.33$, $SD = 1.15$) at a significant level (F (1, 58) = 7.67, $p < .01$, partial $\eta^2 = .12$). The treatment also had a moderate effect on the number of suggestions offered ($M = 3.38$, $SD = 1.74$) when compared to the control group ($M = 2.19$, $SD =$ 1.67) at a significant level (F (1, 58) = 7.23, $p < .01$, partial $\eta^2 = .11$). The treatment had a moderate effect on the quality of coaching as measured by independent raters with incremental theory induction group *(*$M = 3.10$, $SD = 1.39$) receiving higher scores than the control group ($M = 1.88$, $SD = 1.33$) at a significant level (F (1, 58) = 11.93, $p < .01$, partial $\eta^2 = .17$ (Heslin, VandeWalle, & Latham, 2006).

Summary. Implicit person theory was used as part of the theoretical framework for this study because of its potential to explain implicit beliefs of principals possibly influencing their ratings of teachers. Implicit person theories influenced students' predictions of others' future behavior based on previous trait-related behavior (Chiu et

al., 1997). Entity theorists are more likely to make more extreme judgments based on single behaviors, they are less likely to reflect changes in performance over time, and they are less likely to consider external factors in assigning ratings. Entity theorists made more extreme judgments of future behavior than incremental theorists (Chiu et al.) and assigned more extreme judgments, both positive and negative, to schools based on nine examples of student behaviors (Levy & Dweck, 1999). Entity theorists believed more strongly that a single behavior revealed a person's true character (Chiu et al.). Students holding entity theories were more likely to interpret personal traits as the cause of behavior, but students holding incremental theories were more likely interpret external factors as the cause of behavior (Levy & Dweck; Molden et al., 2006). Cognitive load of the evaluator influenced their integration of situational factors in assigning ratings (Molden et al.). Evaluators with an incremental theory were more likely to assign higher second rating when performance improved than evaluators with an entity theory (Heslin, 2003). Implicit person theories of evaluators influenced the quality of coaching, the number of suggestions, and the quality of feedback (Heslin, 2003; Heslin, VandeWalle, & Latham, 2006). Incremental theory induction can influence judgments (Chiu et al., 1997; Heslin, 2003; Levy & Dweck, 1999) suggesting that providing incremental theory induction during training on the evaluation instrument could create evaluators who adopt a more growth-oriented approach to evaluation processes.

Inferences for Forthcoming Study

Limited research has studied factors influencing principal ratings of teachers, but this literature review suggests teacher, student, evaluation process, and principal characteristics influence principal ratings of teachers. No empirical studies were found

addressing the influence of principal dominant leadership frames on principal ratings of teachers, but the literature suggests that principals operating in the human resource frame might be more lenient in order to focus more on a formative approach, build positive relationships with teachers, and protect staff morale (Kluger & DeNisi, 1996; Painter, 2000; Sagona, 2012; Sartain et al., 2011). Implicit person theories have the potential to explain actions and reactions to evaluation processes. Implicit person theories influence principal responses to complex problems encountered during implementation (Chiu et al., 1997), interactions with teachers, judgment of teachers, and investments in professional development (Dweck, 1995; Heslin et al., 2005; Hong et al., 1999; Levy et al., 1998; Levy et al., 2001).

Gaps in the Research

There are numerous studies on evaluation systems; many exploring reactions to and perceptions of evaluation systems. Some studies have examined the relationships between principal ratings of teachers and student achievement data but have not explicitly explored the factors moderating or influencing these relationships. There are numerous studies regarding implicit person theories in the psychology investigating academic pursuits and social interactions of K-12 and postsecondary students. These studies explored responses to complex problems, judgment of others, and judgment of self. There are limited studies on implicit theories and evaluation processes. No studies were found directly examining the relationship between implicit theories and implementation of a professional teacher evaluation system. No studies were found examining the relationship between principal ratings of teachers and principal dominant leadership frame.

Chapter 3: Research Methodology

With high school graduation rates below 70%, the North Carolina State Board of Education commissioned 16 representatives from the North Carolina Department of Public Instruction, 11 school districts, institutions of higher education, and the North Carolina Association of Educators to revise the North Carolina teacher evaluation standards. The resulting North Carolina Professional Teaching Standards laid the foundation for a new teacher evaluation process (North Carolina Public Schools, 2006; 2008). Despite the new evaluation process's goal to focus on student learning, no correlation was found between student growth and teacher evaluation ratings after the first year of statewide implementation (North Carolina Public Schools, 2011). The North Carolina Department of Public Instruction staff concluded the data indicated there were needs for additional professional development on the instrument and the explicit inclusion of student growth through a sixth standard based on Education Value-Added Assessment System (EVAAS) measurements (North Carolina Public Schools, 2011). Incorporating student achievement in the teacher evaluation process was also a component of North Carolina's Race to the Top application (North Carolina Public Schools, 2013). Empirical research suggests factors other than teacher effectiveness influence principal ratings of teachers and should be considered including characteristics of teachers, schools, and principals (Holtzapple, 2005; Jacob & Lefgren, 2008; Kimball et al., 2004; Milanowski, 2004; Sartain et al., 2011).

The overarching framework of this study is social cognitive theory (Bandura, 1986) as described in Chapters 1 and 2. The literature review, found in Chapter 2, focused on three categories of external factors: teacher, student, and evaluation process

characteristics. Instead of using these three categories, this study focused on school characteristics as external factors for the following reasons: (a) The unit of the study was the principal, and aggregate teacher and student characteristics were treated as school characteristics; and (b) Differences in the evaluation process were difficult to quantify because the evaluation process used the same or similar instruments, policies, and training statewide. School characteristics included the grade span of the building as reported by the principal and the school composite value-added data as measured by EVAAS data and reported by North Carolina Public Schools. Internal characteristics were limited to principal characteristics and included the following: (a) Use of leadership frames as measured by Bolman and Deal's (1990) Leadership Orientations Instrument—Self, Section I; beliefs about the malleability of others' personal characteristics and abilities as measured by the Kind of Person Instrument—Others (Dweck et al., 1995); and principal total years' experience as principal. The third aspect of social cognitive theory, motivation and action, was limited to principal mean ratings of teachers on each of Standards 1 through 5. The predictor variables were selected based on a review of the literature related to teacher evaluations for their potential to explain variability in principal ratings of teachers and the fit of the factors to quantitative research with the principal as the unit of study. An illustration of the suggested interactions can be found in Figure 4 in Chapter 1.

Research Questions

The research questions were addressed through a survey of 399 North Carolina principals with stratified, proportional sampling techniques used to select 226 elementary, 88 middle, and 85 high school principals of schools with value-added data. The

following overarching question guided this study: Do principal and school characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument? These analytic questions were used to answer the overarching question:

1. Do principal characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    a. Do principal implicit person theories as measured by the Kind of Person Instrument (Dweck et al., 1995) influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    b. Does principal use of dominant leadership frames as measured by Bolman and Deal's (1990) Leadership Orientations Instrument influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    c. Does principal total years' experience as a principal influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
2. Do school characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    a. Does the school value-added composite measured by EVAAS data influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    b. Does school grade span as reported by the principal influence principal ratings of teachers on the North Carolina teacher evaluation instrument?

Research Design

This research was grounded in the premise principal, teacher, and school factors influence principal subjective ratings of teachers. This study used simple and multiple

regression techniques to examine the extent to which the following factors contribute to principal ratings of teachers: the overall teacher effectiveness as measured by EVAAS value-added data, the grade span of the school, principal years' experience as a principal, principal dominant leadership frames as measured by Bolman and Deal's (1990) Leadership Orientations Instrument—Self, Section I, and beliefs about the malleability of others' personal characteristics and abilities as measured by the Kind of Person Instrument (Dweck et al., 1995). The Leadership Orientations Instrument—Self, Section I, (Bolman & Deal) was chosen because it provides a measure of principal leadership style. The Kind of Person Instrument (Dweck et al., 1995) was chosen as a measure of principal beliefs about the malleability of others' personal characteristics and abilities.

Predictor and Criterion Variables

The predictor variables of this study included external factors—school grade span and overall teacher effectiveness as measured by EVAAS value-added data—and principal factors—principal leadership style as measured by Bolman and Deal's (1990) Leadership Orientations Instrument—Self, Section I, principal beliefs about the malleability of others' personal characteristics and abilities as measured by Kind of Person Instrument (Dweck et al., 1995), and years' experience as a principal. The criterion variables were the principal mean ratings of teachers on each of Standards 1 through 5; Standard 6 is based on EVAAS value-added data not principal ratings. North Carolina's teacher evaluation process does not include an overall teacher rating, so separate regression tests were conducted for each standard and for each question. Principal ratings were assigned a numerical value with *not demonstrated*, *developing*,

proficient, *accomplished*, and *distinguished* assigned values of 0, 1, 2, 3, and 4 respectively.

Delimitations

This study was limited to principals in North Carolina who assigned teachers in their school their most recent summative evaluation ratings. The findings may not be generalizable to first year principals who were nonrenewed, evaluation processes in other states, and nonprincipal evaluators including peer observers, mentors, curriculum coaches, assistant principals, and central office personnel. North Carolina public schools were configured into 57 different grade spans. Because 2012-2013 value-added data were only available for fourth grade through high school, only elementary schools with fourth and/or fifth grade students were included in the study. Because school grade span was a predictor variable, schools not distinctly elementary, middle, or high schools were not included. Elementary schools spanning beyond sixth grade were not included. Middle schools only included schools serving students in sixth, seventh, and/or eighth grades. High schools only included schools serving students in ninth through twelfth grades. Schools designated as early colleges and serving students in Grades 9-13 were not included.

The focus of this study was the influence of principal characteristics on principal ratings of teachers. Factors not addressed but with potential to influence principal ratings include characteristics of the evaluation process; teacher characteristics; barriers, problems, and resources; teacher-principal relationships; and principal perceptions of the standards. Characteristics of the evaluation process and barriers, problems, and resources were not addressed because all districts in North Carolina implemented the same process

with similar resources provided through Race to the Top funds and regional training. Teacher-principal relationships and principal perceptions of the standards were not addressed as the design did not support inclusion of these variables.

Limitations

Limitations of this study included limitations associated with surveying only principals, choosing the principal instead of the teacher as the unit of analysis, and using regression techniques to answer research questions. There were two limitations associated with surveying only principals. Principal dominant leadership orientation frames were based on principal perceptions of their own leadership not on teacher perceptions; only using self-evaluations decreased the validity of the Leadership Orientations Instrument (Bolman, 2000). Principals may choose to report their implicit person theory based on their perceptions of the correct answer not their true theory. The second area of limitation, using the principal as the unit of analysis, was chosen due to the confidentiality of teacher evaluations and the desire to focus on the principal in the evaluation process. Because the principal was chosen as the unit of study, this research was not able to address teacher-principal relationships or examine differences in ratings among teachers and across standards. Being able to examine within school ratings would allow the researcher to identify if principal ratings of a teacher as compared to other teachers in the building were an accurate reflection of teacher value-added data. Because of the research design, the researcher was only able to draw conclusions regarding school mean ratings. Finally, using regression techniques allowed the researcher to identify relationships but not the underlying causes. The design of this study only allowed the researcher to determine if predictive relationships, not causal relationships, existed.

Because the independent variables cannot be manipulated for this study, causal relationships could not be identified.

Participants

The sampling frame of a study is, "those people who have a chance of being included" (Fowler, 2009, p. 20). The sampling frame for this study was based on an Educational Directory and Demographical Information Exchange (EDDIE) custom report. North Carolina had 2,489 public schools with 56 grade span configurations (North Carolina Public Schools, 2013). The custom report from EDDIE was filtered to include only public school principals for schools that included grades 4-12 because value-added data were not available for K-3. This report included school number, school name, grade span, principal names, principal email addresses, and other descriptive indicators related to the school characteristics. A comparison of the 2013-2014 report to the 2012-2013 report was used to identify principals returning to the same school. Because data released in the fall reflects performance and evaluation ratings from the previous year, only principals remaining in the same school were included in the sample.

Because this study addressed a research question related to school grade span, only principals of schools that were distinctly elementary, middle, or high schools were included in the sampling frame. For example, K-8 and 7-12 schools were not included in this study. These configurations were not included because K-8 schools included ratings for elementary and middle school teachers, and 7-12 schools included ratings for middle and high school teachers. For the purposes of this study, elementary schools were limited to schools containing fourth and fifth grade students but no students above sixth grade. Middle schools included schools serving sixth, seventh, and/or eighth grade students.

High schools included only traditional high schools serving students in ninth through twelfth grades. The grade span configurations of North Carolina public schools are included in Table 6 with school grade spans included and school grade spans not included in the sampling frame indicated.

The sample size for this study was obtained using G*Power 3.7.1 for "multiple regression, fixed model R^2 increased" (Soper, 2013). Significance criterion was set at α = .05, and a moderate effect size of f^2 =.15 was selected. The number of test predictors was set at 5 and "a priori" type of power analysis was chosen. The results of the calculation were a noncentrality parameter of $\lambda = 20.7$; a critical F value of 2.28, and $df = 5, 132$. The G*Power calculator recommended a sample size of 138 with an actual power of .9508. A second calculation for sample size was run using *A-Priori Sample Size Calculator for Multiple Regression* (Faul, Erdfelder, Buchner, & Lang, 2009). The same parameters of α = .05, F = .95, f^2 =.15, and K = 5 were used. This calculator also recommended a sample size of 138.

This study used stratified, proportional sampling with three strata—elementary, middle, and high school principals—to address the research question, "Does school grade span influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" Stratified sampling techniques were used because they can produce lower sampling errors than simple random sampling techniques (Fowler, 2009). A meta-analysis of response rates for 56 online surveys reported in 39 studies yielded a mean response rate of 34.6%, SD = 15.7% (Cook, Heath, & Thomson, 2000). Using 34.6% as

Table 6

North Carolina Public Schools Grade Span Configurations

School Grade Span	Number of Schools	Included in Elementary	Included in Middle School	Included in High School	Not Included
K-3	6				X
K-4	18				X
K-5	591	X			
K-6	28				X
K-7	3				X
K-8	75				X
K-10	4				X
K-12	32				X
PK-3	13				X
PK-4	38				X
PK-5	487	X			
PK-6	39				X
PK-7	1				X
PK-8	40				X
PK-12	6				X
PK-13	2				X
1-5	4	X			
1-12	1				X
2-3	2				X
2-4	1				X
2-5	4	X			
2-6	1				X
2-12	1				X
3-5	22	X			
3-8	1				X
3-10	1				X
3-12	3				X
4-5	11	X			
4-6	6				X
4-8	3				X
4-12	1				X
5-6	8				X
5-12	2				X
5-13	1				X
6	1				X
6-7	1				X
6-8	432		X		
6-9	1				X

6-10	2				X
6-12	55				X
6-13	1				X
7-8	25				X
7-9	2				X
7-12	8				X
7-13	2				X
8	1				X
8-12	3				X
8-13	1				X
9	3				X
9-10	1				X
9-11	7				X
9-12	418			X	
9-13	47				X
10-12	2				X
11-12	2				X
SP-ED	18				X
	2.489	1,112	432	418	527

Note: A description of grade span configurations in North Carolina public schools indicating those configurations included in the sampling frame of this study. Adapted from *Performance of All Schools* by North Carolina Public Schools (2012).

the estimated response rate for this study and a desired response from at least 138 principals gave an initial sample size of 399 (138/34.6%).

Because the sample was a stratified, proportional sample with three strata, the proportion of each school level was found by dividing the possible schools at each level by the total number of possible schools: elementary, 1,112/1,962 = 56.7%; middle school, 432/1,962 = 22.0%; and high school, 418/1,962 = 21.3%. These percentages were multiplied by the desired sample size of 399 to find the number of schools to sample at each level: 226 elementary, 88 middle schools, and 85 high schools. The *Performance of All Schools* file was downloaded from North Carolina Public Schools accountability page. This file included all schools in North Carolina, their school codes, their grade spans, their system name, and their accountability status. A pivot table was used to group

schools by grade span and filter school grade spans that were not addressed. The interval for sampling is the inverse of the sampling fraction which is the sample size divided by population or the population divided by the sample size. A randomly selected *n*th term was the starting point with every *k*th term selected; the data were addressed as circular, returning to the top of the list until the sample was complete. If a school was selected and the principal was not a returning principal, the next school on the list was selected.

Instrumentation

The instrument for this study was an online survey administered through Survey Monkey. The survey included three sections. The first section included 32 items with frequency rating scales of 1 (*never)* to 5 (*always*) from Section I of the Leadership Orientations Instrument (Bolman & Deal, 1990). The second section included three items with Likert rating scales of 1 (*strongly agree)* to 6 (*strongly disagree)* from the Kind of Person Instrument—Others (Dweck et al., 1995). The final section collected information on the professional and demographic characteristics of the respondent.

Leadership Orientations Instrument. The first section of the survey included the 32 items from Section I of the Leadership Orientations Instrument—Self (Bolman & Deal, 1990). The Leadership Orientations Instrument (Bolman & Deal) items were used to collect data related to principal dominant leadership frames. Principals were asked to rate the frequency of their behaviors associated with the four leadership frames in implementing the North Carolina teacher evaluation process. This instrument was chosen because of the potential relationships of each of the four frames—structural, human resource, political, and symbolic—to evaluation processes. For example, structural leaders develop effective management systems to manage the requirements related to the

evaluation process; human resource leaders coach, develop, and empower teachers through the evaluation process; political leaders deal comfortably with conflicts that may arise in the evaluation process; and symbolic leaders provide inspiration and motivate individuals to commit to the organizational mission through the evaluation process theoretically (Bolman & Deal). There are two versions to the Leadership Orientations Instrument: a self-survey and a survey to be completed by others. This study used only Section I of the self-survey which also includes two additional sections on leadership plus another section collecting personal data.

In the first section, principals were asked to rate their use of certain behaviors in implementing the teacher evaluation process on a 5-point frequency rating scale with a range of 1 (*never*) to 5 (*always*). The instructions asked principals to rate their use of behaviors specifically in the implementation of the evaluation process. The items were not altered. This section included 32 actions with eight actions related to each of the four frames. "Show high levels of support and concern for others" is an example of a human resource action. "Am able to be an inspiration to others" is an example of a symbolic action. "Develop alliances to build a strong base of support" is an example of a political action. "Set specific, measurable goals and hold individuals accountable for results" is an example of a structural action (Bolman & Deal, 1990).

The reliability of Section I of Bolman and Deal's Leadership Orientations Instrument had been previously established using 1,309 colleague ratings of managers of multiple sectors of business and education. Test score statistics are described in Table 7. High reliability had been established using Cronbach's alpha and the Spearman Brown

Table 7

Test Score Statistics for Section I of Leadership Orientations Instrument

Frame	Mean	Standard Deviation	Standard Error	Cronbach's Alpha	Spearman Brown Coefficient	Total Cases
Structural	32.493	5.703	0.158	.92	.93	1,309
Human resource	32.458	6.303	0.173	.93	.92	1,331
Political	31.391	5.739	0.161	.91	.91	1,268
Symbolic	31.382	6.325	0.174	.93	.93	1,315

Note: Test score statistics for reliability of Bolman and Deal's Leadership Orientations Instrument. Adapted from *Research Using Leadership Orientations Instrument* by Bolman (2013).

coefficient. Cronbach's alpha for the four frames ranged from .91 to .93, and Spearman Brown coefficient for the four frames ranged from .91 to .93 (Bolman, 2010). The Leadership Orientations Instrument—Self (Bolman & Deal, 1990) had been used in research related to elementary and secondary schools supporting its validity as a measure of principal leadership approach (Goodall, 2008; Henrikson, 2007; Penix, 2009; Williams, 2008).

Previous dissertations have examined the relationships between school climate and principal leadership frames, successful principal succession and leadership frames, and school performance and leadership frames (Goodall, 2008; Henrikson, 2007; Penix, 2009). In a survey of principals and teachers from 35 schools, the Pearson correlations between teachers perceptions of principal leadership frames and school climate ratings were significant at the .01 level (two-tailed): structural, r = .748; human resource, r = .796; political, r = .729; and symbolic, r = .790 (Goodall). Goodall's study suggests that there is a correlation among teacher ratings of principal leadership orientations and their ratings of the school.

Another study examined the issue of principal succession using the four frames. Among principals surveyed ($N = 302$), beginning principals and second year principals used significantly fewer frames than principals with 10 or more years' experience ($p = .03$; $p = .02$). This study also found significant relationships ($p < .0005$) among the use of the four frames, the use of multiple frames, meeting challenges, and composite versatility with effect size ranging from .39 to .88. The greatest effect size, .88, was found between political frame and composite versatility; and composite versatility and number of frames used (Henrikson, 2007).

A third study of principals and teachers in West Virginia elementary schools ($n_{principals} = 42$, $n_{teachers} = 247$) compared leadership frames of principals in high performing schools to principals in low performing schools. Teachers from the high performing schools perceived their principals as using each of the four frames more frequently (structural frame, $t = 4.76$, $p < .001$; human resource, $t = 2.78$, $p < .01$; political frame, $t = 3.58$, $p < .001$; symbolic frame, $t = 2.23$, $p < .05$) than did teachers from low performing schools (Penix, 2009). Although empirical research examining the influence of leadership orientation frames on principal mean ratings of teachers was not found, this research suggest that the Leadership Orientations Instrument is an appropriate measure of principal leadership style. Leadership orientation frames have been used to capture relationships between principal characteristics and school climate, teacher perceptions of leadership and school achievement level, and principal experience and use of multiple frames (Goodall, 2008; Henrikson, 2007; Penix).

Regression analysis was used to validate the instrument. A gender analysis of preferences for the four frames was also conducted with no significant differences found

among males and females in the higher education and Singapore samples on other and self-ratings. Among American school administrators the only significant differences were the rating of women significantly higher than men on the structural, political, and symbolic frames by others and the rating of women significantly lower on the political frame by self with $p \leq .05$, two-tailed test (Bolman & Deal, 1991).

Ordinal data was treated as interval data, and using standard practices recommended by Bolman and Deal (1991), the mean of principal self-ratings on the eight items related to each frame were used to determine the dominance of the frame. The rating scale was 1 to 5 with the highest ratings being often and always. Structural frame items were 1, 5, 9, 13, 17, 21, 25, and 29; human resource frame items were 2, 6, 10, 14, 18, 22, 26, and 30; political frame items were 3, 7, 11, 15, 19, 23, 27, and 31; and symbolic frame items were 4, 8, 12, 16, 20, 24, 28, and 32. The number of dominant leadership frames was determined by standardizing the mean rating for principals on each subscale. Ratings meeting the criteria $M + .25SD$ were coded as dominant. SPSS was used to calculate the number of dominant frames and recode into new variable.

Kind of Person Instrument. The second section of the survey used the three-item Kind of Person Instrument, a universal implicit person theory survey (Dweck et al., 1995). This theory measures individual beliefs about the malleability of others' personal characteristics and abilities. Individuals who believe personal characteristics of others are malleable possess an *incremental theory,* but individuals who believe personal characteristics of others are not malleable possess an *entity theory.* Individual beliefs are measured on a continuum. These beliefs are referred to as *implicit person theories* because they are usually difficult for individuals to articulate (Dweck et al., 1995).

This study used three entity theory items from the instrument, "The kind of person someone is is something very basic about them and it can't be changed much;" "People can do things differently but the important parts of who are can't be really be changed;" "Everyone is a certain kind of person and there is not much that can be done to really change that." Only entity theory items were used to minimize the length of the survey. Levy and Dweck (1996) compared the use of only the incremental items to the use of both incremental and entity items. Participants ($N = 101$) completed the survey using both types of items and using only the incremental items with a delay of a week or less between the first and second surveys. There was a correlation of classifying individuals as incremental or entity theorists of .83.

Self-theory research originally addressed theories of intelligence but evolved to also address the following question: Do self-theories influence judgments of others? Specifically, "Would entity theorists, make more extreme judgments about others' traits from a small sample of behavior in the same way that they made global inferences about themselves?" (Levy, et al., 1988, p. 1425). Entity theorists were more likely to make trait judgments without considering the context and less likely to adjust ratings over time (Erdley & Dweck, 1993). Incremental theorists were less likely to use traits to predict behavior (Chiu, et al., 1997). The Kind of Person Instrument was chosen to measure how principal beliefs about others influence their actions because of previous literature demonstrating the influence of implicit person theories on judgments of others. Dweck et al. developed domain-specific surveys regarding morality and intelligence as well as the cross-domain, universal Kind of Person Instrument. In this study the Kind of Person Instrument was chosen because of its potential to capture principal beliefs about the

malleability of others' personal characteristics and abilities. The Kind of Person Instrument uses only three questions, potentially affecting reliability, but the instrument has high internal reliability. Six studies with a total sample size of 638 (n = 69, 184, 139, 121, 93, and 32) were used to test the reliability and validity of the implicit person theory instruments and to compare ratings on measures of implicit person theories related to intelligence, kind of person, and morality. Only four of the studies (n = 69, 184, 93, and 32) used the Kind of Person instrument. No information was provided on the selection of participants for these studies. The Cronbach's alphas ranged from .90 to .96 for the Kind of Person Instrument. Test-retest reliability was weakened by only a two-week interval, but the result was .82 (Dweck et al., 1995).

Locus of control, a theory also emerging from social cognitive theory and self-efficacy, was considered for this study. In selecting the Kind of Person Instrument, the following factors were considered. Locus of control instruments have been more widely used and tested for validity and reliability in research across many fields including health sciences and education. The Kind of Person Instrument used in educational research has been limited with much of the available research focused on student behaviors. The locus of control instrument most closely aligned to purposes of this study was the teacher locus of control instrument developed by Rose and Medway (1981). This instrument includes items such as: "When the grades of your students improve, it is more likely: a. because you found a way to motivate the students, or because the students were trying really hard" (Rose & Medway, p. 181). For purposes of this study, the Kind of Person Instrument items better captured principal attitudes related to the potential for teachers to improve. Additionally, the Kind of Person Instrument required only three items, and

Rose and Medway's instrument consisted of 11 items. The Kind of Person Instrument was chosen because of its potential to better capture principal attitude toward others' abilities and its length.

To measure discriminant validity, Dweck et al. (1995) compared individual responses to other individual characteristics to determine variables other than implicit person theory possibly explaining responses to the Kind of Person Instrument. According to the researchers, none of the tested variables demonstrated a significant relationship to individual responses: gender; age; self-presentation concerns as measured by the Self-Monitoring Scale (Snyder, 1974) and the Social Desirability Scale (Paulhus, 1984); cognitive ability as measured by Scholastic Aptitude Test scores; confidence in intellectual ability as measured by author-created measure; self-esteem (Coopersmith, 1967); view of others and the world (Chiu & Dweck, 1994); and political attitudes (Altemeyer, 1981; Kerlinger, 1984). The value of *r* ranged from .16 to .18, but the significance criteria were not given (Dweck et al.)

The survey only included questions addressing the entity theory. Questions addressing the incremental theory directly were not included because they were found to be "highly compelling" (Dweck et al., 1995, p. 270) and were not necessary. Although revisions to the incremental theory questions increased the reliability of using both incremental and entity theory statements, the instrument can be used with only the statements directly related to entity theory or with statements related to both. Using only the entity theory items, individuals with ratings below 3.0 are considered to hold an incremental theory, and individuals with ratings above 4.0 are considered to hold an entity theory (Dweck, 2000).

Demographic and professional information. For Question 1, principals were asked to provide their North Carolina six-digit school code. For questions 37-41, they were asked to respond to multiple choice and open-ended items asking for their school grade span; growth status; years' experience as a principal, administrator, and teacher; growth status; gender; and ethnicity.

School code. The first question of the survey asked the principals to enter their six-digit school code. Providing this code allowed the researcher to match information regarding teacher ratings from the North Carolina School Report Card and EDDIE to the principal survey responses. Other information available on school report cards included the number of teachers receiving *not demonstrated*, *developing*, *proficient*, *accomplished*, and *distinguished* ratings on each of Standards 1 through 5. Each rating was assigned a value from 0 to 4 with *not demonstrated* being 0 and *distinguished* being 4. The mean rating of each standard was then calculated and used as the criterion variable.

School grade span. Principals were asked to provide their school grade span and classify the school as either an elementary, middle, or high school. Data on grade span were reported using descriptive statistics. Grade span categories of elementary, middle, and high school were a predictor variable coded as dummy variables. Grade level and grade span had influenced value-added data in multiple studies as described in the literature review (Kane & Staiger, 2012; Borman & Kimball, 2005; Eady & Zepeda, 2007; Holtzapple, 2005; Jacob & Lefgren, 2006; Jacob & Lefgren, 2008; Kimball et al., 2004; New Teacher Project, 2009).

Value-added composite. Principals were asked to provide their school value-added composite, or growth status, based on EVAAS value-added models. The value-

added composite is based on SAS EVAAS statistical models used to predict student performance and compare the predicted performance to the actual performance. Statistical concerns with value-added models are described in Chapter 2. WestEd (2010) evaluated value-added models and recommended the EVAAS Univariate Response Model (URM) and the EVAAS Multivariate Response Model (MRM). One advantage of the MRM was multiple teachers could be assigned to a single student (WestEd). The multivariate response model (MRM) is used when possible. When sufficient data to use MRM are not available, univariate response model (URM) is used (Wright et al., 2010). To reduce potential bias with the URM, at least three prior scores must be available (Wright et al.) Test scores are compared using Normal Curve Equivalents (NCE) with the mean for each grade and subject equal to 50. The growth standard for expected growth is zero (Wright et al.). Although these data are now a component of teacher and principal summative evaluations, SAS recommended, "The use of value-added measures as one component of accountability systems is important, but in our view, the diagnostic information is of greater importance" (Sanders, Wright, Rivers, & Leandro, 2009, p. 9). Growth status—did not meet expected growth, met expected growth, and exceeded expected growth—were posted on school report cards and public data. The actual school value-added composite was not publicly available information for 2012-2013.

Experience. In 2011-2012, the mean total years' experience for public school principals in the United States was 7.2 years and the mean years at current school was 4.2 years compared to a mean total years' experience of 5.4 years and a mean years at current school of 2.7 years in North Carolina. Standard deviation was not provided. Among public school principals in the United States, 34% were in their first two years as

principals at their current school (National Center for Education Statistics, 2012). Principals were asked to provide the length of their experience as a principal at any school, as an administrator, and as a teacher rounded to the nearest year. Other studies examined the types of experience prior to becoming an administrator and competence as perceived by the teachers (Ing, 2010; Kimball, 2002: Wise et al., 1984), but type of previous experience was not addressed because the studies suggesting a possible influence on the evaluation process focused on the influence of these previous experiences on teacher reactions. This study focused on the principal as the unit of analysis not the teacher.

Gender and race. Principals were asked to identify their gender, ethnicity, and race. Descriptive data were used to summarize the demographics of the sample. The gender item was dichotomous with choices of male and female. In the United States, 49% of school principals were male and 51% were female, but among elementary principals 41% were male and 59% were female, among secondary schools 72% were male and 29% were female, and among combined grade level schools 60% were male and 40% were female (National Center for Education Statistics, 2008b). Although no studies suggested gender influenced the evaluation process, there were mixed results regarding the influence of gender on leadership orientation. Two studies described earlier in the literature review suggested a relationship between gender and orientation to certain frames with female principals demonstrating a stronger orientation toward human resource (Meade, 1992; Suzuki, 1994) and symbolic frames (Meade). Although these studies suggested a relationship between gender and the human resource frame, a gender analysis conducted by Bolman and Deal (1991) found no significant differences among

leadership orientations of males and females in higher education and among Singapore school administrators. Among the sample of American school administrators women were ranked significantly higher than men by others on three of the frames, and the women rated themselves significantly lower than men on the political frame ($p \leq .05$, two-tailed test).

Items addressing ethnicity and race used categorical responses based on the minimum categories of race and ethnicity established by the U.S. Office of Management and Budget. These categories are based on social categories and were self-identified. The five racial categories were (a) White; (b) Black or African American; (c) American Indian or Alaska Native; (d) Asian; and (e) Native Hawaiian or other Pacific Islander (U.S. Census Bureau, 2012). Individuals were allowed to check multiple races. The required categories for ethnicity, Hispanic or Latino and not Hispanic or Latino, were included as a separate item. The U.S. government, "considers race and Hispanic origin to be two separate and distinct concepts" (U.S. Census Bureau, n.d., heading "Ethnic groups," para. 2).

Procedures

This crosssectional nonexperimental survey design study examined the relationships among principal internal characteristics, contextual or external characteristics, and principal ratings of teachers on the North Carolina teacher evaluation instrument. This section describes procedures for collecting data, handling data, storing data, cleaning and entering data, transforming data, and analyzing data.

Data Collection

Because data were collected from principals during the 2013-2014 school year based on data for the year 2012-2013, only principals who participated in the summative evaluation process the previous year were included in the sampling frame. The database of principals was created from North Carolina's Education Directory and Demographical Information Exchange (EDDIE). A spreadsheet was created using the "Create Custom Reports—Schools" with the following columns included: school number, school name, grade level current, principal first name, principal middle name, principal last name, and principal email address. Schools were organized in this list by school number. The spreadsheet from 2013-2014 was compared to the spreadsheet for 2012-2013. Only 2013-2014 principals who were principals of the same school in 2012-2013 were included in the study. The spreadsheet was sorted by the six-digit school number. The following intervals were used for sampling: elementary schools, $k = 8.36$; middle schools, $k = 3.25$; and high schools, $k = 3.14$. Randomly selected nth terms were the starting points with every kth term selected; the data were addressed as circular, returning to the top of the list until each sample was complete.

Principals were emailed a pre-notification. This letter included the purpose of the study, the relevance of the study, and the importance of their participation in the study. The email met requirements as set forth by The George Washington University's Institutional Review Board. The email described the methods for maintaining confidentiality and provided contact information for recipients to ask further questions. The email explained the school code would only be used to (a) monitor the return surveys and avoid sending reminders to those individuals quickly returning the survey, (b) match

survey responses with data from the school report card, and (c) provide participants with summary results upon completion of the study. Participants were directed to a survey link created through SurveyMonkey. The email requested principals provide their 6-digit school code to facilitate the matching of public data to the principal survey responses and to identify principals not yet responding. Participants were informed the survey was voluntary and the responses were confidential. Once data were matched based on the school code, the school codes were deleted from the database, the principals of the school were removed from a list to receive a follow-up reminder, and they were added to a list to receive a thank you e-mail.

In a meta-analysis of electronic survey research, Cook, Heath, and Thompson (2000) concluded the number of precontacts, number of contacts, personalized contacts, salience, and sponsorship were strategies most related to higher response rates. The following steps were taken to increase return rates:

(1) Precontact was made by email to participants. A paper prenotification is not associated with higher return rates on electronic surveys. Schaefer and Dillman (1998) suggested this might be caused by the failure of participants to associate a paper prenotification with an electronic survey.

(2) All contacts to principals were personalized using cut-and-paste functionality.

(3) Two follow-up contacts were made at two week intervals to individuals not responding.

(4) The salience of this study, prominence and importance, were described in the e-mail.

Data was collected in January and February of 2014.

Data Handling

Storage. Data were collected from www.surveymonkey.com with built-in security measures. Survey Monkey requires unique user names and passwords be entered each time a user logs in; session cookies do not include user names and passwords; and Secure Sockets Layer (SSL) technology authenticates servers and encrypts data. Network security is protected through firewalls, intrusion detection systems, and McAfee programs. Data security is guarded by storage facilities located in the United States; data backups occur hourly and are encrypted. Further information on security of data stored through Survey Monkey can be found at http://www.surveymonkey.com/mp/policy/security/. The data were downloaded into an Excel spreadsheet and stored in a password protected document on a single computer and on a secure external hard drive. Data were analyzed using IBM SPSS (Statistical Product and Service Solutions) Statistics GradPack for Windows.

Cleaning and entry. Of the 73 responses to the survey, 5 were deleted due to either a missing school code preventing the matching of principal data to school data or a response rate less than 80%. Among the other 68 responses, there were no missing data from Section 1 (leadership orientation) and Section 2 (implicit person theory). Two principals provided incorrect responses to principal years' experience—adding years' experience as administrator to years' experience as teacher. Listwise deletion was used to deal with these two cases on regression tests using these variables. Five participants did not respond to questions related to years' experience as administrator and four participants did not respond to questions related to years' experience as a principal. Listwise deletion was used to deal with these cases on tests using these variables. Data

were entered in Excel spreadsheet downloaded from Survey Monkey. The district and school name were used to locate the percent of teachers receiving each rating on each standard using the North Carolina educator effectiveness data available from http://apps.schools.nc.gov/pls/apex/f?p=155:4:0::NO. North Carolina policy prohibits the publishing of data on evaluation standards when less than five teachers in the school were rated on that standard. Policy also requires the ratings of all teachers on Standards 1 and 4. Principals are required to rate career status teachers on Standards 2, 3, and 5 during licensure renewal years. Three principals in this study were associated with schools with less than five teachers receiving ratings on Standards 2, 3, and 5. Listwise deletion was used to handle these cases.

Publicly available data related to school context were matched using the school code. The following data were collected related to school context: Title I status, school size, percent of teacher who were fully licensed, percent of classes taught by highly qualified teachers, percent of teachers with advanced degrees, percent of teachers with less than three years' experience, and teacher turnover. Prior research had not addressed these school characteristics as influencing principal ratings of teachers, but the data were collected to provide a deeper understanding of school context.

Data transformation. Item 1 was used to identify distribution of school ratings from http://apps.schools.nc.gov/pls/apex/f?p=155:4:723532860027701::NO::P4_ SCHOOL:010303 using an Excel spreadsheet. North Carolina publishes the number of teachers rated at each level for each standard. The number of ratings at each level was multiplied by the following values: not demonstrated, 0; developing, 1; proficient, 2; accomplished, 3; distinguished, 4. The mean rating was calculated by dividing the sum

Table 8

Distribution of North Carolina Teacher Evaluation Instrument Ratings 2011-2012

	Not Demonstrated	Developing	Proficient	Accomplished	Distinguished	Mean
Standard 1	54 (0.1%)	*1512 (1.7%)*	*33,208 (36.3%*	*44305 (48.4%)*	*12396 (13.6%)*	*M* =2.74 *SD* = .78
Standard 2	113 (0.2%)	1,292 (2.4%)	21,212 (38.7%)	26,478 (48.3%)	5673 (10.4%)	*M* = 2.66 *SD* = .79
Standard 3	105 (0.2%)	1,406 (2.6%)	25,509 (47.1%)	22,427 (41.4%)	4,684 (8.7%)	*M* = 2.56 *SD* = .80
Standard 4	45 (0%)	1,967 (2.1%)	34,423 (37.6%)	47,015 (51.4%)	7,998 (8.7%)	*M* = 2.67 *SD* = .77
Standard 5	97 (0.2%)	1,431 (2.6%)	25,966 (48.1%)	21,488 (39.8%)	5,035 (9.3%)	*M* = 2.55 *SD* = .80

Note: Distribution of ratings on North Carolina teacher evaluation instrument 2011-2012 with mean. Raw data retrieved from http://apps.schools.nc.gov/pls/apex/f?p=155:5:0::NO on August 3, 2013.

of the products described by the number of ratings on the standard. The 2011-2012 distribution of ratings for North Carolina is provided in Table 8 above. The means of each standard were used because North Carolina does not have a formula for assigning an overall rating.

Principal self-ratings on items 2 through 33 were transformed by recoding into a new variable and using the computing function to calculate the standardized mean of item ratings related to each leadership frame: structural frame items were 1, 5, 9, 13, 17, 21, 25, and 29; human resource frame items were 2, 6, 10, 14, 18, 22, 26, and 30; political frame items were 3, 7, 11, 15, 19, 23, 27, and 31; and symbolic frame items were 4, 8, 12, 16, 20, 24, 28, and 32. Frames were identified as dominant using the criterion $M + .25SD$.

The mean ratings of items 33-35, scale of 1 to 6, were used to identify the principal implicit person theory regarding the malleability of others' personal characteristics and abilities following procedures previously established in research addressed in the literature review (Chiu et al., 1997; Chiu et al., 1999; Gutshall, 2003;

Heslin, 2003; Hong et al., 1997; Levy et al., 1988; Levy & Dweck, 1999). If the mean rating was greater than 4.0, SPSS was used to transform the variable into a new variable identifying the principal as an entity theorist with a dummy code of 1. If the mean rating was less than 3.0, SPSS was used to transform the variable identifying incremental theorists with a dummy code of 1. Respondents with a mean scale between 3.0 and 4.0 were considered neither entity nor incremental theorists and received a dummy code of 0 for both variables.

Data Analysis

The unit of analysis for this study was the principal. Descriptive statistics were used to provide each principal's mean rating for Standards 1 through 5, a demographic profile of principals participating in the survey, a leadership profile of principals participating in the survey, and a profile of the schools represented by the principals participating in the survey. Principal personal characteristics including ethnicity, gender, experience, and content background were collected. Principal number of dominant leadership orientation frames, principal dominant implicit person theory, principal years' experience, school growth status, and school grade span served as predictor variables. Principal mean ratings of teachers on each standard served as the criterion variables. Simple and multiple regression techniques were used to identify the factors significantly predicting principal ratings of teachers.

Multiple Regression/Correlation Analysis (MRC) is "a highly general and therefore very flexible data-analytic system that may be used whenever a quantitative variable (the dependent variable) is to be studied as a function of, or in relationship to, any factors of interest (expressed as independent variables)" (Cohen, Cohen, West, &

Aiken, 2003, p. 3). Regression techniques were chosen because they do not limit the types of variables to be used (Cohen et al.). The variables may be the result of an experiment or naturally occurring (Cohen et al.). The dependent, or criterion, variables of this study were principal dominant leadership frames as measured by Bolman and Deal's (1990) Leadership Orientations Instrument; principal dominant implicit person theories as measured by the three-item universal implicit person theory Kind of Person Instrument (Dweck et al., 1995); principal years' experience as a principal; school grade span; and the growth status based on aggregate teacher value-added data for the school. Other data collected for exploratory reasons were also tested: principal years' experience as an administrator, school Title I status, the percent of teachers in the school with less than three years' experience, and school teacher turnover rate. The criterion variables were the principal mean ratings of teachers on Standards 1 through 5. Simple regression techniques were used to analyze the relationship between principal years' experience as a principal and as an administrator, principal number of dominant leadership orientation frames, percent of teachers with less than three years' experience, and school teacher turnover rate. The use of dummy variables with more than two levels caused multiple regression techniques to be required to analyze the influence of the following variables: implicit person theory, school growth status, and school grade span. Multiple regression techniques were also used to analyze the influence of multiple predictor variables on principal mean ratings of teachers.

Multiple regression techniques require a dependent variable that is measured on a continuous scale (Cohen et al., 2003). For this study, the principal mean rating on each standard were treated as a continuous variable. Five independent variables served as

predictor variables. Linearity of relationships between each predictor variable and criterion variable as well as among all predictor variables and the criterion variable were tested using scatterplots and partial regression plots. Independence and normality were supported by random selection of principals from a pool of principals from all school districts, and independence of observations was tested using SPSS's Durbin-Watson statistic. Normality assumptions were tested by using SPSS to calculate skew and kurtosis, examining histograms, and verifying SPSS tests of normality.

Descriptive statistics were used to compare the demographics of the sample to the demographics of the population of North Carolina principals as contained in EDDIE. Chi square goodness-of-fit tests were used where possible to compare the observed distribution to the expected distribution. The chi square value was calculated using SPSS. The standardized residual for each variable was calculated separately because SPSS does not calculate the standardized residuals for chi square goodness-of-fit tests. Scatterplots of each predictor and criterion variable, residuals and histograms, and partial regression plots were used to identify potential outliers and influential points. Residual versus fitted plots were used to determine homoscedasticity. Tolerance and Variance Inflation Factor (VIF) were used to test for collinearity.

Reliability of instruments was determined using SPSS results. Cronbach's alpha measured the internal reliability of the scales used in the survey. The following parameters were used for reliability analysis: descriptive statistics for item, scale, scale if item deleted, and interitem correlations. Regression tests were conducted using SPSS. The model summary table data, r, R^2, and adjusted R^2 were used to identify the goodness-of-fit of the model; the F ratios in the ANOVA table were used to determine the

significance of the model; and *t* tests determined the significance of the relationship between each predictor and criterion variable.

Chapter 4: Presentation of Results

This chapter presents the results of this study and provides an analysis of the influence of principal and school characteristics on principal ratings of teachers. The predictor variables in the original study design included principal years' experience, principal use of multiple leadership orientation frames, principal beliefs about the malleability of others' personal characteristics and abilities, school grade span, and school growth status. The criterion variables were the mean principal rating of teachers on each of the five teacher evaluation standards with ratings assigned by the principal. Participants in this study consisted of North Carolina principals who assigned teachers their most recent summative evaluation rating; they were selected using stratified, proportional sampling techniques. This chapter presents descriptive statistics related to the population and sample, the predictor variables, and the criterion variables followed by sections summarizing inferential statistics related to each of the research questions. The following overarching question guided this study: Do principal and school characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument? These analytic questions were used to answer the overarching question:

1. Do principal characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?
    a. Do principal implicit person theories as measured by the Kind of Person Instrument (Dweck et al., 1995) influence principal ratings of teachers on the North Carolina teacher evaluation instrument?

b. Does principal use of dominant leadership frames as measured by Bolman and Deal's (1990) Leadership Orientations Instrument influence principal ratings of teachers on the North Carolina teacher evaluation instrument?

c. Does principal total years' experience as a principal influence principal ratings of teachers on the North Carolina teacher evaluation instrument?

2. Do school characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?

   a. Does the school value-added composite measured by EVAAS data influence principal ratings of teachers on the North Carolina teacher evaluation instrument?

   b. Does school grade span as reported by the principal influence principal ratings of teachers on the North Carolina teacher evaluation instrument?

Descriptive Statistics

The data analyzed in this study were collected through an online survey distributed to a stratified, proportional sample of 399 North Carolina public school principals: 226 elementary, 88 middle school, and 85 high school principals. A response rate of 35% was expected based on a meta-analysis conducted by Cook, Heath, and Thompson (2000) of 56 online surveys reported in 39 studies. Although 399 principals were invited to participate in the survey and two follow-up emails were sent to nonrespondents, only 73 principals chose to participate in the survey. Listwise deletion was used to remove five cases because the participants either did not provide a usable school code or did not respond to 80% of the items. Listwise deletion was used to handle missing data for each series of tests as needed. Although a return rate of 35% was

expected, the return rate for this survey was 21% with only 17% responses meeting criteria to be used in data analysis. This section reports frequencies and descriptive statistics related to demographic, predictor, and criterion variables.

Demographic Variables

All demographic information was not completed by all participants. Demographic information collected included gender; race; ethnicity; years' experience as principal, administrator, and teacher; and school grade span. School board region, location type, teacher turnover rate, percent of teachers with less than three years' experience, and Title I status were collected by matching school code to publicly available data.

Among principals responding to the survey, 38 (56%) were from Title I schools. Respondents represented all eight state board of education regions with between 8 and 12 respondents from each region. Respondents were from schools located in cities (19%), suburbs (9%), and towns (10%), but most respondents were from rural schools (62%). This was not statistically significantly different than the expected distribution of 32% city, 51% rural, 9% suburban, and 8% town, $\chi^2 = (5.521, N = 68)$, $p = .137$.

Not all respondents chose to respond to questions regarding race (3%), but the majority of respondents were White (81%) followed by Black/African American (10%). More females (56%) than males (43%) responded. The ethnicity and gender of the population of principals was not available through EDDIE. Respondents represented schools varying greatly in size ($M = 664$, $SD = 426$) compared to a mean of 576 and standard deviation of 383 for all schools in North Carolina.

Predictor Variables

The predictor variables in this study included grade span, school growth status, principal implicit person theory, number of dominant leadership orientation frames, and principal years' experience. Grade span, school growth status, and principal implicit person theory were coded as dummy variables; number of dominant leadership orientation frames and principal years' experience were continuous variables. The following sections provide summaries of descriptive statistics related to each predictor variable.

School grade span. Stratified, proportional sampling was used to select the principals to receive the surveys: 226 elementary, 88 middle, and 85 high school principals of North Carolina public schools responsible for assigning teachers their most recent summative ratings. A chi-square goodness-of-fit test was used to determine if the distribution of principals by grade span followed the hypothesized distribution of 57% elementary, 22% middle school, and 21% high school. The chi-square goodness-of-fit test indicated that the distribution of principals by grade span choosing to participate in this study was not statistically significantly different from the hypothesized distribution, χ^2 (2) = 3.73, p = .155, with 31 (46%) elementary, 18 (27%) middle school, and 19 (28%) high school principals participating.

School growth status. In North Carolina, value-added data are used to classify schools into three categories based on their growth index: did not meet expected growth, met expected growth, and exceeded expected growth. To standardize growth measures across different grade levels and subject areas, SAS calculates the school's growth index by dividing the school's composite value-added by the standard error of measure (SAS,

n.d.). North Carolina classifies schools with a growth index between -2 and +2 as meeting expected growth, schools with a growth index less than -2 as not meeting expected growth, and schools with a growth index greater than +2 as exceeding expected growth. Statewide, 29% of schools exceeded expected growth, 43% of schools met expected growth, and 29% of schools did not meet expected growth. A chi-square goodness-of-fit test indicated that the distribution of principals choosing to participate in this study by school growth status was statistically significantly different from the hypothesized distribution, χ^2 (2) = 9.11, p = .011, with 44% (n = 30, standardized residual = 2.38) of principals representing schools that exceeded growth. The percent of principals representing schools exceeding growth, 28%, was not statistically significantly different, and the percent of principals representing schools not meeting growth, 28%, was not statistically significantly different.

Implicit person theory. Principals were asked to respond to three items related to their beliefs about the malleability of others' personal characteristics and abilities. These items, used with permission from the Kind of Person Instrument (Dweck et al., 1995), were Likert rating scales of 1 (strongly agree) to 6 (strongly disagree). Individuals were classified into three categories based on their mean rating on these three items: mean ratings greater than or equal to 4, incremental theorist; mean ratings less than or equal to 3, entity theorists; and mean ratings between 3 and 4, undetermined. Dweck et al. found that approximately 40-45% of individuals held entity theories, 40-45% held incremental theories, and approximately 10-15% of participants were undetermined. A chi-square goodness-of-fit test using expected distributions of 45% entity, 45% incremental, and 10% undetermined was conducted using SPSS nonparametric testing.

This test indicated that the distribution of principals choosing to participate in this study by implicit person theory was not statistically significantly different from the hypothesized distribution, χ^2 (2) = 0.42, p = .813, with 49% (n = 33) of principals identified as holding an incremental theory, 41% (n = 28) identified as holding an entity theory, and 10% (n = 7) identified as undetermined.

Mean years' experience. Two principals incorrectly responded to the question related to years' experience as a principal. For this question, they added their years' experience as an administrator and their years' experience as a teacher. Their reported years' experience as a principal were greater than their years' experience as an administrator. Listwise deletion was used to deal with these two cases and the five principals who did not respond to this question. In this study, 36% (n = 21) of responding principals (N = 61) had 10 or more years' experience (M = 8.31, SD = 5.74). According to the Schools and Staffing Survey (National Center for Education Statistics, 2009), the mean years' experience of principals in North Carolina was 5.4 years with only 4% of principals having more than 10 years' experience. A t-test was conducted to determine if there was a statistically significant difference between years' experience as a principal of the participants compared to all public school principals in North Carolina. The mean years' experience of participants was greater (M = 8.31, SD = 5.74) than the population (μ = 5.4, σ not available) with a mean difference of 2.91 and a 95% CI [1.44, 4.38].

Dominant leadership frames. Participants responded to the 32-item Leadership Orientations Instrument—Self, Section I, rating their frequency of leadership behaviors (Bolman & Deal, 1990). This scale consists of four subscales that measure participant

Table 9

Criteria Used to Identify Frames as Dominant or Nondominant

Frame	Mean	*SD*	.25**SD*	Mean + .25**SD*
Structural	4.17	0.43	0.11	4.28
Human Resource	4.19	0.45	0.11	4.30
Political	3.78	0.47	0.12	3.90
Symbolic	3.74	0.52	0.13	3.87

frequency of behaviors related to structural, human resource, political, and symbolic leadership frames. To identify dominant leadership frames, principal mean ratings on each subscale were standardized using SPSS. Principals with a mean self-rating on leadership frame subscales .25*SD* above the mean were identified as dominant in that frame. Others were identified as nondominant. The SPSS function "recode into a different variable" was used to assign a value of 1 to dominant frames and a value of 0 to nondominant frames. The means, standard deviations, and ratings used to classify principal leadership orientation frame are provided in Table 9 above. The compute variable function in SPSS was used to calculate the number of dominant frames by adding the values of 0 or 1 assigned to the four frames. Among the principals responding, 23 (34%) had no dominant leadership orientation frame, 12 (18%) had one dominant leadership orientation frame, 13 (20%) had two dominant leadership orientation frames, 14 (21%) had three dominant leadership orientation frames, and 6 (9%) met the criteria to be identified as dominant in all four frames. Frequencies of each dominant frame and combination of dominant frames are displayed in Table 10.

Internal reliability analyses of the Leadership Orientation—Self, Section I, (Bolman & Deal, 2000) for this study and the pilot were conducted using SPSS. The

Table 10

Frequency of Dominant Frames

Dominant Frames	Frequency	Percent
No Frames	**23**	**34%**
One Frame	**12**	**18%**
Human Resource	5	7%
Political	4	6%
Symbolic	3	4%
Two Frames	**13**	**20%**
Political, Symbolic	4	6%
Human Resource, Political	3	4%
Human Resource, Symbolic	2	3%
Human Resource, Structural	2	3%
Structural, Political	2	3%
Three Frames	**14**	**21%**
Human Resource, Political, Symbolic	9	13%
Structural, Political, Symbolic	3	4%
Structural, Human Resource, Political	1	2%
Structural, Human Resource, Symbolic	1	2%
Four Frames	**6**	**9%**

findings were compared to results published by Bolman (2013). Bolman analyzed the internal reliability of each subscale and found Cronbach's alpha values between .91 and .93 for total cases ranging from 1,268 and 1,331. The observed values in this study were Cronbach's alpha between .75 and .80. Details are provided in Table 11. Bolman and Deal's data included participants from education, business, and industry. Some indicators might appear to be more relevant to groups outside of education. For example, the survey item with the lowest mean rating ($M = 3.52$, $SD = .91$) was "am highly imaginative and creative." The survey item with the highest mean rating ($M = 4.38$, $SD = .49$) was "think very clearly and logically." Principals could perceive some items more relevant to their leadership especially in the context of evaluations. Other studies using the Leadership Orientations Instrument have found similar Cronbach's alpha values to both Bolman and Deal's data and observed data in this study: .943 to .951 (Penix, 2009), 247 teacher ratings of 42 principals; .75 to .93 (King, 2009), teacher and community

Table 11.

Bolman and Observed Test Score Statistics for Section I of Leadership Orientations Instrument

Frame	Cronbach's α	Total Cases	Items
Bolman (2013)			
Structural	.92	1,309	1, 5, 9, 13, 17, 21, 25, 29
Human Resource	.93	1,331	2, 6, 10, 14, 18, 22, 26, 30
Political	.91	1,268	3, 7, 11, 15, 19, 23, 27, 31
Symbolic	.93	1,315	4, 8, 12, 16, 20, 24, 28, 32
Observed			
Structural	.75	68	3, 7, 11, 15, 19, 23, 27, 31
Human Resource	.79	68	4, 8, 12, 16, 20, 24, 28, 32
Political	.80	68	5, 9, 13, 17, 21, 25, 29, 33
Symbolic	.80	68	6, 10, 14, 18, 22, 26, 30, 34

Note: Test score statistics for reliability of Bolman and Deal's Leadership Orientations Instrument—Self, Section 1, from Bolman (2013) and data observed in this study in 2014.

member ratings of principals; and .83 to .86 (Suzuki, 1994) included 124 principals. DeVellis (2003) described values between .70 and .80 as respectable, values between .80 and .90 as very good, and values much higher than .90 as indicating the scale might be too long. DeVellis acknowledges the ranges are "personal and subjective" (DeVellis, pp. 95-96).

Criterion Variables

The criterion variables for this study were principal mean ratings of teachers on Standards 1 through 5 of the North Carolina teacher evaluation instrument. Principals rated teachers on the following scale: not demonstrated, developing, proficient, accomplished, and distinguished. For this study, the ratings were assigned values of 0, 1, 2, 3, and 4, respectively. The principal mean rating for each standard was then found by multiplying the number of developing ratings by 1, the number of proficient ratings by 2, the number of accomplished ratings by 3, and the number of distinguished ratings by 4 and dividing by the number of teachers receiving a rating on that standard. Principals were required to rate beginning teachers and probationary teachers on Standards 1

through 5 each year, but tenured teachers were only required to be observed on Standards 1 and 4 in nonrenewal years of their licensure cycle. North Carolina only reported data when five or more teachers received ratings for a standard. Due to fewer teachers being rated on Standards 2, 3, and 5, and state policy prohibiting the publication of data for less than five teachers, ratings for three schools were not published for Standards 2, 3, and 5. For these cases, listwise deletion was used to deal with missing mean ratings. There were no missing data for Standards 1 and 4.

Data were examined using the SPSS explore function. Histograms were used to examine mean ratings for each standard for normality. Box and whisper plots were used to identify outliers. Only one outlier was identified, the maximum value for Standard 5: mean rating (3.94). The next highest value was 3.77. The outlying value was not modified. On each standard, at least 97.83% of teachers received the three highest ratings: proficient, accomplished, or distinguished. One principal in the sample assigned one teacher a rating of not demonstrated on Standard 5. On each standard, between 1.34% and 2.17% of teachers were assigned a rating of developing. Distribution of principal ratings can be seen in Table 12.

Inferential Statistics

Regression tests were conducted to analyze the amount of variability in principal ratings of teachers explained by principal and school characteristics. A combination of simple and multiple regression tests were applied to answer the research questions. For each question, separate regression analyses were conducted for each of the first five standards of the evaluation instrument.

Table 12

Distribution of Population Sample Ratings on the North Carolina Evaluation Instrument 2013-2014

Rating	N. D.		Developing		Proficient		Accomplished		Distinguished	
Assigned value	0		1		2		3		4	
Standard	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
1* (*M* = 2.84, *SD* = .42)	0	0%	38	1%	921	32%	1376	49%	499	18%
2** (*M* = 2.73, *SD* = .46)	0	0%	30	2%	621	39%	721	45%	238	15%
3** (*M* = 2.60, *SD* = .46)	0	0%	35	2%	750	47%	634	39%	191	12%
4* (*M* = 2.71, *SD* = .44)	0	0%	48	2%	1044	37%	1427	50%	315	11%
5** (*M* = 2.65, *SD* = .48)	1	0%	29	2%	715	44%	621	39%	244	15%

Note. *N = 68. **N = 65. N. D. represents not demonstrated

Assumptions for Linear Regression Models

Each series of simple regression tests started by verifying that the data met the assumptions for linear regression: linearity of relationships, no significant outliers, independence of residuals, homoscedasticity of residuals, and normality of residuals. Multiple regression tests included the verification of these assumptions as well as noncollinearity of points (Lund & Lund, 2012). The first assumption, variables should be interval or ratio, was met through the design of the study including the creation of dummy variables for variables that started as nominal for school growth status, school grade span, principal implicit person theory, and Title I status. The criterion variables were the means of ordinal evaluation ratings of teaches treated as interval variables. The second assumption, a linear relationship exists between the variables, was verified by plotting the dependent variables against the independent variables using SPSS and by plotting the studentized residual against the unstandardized predicted values. Most scatterplots were either approximately linear or somewhat linear except as described in the results examining the relationship between principal ratings of teachers and principal number of dominant leadership frames. The third assumption, no significant outliers, was examined using SPSS's casewise diagnostics with ±3 *SD* entered as the criterion for

identifying outliers. For multiple regression tests, influential statistics were identified using Cook's statistics and examined. Independence of residuals was checked using the Durbin-Watson statistics. Durbin-Watson statistics values ranged from 1.80 and 2.00 indicating independence of residuals. Homoscedasticity of residuals was verified using the scatterplots graphing the studentized residual against the unstandardized predicted variables. For multiple regression tests, noncollinearity of dependent variables was verified using correlations of less than .70 and tolerance values of less than 0.10. Normality of residuals was verified by inspecting histograms and P-P plots.

Influence of Principal Characteristics on Principal Mean Ratings of Teachers

The influence of principal implicit person theory, the number of dominant leadership frame, and previous experience on principal mean ratings of teachers were examined through three subquestions to answer Question 1, "Do principal characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" A series of separate simple and multiple regression tests were conducted to answer these questions. This section presents the findings of the statistical tests for each subquestion.

Implicit person theory. Question 1a examined the influence of principal implicit person theory related to principal beliefs about the malleability of personal characteristics and ability: "Do principal implicit person theories as measured by the Kind of Person Instrument (Dweck et al., 1995) influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" To answer research question 1a, a series of separate multiple regression tests were conducted to address the percent of variability in principal ratings of teachers on Standards 1 through 5 explained by principal implicit

person theory related to their beliefs about the malleability of others' personal characteristics and abilities. The mean of each principal's self-ratings on the three items from the Kind of Person Instrument (Dweck et al.) was recoded into dummy variables based on the ranges established by Dweck et al. for identifying individuals with entity (≤ 3) and incremental (≥ 4) beliefs about others. Mean ratings between 3 and 4 were recoded as undetermined. Dummy variables were created and multiple regression tests were conducted using SPSS. Principal beliefs about the malleability of others' personal characteristics and abilities did not significantly influence their mean ratings of teachers on the North Carolina teacher evaluation instrument. A summary of results is found in Table 13. *F* statistics for Standards 1, 2, 4, and 5 were less than 1. See Appendix 7 for implicit person theory scatterplots for each standard.

Dominant leadership orientation frames. Principal responses to the 32 items from the Leadership Orientations Instrument—Self, Section I (Bolman & Deal, 1990) were analyzed to answer Question 1b: "Does principal use of dominant leadership frames as measured by Bolman and Deal's (1990) Leadership Orientations Instrument influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" Principals responded to the 32 items from the Leadership Orientations Instrument—Self, Part I (Bolman & Deal, 1990). The standardized means of the 8 items associated with the four subscales were calculated using SPSS. Respondents with a standardized mean greater than .25*SD* above the subscales mean rating were coded as dominant. SPSS was used to calculate the total number dominant frames to examine the influence of number of dominant frames on principal mean ratings of teachers. Scatterplots graphing the

Table 13

Regression Results Analyzing Influence of Principal Implicit Person Theory on Principal Mean Ratings of Teachers

	Standard 1			Standard 2			Standard 3			Standard 4			Standard 5		
Variable	*B*	*SE*	B	*B*	*SE*	B	*B*	*SE*	B	*B*	*SE*	β	*B*	*SE*	β
Cases	(N = 68, df = 2, 65)			(N = 65, df = 2, 62)			(N = 65, df = 2, 62)			(N = 68, df = 2, 65)			(N = 65, df = 2, 62)		
Entity	.20	.18	.24	.28	.20	.29	.27	.20	.29	.17	.19	.19	.27	.21	.27
Incremental	.15	.18	.18	.24	.20	.26	.16	.19	.18	.08	.19	.10	.26	.21	.26
R^2		.02			.03			.03			.02			.03	
Adj. R^2		-.01			-.01			<.01			-.01			-.01	
F		.65			.97			1.08			.54			.88	
p		.527			.384			.345			.588			.421	

Note: No results were significant at the .05 level.

Table 14.

Regression Results Analyzing Influence of Number of Dominant Leadership .Orientation Frames on Principal Mean Ratings of Teachers

	Standard 1			Standard 2			Standard 3			Standard 4			Standard 5		
Variable	*B*	*SE*	*B*	*B*	*SE*	*B*	*B*	*SE*	*B*	*B*	*SE*	*B*	*B*	*SE*	*B*
(*N* = 68, *df* = 1, 66)															
# of frames	.01	.04	.04	-.01	.04	.02	.02	.04	.07	.01	.04	.05	-.004	,94	-.01
R^2		<.01			<.01			<.01			<.01			<.01	
Adj. R^2		-.01			-.02			-.01			-.01			-.02	
F		.08			.04			.28			.14			.01	
p		.777			.850			.600			,706			,921	

Note: No results were significant at the .05 level.

principal mean ratings of teachers against the number of dominant leadership frames did not depict a linear trend. See Table 14 on previous page for a summary of results and Appendix 8 for scatterplots. Nonlinear tests were also conducted but did not improve the fit. No statistically significant relationships were found.

Principal years' experience. Data collected on principal years' experience as a principal and principal years' experience as an administrator were analyzed to answer Question 1c: "Does principal total years' experience as a principal influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" Regression models were used to explore the influence of principal years' experience on principal mean ratings of teachers on the five standards of the teacher evaluation instrument. Two principals provided incorrect responses for years' experience as a principal; they added their years' experience as an administrator and their years' experience as a teacher. Listwise deletion was used to handle these two cases, five cases where principals did not provide their years' experience, and three cases on Standards 2, 3, and 5 where principal mean ratings of teachers was not available yielding $n = 61$ for Standards 1 and 4, and $n = 58$ for Standards 2, 3, and 5. Principal years' experience significantly influenced principal ratings of teachers on Standard 1, $F(1, 59) = 4.77$, $p = .033$, adj. $R^2 = .06$. For Standards 2 through 5 the principal years' experience as a principal did not significantly predict principal mean ratings of teachers.

Although the original design of the study considered principal years' experience as a principal, experience as an administrator was also collected and more strongly predicted principal mean ratings. A separate series of tests were conducted to analyze the influence of years' experience as an administrator on principal mean years' experience.

Four principals did not provide their years' experience as an administrator. Listwise deletion was used to handle these four cases. Experience as an administrator significantly influenced principal ratings of teachers on Standard 1, 2, and 4 with adj. R^2 values of .18, .06, and .09. See Table 15 for a summary of data related to principal years' experience as a principal and as an administrator.

Influence of School Characteristics on Principal Mean Ratings of Teachers

The second question of this study examined the influence of school characteristics on principal mean ratings of teachers on the North Carolina teacher evaluation instrument: "Do school characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" This question consisted of two subquestions exploring the influence of school growth status and school grade span influencing principal mean ratings of teachers on evaluation instrument.

School growth status. To answer question 2a "Does the school value-added composite measured by EVAAS data influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" principals were asked to respond to an item asking for school growth status as determined by state accountability standards. (Raw composite data were not released publicly for the 2012-2013 school year.) North Carolina classified schools into three categories based on their distance above or below the state mean: met expected growth, exceeded expected growth, or did not meet expected growth. Schools with an index of +2 or greater were classified as exceeded expected growth, schools with an index of -2 or less were classified as did not meet expected growth, and all other schools were classified as met expected growth based on North Carolina's policies for reporting school growth. Only growth status, not growth

Table 15.

Regression Results Analyzing Influence of Principal Years' Experience on Principal Mean Ratings of Teachers

	Standard 1			Standard 2			Standard 3			Standard 4			Standard 5		
Variable	*B*	*SE*	B	*B*	*SE*	B	*B*	*SE*	B	*B*	*SE*	β	*B*	*SE*	β
Years' Experience as Principal															
Cases	(*N* = 61, *df* = 1, 59)			(*N* = 58, *df* = 1, 56)			(*N* = 58, *df* = 1, 56)			(*N* = 61, *df* = 1, 59)			(*N* = 58, *df* = 1, 56)		
Exp._Princ.	.02	.01	.27	.02	.01	.23	.01	.01	.16	.02	.01	.25	.02	.01	.17
R^2		.08			.05			.02			.06			.03	
Adj. R^2		.06			.04			.01			.04			.01	
F		4.77*			3.04			1.38			3.95			1.67	
p		.033			.087			.246			.051			.202	
Years' Experience as an Administrator															
Cases	(*N* = 64, *df* = 1, 62)			(*N* = 61, *df* = 1, 59)			(*N* = 61, *df* = 1, 59)			(*N* = 64, *df* = 1, 62)			(*N* = 61, *df* = 1, 59)		
Exp_Admin	.03	<.01	.36	.02	.01	.27	.02	.01	.19	.03	.01	.33	.02	.01	.24
R^2		.13			.07			.04			.11			.06	
Adj. R^2		.11			.06			.02			.09			.04	
F		9.01*			4.73*			2.17			7.46*			3.70	
p		.004			.034			.146			.008			.059	

Note: *Significant at the .05 level.

index, was publicly available. Dummy variables were created and SPSS multiple regression tests were conducted to analyze the influence of school growth status on principal ratings of teachers. Met expected growth was used as the reference variable using one strategy for identifying the reference variable: select the middle variable as reference variables and use extreme variables in model (Garson, 2006). School growth status did not significantly predict principal mean ratings of teachers on any standard. Complete results are located in Table 16.

School grade span. Question 2b asked, "Does school grade span as reported by the principal influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" To answer this question, schools were classified as elementary, middle, or high school. Dummy variables were created and multiple regression tests were applied using SPSS. Middle school was selected as the reference variable using one strategy for identifying the reference variable when the study does not have a control group: use the middle category for comparisons rather than the extremes (Garson, 2006). School grade span did not significantly predict principal mean ratings of teachers on any standard. A summary of results is located in Table 17. All *F* values were less than 1 which can be the result of lack of fit or a potential violation of assumptions. Scatterplots depicting the relationship between principal mean ratings and school grade span are displayed in Appendix 9.

Additional school characteristics as exploratory factors. In addition to school grade span and growth status, data were matched to Title I status, school turnover rate, and the percent of beginning teachers in the school. A series of regression tests were conducted to analyze the influence of Title I status on principal ratings of teachers. Title

Table 16.

Regression Results Analyzing Influence of School Growth Status on Principal Mean Ratings of Teachers

	Standard 1			Standard 2			Standard 3			Standard 4			Standard 5		
Variable	*B*	*SE*	B	*B*	*SE*	β	*B*	*SE*	B	*B*	*SE*	B	*B*	*SE*	B
	(N = 68, df = 2, 65)			(N = 65, df = 2, 62)			(N = 65, df = 2, 62)			(N = 68, df = 2, 65)			(N = 65, df = 2, 62)		
Did not meet	-.07	.14	-.08	-.15	.16	-.15	-.06	.15	-.06	-.07	.15	-.07	-.07	.17	-.06
Exceeded	-.10	.12	-.12	-.16	.14	-.17	-.22	.14	-.24	-.16	.13	-.18	-.13	.15	-.13
R^2		.01			.02			.05			.02			.01	
Adj. R^2		-.02			-.01			.02			-.01			-.02	
F		0.33			0.74			1.54			0.78			0.37	
p		.719			.481			.223			.463			.693	

Note: *Significant at $p < .05$.

Table 17.

Regression Results Analyzing Influence of School Grade Span on Principal Mean Ratings of Teachers

	Standard 1			Standard 2			Standard 3			Standard 4			Standard 5		
Variable	*B*	*SE*	B	*B*	*SE*	β	*B*	*SE*	B	*B*	*SE*	B	*B*	*SE*	B
	(N = 68, df = 2, 65)			(N = 65, df = 2, 62)			(N = 65, df = 2, 62)			(N = 68, df = 2, 65)			(N = 65, df = 2, 62)		
Elementary	.04	.13	.05	.07	.15	.07	.01	.14	.01	-.01	.13	-.01	.14	.15	.14
High school	-.06	.14	-.06	-.08	.16	-.07	.06	.16	.06	-.03	.15	-.03	-.04	.16	-.04
R^2		.01			.02			< .01			<. 01			.03	
Adj. R^2		-.02			-.02			-.03			-.03			>-.01	
F		0.31			0.52			0.10			0.02			0.86	
p		.738			.596			.907			.976			.429	

Note: *Significant at $p < .05$.

I status was not a significant predictor of principal ratings of teachers on any standard. All *F* values were less than 1. The assumption of linearity was not clearly met. Scatterplots graphing principal mean ratings against Title I status are depicted in Appendix 10.

A series of regression tests were also conducted to analyze the influence of teacher turnover rate on principal ratings of teachers. School teacher turnover rate significantly predicted principal ratings of teachers on all standards with adjusted R^2 values of .11, .14, .12, .17, and .06. A third series of regression tests were conducted to explore the influence of the percent of beginning teachers (teachers with less than three years' experience) on principal mean rating of teachers. Results were significant on Standards 1, 2, and 4 with adjusted R^2 values of .05, .09, and .08. See Table 18 for a summary of results.

Influence of Principal and School Characteristics on Principal Ratings of Teachers

The overarching question "Do principal and school characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" was answered using the criterion variables demonstrating a statistically significant influence on principal mean ratings of teachers: principal years' experience as an administrator, teacher turnover rate, and percent beginning teachers. Principal years' experience as an administrator was chosen instead of years' experience as a principal because of greater statistically significant influence on principal mean ratings. The models were statistically significant for Standards 1 through 4 with adjusted R^2 values of .15, .20, .13, and .23, but only the variable teacher turnover rate was significant at the .05 level for each model as indicated by the *t*-tests A summary of results are in Table 19.

Table 18.

Regression Results Analyzing Influence of Exploratory School Characteristics on Principal Mean Ratings of Teachers

	Standard 1			Standard 2			Standard 3			Standard 4			Standard 5		
Variable	*B*	*SE*	B	*B*	*SE*	β	*B*	*SE*	B	*B*	*SE*	B	*B*	*SE*	B
Influence of Title I Status															
Cases	(*N* = 68, *df* = 1, 66)			(*N* = 65, *df* = 1, 63)			(*N* = 65, *df* = 1, 63)			(*N* = 68, *df* = 1, 66)			(*N* = 65, *df* = 1, 63)		
Title_I=Y	.06	.10	.07	.09	.12	.09	<.01	.12	<.01	.06	.11	.07	.08	.12	.08
R^2		< .01			< . 01			< .01			< .01			.01	
Adj. R^2		-.01			-.01			-.02			-.01			-.01	
F		0.31			0.53			<0.01			0.31			0.40	
p		.581			.469			.985			.580			.530	
Influence of Teacher Turnover Rate															
Cases	(*N* = 68, *df* = 1, 66)			(*N* = 65, *df* = 1, 63)			(*N* = 65, *df* = 1, 63)			(*N* = 68, *df* = 1, 66)			(*N* = 65, *df* = 1, 63)		
TTO Rate	-.02	.01	-.36	-.02	.01	-.40	-.02	.01	-.37	-.02	.01	-.43	-.02	.01	-.27
R^2		.13			.16			.13			.18			.07	
Adj. R^2		.11			.14			.12			.17			.06	
F		9.12*			11.23*			9.31*			14.10*			4.55*	
p		.004			.001			.003			<.001			.037	
Influence of % Beginning Teachers															
Cases	(*N* = 68, *df* = 1, 66)			(*N* = 65, *df* = 1, 63)			(*N* = 65, *df* = 1, 63)			(*N* = 68, *df* = 1, 66)			(*N* = 65, *df* = 1, 63)		
% BT	-.01	<.01	-.25	-.01	.01	-.33	-.01	.01	-.22	-.01	< .01	-.30	-.01	.01	-.22
R^2		.06			.11			.05			.09			.05	
Adj. R^2		.05			.09			.03			.08			.03	
F		4.15*			7.32*			3.05			6.32*			2.94	
p		.046			.009			.086			.014			.092	

Note: *Significant at $p < .05$.

Table 19.

Regression Results Analyzing Influence of Principal and School Characteristics on Principal Mean Ratings of Teachers

	Standard 1			Standard 2			Standard 3			Standard 4			Standard 5		
Variable	*B*	*SE*	B	*B*	*SE*	β	*B*	*SE*	B	*B*	*SE*	B	*B*	*SE*	B
	(*N* = 61, *df* = 3, 57)			(*N* = 58, *df* = 3, 54)			(*N* = 58, *df* = 3, 54)			(*N* = 61, *df* = 3, 57)			(*N* = 58, *df* = 3, 54)		
Teacher Turnover	> -.01	< .01	-.11	-.01	.01	-.21	> -.01	.01	-11	-.01	<.01	-.17	-.01	.01	-.11
% Beginning Teachers	-.02	.01	-.31	-.02	.01	-.37	-.02	.01	-.37	-.02	.01	-.40	-.01	.01	-.25
Experience as Admin.	.02	.01	.20	.01	.01	.07	< .01	.01	.01	.01	.01	.14	.01	.01	.09
R^2		.19			.24			.17			-.27			.11	
Adj. R^2		.15			.20			.13			.23			.06	
F		4.54*			5.71*			3.729*			6.89*			2.23	
p		.006			.002			.016			<.001			.095	

Note: Results significant at the .05 level.

Summary

The purpose of this study was to examine the influence of principal and school characteristics on principal mean ratings of teachers on the North Carolina teacher evaluation system. The original design of the study addressed five predictor variables selected based on the literature review: principal years' experience as a principal, principal use of dominant leadership frames, principal implicit person theory, school growth status, and school grade span. Other variables collected included school teacher turnover rate, the percent of teachers in the school who were beginning teachers, principals years' experience as an administrator, and school Title I status. Of the 399 principals invited to participate in the survey, only 68 usable responses were returned.

For principal characteristics, principal years' experience as an administrator showed a significant relationship to principal ratings. The results of this series of regression tests indicates principal experience as a principal was a predictor of a small amount (6%) of the variability in principal ratings of teachers on the North Carolina teacher evaluation instrument for Standard 1, $F\,(1, 59) = 4.77$, $p = .033$. Experience as an administrator was a predictor of a small amount (4-10%) of the variability in principal ratings of teachers on the North Carolina teacher evaluation instrument significant at the .05 level with adj. R^2 values on Standards 1, 2, and 4 of .13, .06, and .11. More years' experience as an administrator resulted in higher mean ratings of teachers. The number of dominant leadership orientation frames and principal implicit person theory were not predictors of principal mean ratings of teachers on the North Carolina Teacher evaluation instrument at the .05 level.

For school characteristics, the predictor variables in the original model—school grade span and growth status—were not significant predictors of principal ratings of teachers at the .05 level. Factors collected by matching school data with significant relationships to principal ratings of teachers were teacher turnover rate and the percent of teachers with less than three years' experience. Title I status did not influence principal ratings of teachers at the .05 level. Teacher turnover rate and the percent of teachers who were beginning teachers influenced principal ratings of teachers at the .05 level. Teacher turnover rate significantly influenced principal mean ratings of teachers on Standards 1 through 5: Standard 1, F (1, 63) = 9.18, p = .004; Standard 2, F (1, 60) = 11.23, p = .001; Standard 3, F (1, 60) = 9.31, p = .003; Standard 4, F (1, 63) = 14.10, p < .001; and Standard 5, F (1, 60) = 4.55, p = .037. Higher teacher turnover rates led to lower mean ratings. The percent of teachers in a school who were beginning teachers significantly influenced principal mean ratings of teachers on Standards 1, 2, and 4: Standard 1, F (1, 63) = 4.15, p = .046; Standard 2, F (1, 60) = 7.32, p = .009; and Standard 4, F (1, 63) = 6.32, p = .014. A higher percent of beginning teachers resulted in lower mean ratings.

This was an exploratory study and identifying significance at .05 instead of a more restrictive value increased the potential for Type I errors. There were F values of less than 1 which can indicate a violation of assumptions. Scatterplots are provided in Appendixes 7-10 for tests with F values of less than 1.

Chapter 5: Interpretations, Conclusions, and Recommendations

Limited research has examined the influence of principal and school characteristics on principal ratings of teachers. Most research examining factors influencing principal ratings of teachers has focused on student characteristics such as socioeconomic status and prior achievement (Borman & Kimball, 2005; Jacob & Lefgren, 2008; Kimball, 2002) or teacher characteristics such as content area, grade level, or career status (Gallagher, 2004; Harris & Sass, 2009; Holtzapple, 2005; Kane et al., 2011; Wise et al., 1984). The purpose of this study was to analyze the influence of school and principal characteristics on principal mean ratings of teachers on the North Carolina teacher evaluation instrument.

The conceptual framework for this study integrated social cognitive theory (Bandura, 1986), implicit person theory (Dweck et al., 1995), leadership orientation frames (Bolman & Deal, 1990), and the North Carolina teacher evaluation system. The overarching framework for this study was social cognitive theory and Bandura's (1986) premise that human behavior (principal ratings of teachers) cannot be separated from personal characteristics of the individuals involved (the principals) and the environmental characteristics of the context (the school). Two theoretical frameworks were incorporated to explore specific principal characteristics: implicit person theories to examine the influence of principal beliefs about the malleability of others' personal characteristics and abilities (Dweck et al., 1995) and dominant leadership orientation frames (Bolman & Deal, 1990) to examine the influence of principal leadership behaviors particularly the use of a multi-frame approach. The North Carolina teacher evaluation instrument provided the dependent variables of the study. The evaluation instrument

contains six standards with ratings on the first five standards assigned by the principal and ratings on the sixth standard based on teacher value-added data (North Carolina Public Schools, 2012).

Predictor variables included principal years' experience as a principal, principal number of dominant leadership orientation frames, principal implicit person theory related to the malleability of others' personal characteristics and abilities, school grade span, and school growth status. Other data were collected related to school characteristics and explored in Chapter 4 of this study. These data included school Title I status, school teacher turnover rate, principal years' experience as an administrator, and the percent of teachers in a school with less than three years' experience. Participants were selected for this study using a stratified, proportional sample of principals who assigned teachers their most recent summative ratings on the North Carolina teacher evaluation instrument. Although 399 principals were invited to participate in the survey, only 68 usable responses were received and not all of these cases contained needed information for all regression tests. The participants were asked to complete a survey that included 32 items from the Leadership Orientations Instrument—Self, Section I (Bolman & Deal, 1990), 3 items from the Kind of Person Instrument—Others (Dweck et al., 1995), 3 items related to years' experience, 1 item related to grade span, 1 item related to growth status, and 3 items related to demographics. They were also asked to provide their school codes which were used to match school data available from North Carolina school report cards and the Educational Directory and Demographical Information Exchange (EDDIE). Data were collected in January and February 2014 using SurveyMonkey's online survey.

This chapter provides a contextual interpretation of the findings related to the research questions addressed in Chapter 4. The theoretical and conceptual frameworks were used as the basis for interpreting these findings. Potential policy implications and considerations, limitations of this study, and suggestions for further research are presented.

Findings and Interpretations

This section examines the critical findings of the study and interprets findings based on the literature reviewed in Chapter 2 and the theoretical and conceptual frameworks for the study. Critical findings related to the study sample, principal characteristics, school characteristics, and exploratory factors are discussed in the following section.

Study Sample

The response rate for this survey, 17%, was lower than expected when compared to an expected response rate of 35% (Cook, Heath, & Thomson, 2000). Several factors possibly contributed to the lower than expected response rate. Three large, urban districts in North Carolina held policies limiting access to personnel for research projects possibly contributing to the lower than expected response. Among those responding, 62% of respondents represented rural schools. Race to the Top and North Carolina's Accountability and Curriculum Reform Effort led to simultaneous changes in math, science, social studies, language arts, healthful living, world languages, and arts education curricula; teacher, principal, and support staff evaluation processes; and state accountability requirements. The changes were accompanied by budget cuts and loss of personnel due to the recession. The weather possibly contributed to lower response rates

with higher than normal snowfall rates and districts across the state missing days each week of the window for holidays, teacher workdays, or snow days. Other reasons possibly limiting response rates included the request that principals provide their school codes to allow for matching of their responses to public data. Although school codes were deleted once data were matched, some principals might have held reservations regarding anonymity. Finally, emails inviting principals to participate were from The George Washington University. Principals might have felt more comfortable responding especially as related to providing their school codes if the invitation to participate had been from a North Carolina school. Restricted access to principals in larger districts, sweeping changes, reduced resources, weather, concerns about anonymity, and invitation from out-of-state source possibly contributed to the lower than expected response rate.

Principal mean years' experience was higher than expected when compared to mean years' experience for all principals in North Carolina. This study only included principals responsible for assigning the most recent summative ratings to teachers. The assignment of ratings occurred in May and June of 2013. Data collection procedures did not begin until January 2014. This gap in time meant principals who left at the end of the 2012-2013 school year or during first semester of the 2013-2014 school year were not included in this study. The time lapse eliminated principals in their first year as principals or in their first year at their schools from the sample possibly explaining some of this discrepancy. The percent of principals from schools exceeding expected growth was statistically significantly higher than expected as described possibly affecting the results of the study.

Principal Characteristics

The influence of principal implicit person theory, principal use of dominant leadership orientation frames, and principal years' experience on principal mean ratings of teachers was explored in this study. The first two principal characteristics tested did not significantly influence principal mean ratings. The first characteristic examined principal beliefs about the malleability of others' personal characteristics and abilities. Principals responded to three items from the Kind of Person Instrument—Others (Dweck et al., 1995). Dweck et al.'s procedures for classifying individuals as having an entity theory or incremental theory were used. Entity theorists believe a person's characteristics and ability cannot be changed much, but incremental theorists believe a person's characteristics and ability can be changed over time (Dweck et al.). The distribution of principals choosing to participate in this study by implicit person theory was not statistically significantly different from the hypothesized distribution with 49% holding an incremental theory, 41% holding an entity theory, and 10% identified as undetermined. Distribution of implicit person theories among principals revealed a pattern of beliefs similar to the general population unlike earlier implicit person theory research with teachers as participants with a higher percent of teachers holding incremental theories (Gutshall, 2013).

The influence of principal implicit person theory on principal mean ratings of teachers was not statistically significant on any standard. The literature suggested entity theorists and incremental theorists would evaluate others differently with entity theorists more likely to stereotype others, assign more extreme ratings, and incremental theorists more likely to change judgments of others over time (Levy et al., 1998; Levy et al., 2001;

Chiu et al., 1997). Other data related to principal ratings of teachers could be examined for influences particularly differences in ratings of individual teachers over time. Dispersion statistics such as standard deviation and interquartile ranges could be examined in future studies to analyze the influence of implicit person theories on the dispersion of ratings.

The second series of tests explored the influence of principal self-reported leadership behaviors. Bolman and Deal (1990) identified four leadership frames—structural, human resource, symbolic, and political—used by leaders to make meaning of data. Bolman and Deal wrote successful leaders must be able to use multiple frames (Gladwell, 2005; Bolman & Deal, 2003). Bolman and Deal (1991) found significant correlations at the .001 level between effectiveness as a manager and the use of multiple frames, $R^2 = .71$, and effectiveness as a leader and the use of multiple frames, $R^2 = 72$. Principals responded to 32 items from the Leadership Orientations Instrument—Self, Section I (Bolman & Deal, 1990). Total number of dominant frames was calculated for each principal using standardized means with 38% of principals holding no dominant frame, 18% of principals of holding one dominant frame, 20% of principals holding two dominant frames, 21% of principals holding three dominant frames, and 9% of principals holding four dominant frames. Principal mean ratings of teachers were plotted against the number of dominant frames used. No linear trend was observed. Nonlinear tests were also conducted with no significant findings. Principal beliefs about the malleability of others' personal characteristics and abilities and their use of leadership behaviors from multiple frames could influence principal ratings in other ways including the dispersion of ratings across standards and among teachers.

The final two series of tests examining the influence of principal characteristics explored the influence of principal experience on principal mean ratings. Principals were asked to provide their years' experience as a principal and administrator. Participant years' experience as a principal (M = 8.31, SD = 5.74) predicted a small amount (6%) of variability in principal ratings on Standard 1, but did not significantly influence principal ratings on Standards 2 through 5. Participant years' experience as an administrator (M = 12.55, SD = 5.53) predicted 13% of the variability in principal mean ratings of teachers on Standard 1, 6% of the variability in principal ratings on Standard 2, 11% of the variability in principal ratings on Standard 4, and 6% of the variability in principal ratings on Standard 5. Other explanations of these relationships could be that principals with more experience could be at schools with more effective teachers, principals with more experience might create schools with more effective teachers, or principals might stay longer at schools with better teachers.

Previous studies suggested three reasons for this influence: (a) different purposes among veteran versus novice evaluators (Bryan & Currin, 1995); (b) more lenient ratings when evaluator is focused on formative purposes and building relationships (Piggot-Irvine, 2003; Rowe, 2000; Sartain et al., 2011); and (c) different attitudes toward the evaluation system (Tziner, et al., 2002). Bryant and Currin found differences in how experienced and novice evaluators rated teachers. Experienced evaluators focused more on providing formative feedback and building partnerships with teachers, but novice evaluators focused more on acting as a monitor. Other studies from the literature review suggested principals would rate more leniently when motivated by improving teaching and preserving relationships (Piggot-Irvine; Rowe; Sartain et al.). Veteran North

Carolina administrators would have used the Teacher Performance Appraisal Instrument-Revised (TPAI-R) from 2000 to 2011 (NC Public Schools, n.d.), and principals comfortable with the TPAI-R instrument might approach the new evaluation system differently. Tziner, et al. (2002) concluded that evaluator attitudes toward evaluation instruments can influence their ratings of employees. This finding suggests veteran principals who were comfortable with the previous instrument might perceive the new instrument differently than novice principals influencing their ratings of teachers.

School Characteristics

The influence of school growth status and school grade span on principal mean ratings of teachers was examined in this study. The first school characteristic explored in this study was school growth status based on school value-added composite. School growth status did not significantly predict principal mean ratings of teachers, but possibly influenced who chose to participate in the study. A higher percentage of principals from schools labeled as exceeding expected growth by North Carolina responded than expected with 44% of principals representing schools that exceeded growth, 28% representing schools that met growth, and 28% representing schools that did not meet growth. Although school growth status as determined by EVAAS value-added models did not significantly influence principal mean ratings of teachers, this study did not examine the influence of the percent of students scoring a Level 3 (proficient) or higher. Jacob and Lefgren (2008) found the percent of students scoring proficient or higher had a stronger relationship to principal ratings of teachers than growth based on value-added data.

Although prior research suggested principals rated teachers in different grade levels differently, school grade span did not significantly predict principal mean ratings of teachers on any standard. Earlier studies found differences in correlation between principal rating and teacher effect at different grade levels. Because only school data were collected, only comparisons among school grade spans could be made. Teacher evaluation data is only published by school. Grade level data was not available. The distribution of principals by grade span participating in this study was not statistically significantly different from the hypothesized distribution, with 31 elementary, 18 middle school, and 19 high school principals participating. In 2014, 57% of public schools in North Carolina were elementary, 22% were middle, and 21% were high schools.

Exploratory Factors

Additional tests were conducted to explore the influence of school Title I status, school teacher turnover rate, and the percent of teachers in a school with less than three years' experience. Title I status was not a significant predictor of principal ratings of teachers on any standard, but school teacher turnover rate and the percent of teachers with less than three years' experience in a school did produce statistically significant results. School teacher turnover rate significantly predicted between 11% and 17% of the variability in ratings for Standards 1 to 4, but only 6% of the variability in Standard 5 ratings. Standard 5 addresses teacher reflection and no indicators are considered observable. Standards 1 to 4 each contain some observable indicators related to leading in their classrooms, establishing a welcoming environment, knowing their content, and facilitating learning. The instrument can be found at http://www.ncpublicschools.org /docs/effectiveness-model/ncees/instruments/teach-eval-manual.pdf.

The percent of teachers with less than three years' experience significantly influenced principal mean ratings of teachers on Standards 1, 2, and 4 with adjusted R^2 values of .04, .09, and .03. Although there was not a statistically significant correlation between teacher turnover rates and percent beginning teachers, schools with higher turnover rates and higher percent beginning teachers have more teachers new to the school. Principals may perceive veteran teachers as having more ability than novice teachers (Holtzapple, 2005; Sartain et al., 2011), assign higher ratings on all standards to veteran teachers they consider to be good teachers (Harris & Sass, 2009), and allow teacher performance from previous years to influence their current ratings (Whitehurst, Chingos, and Lindquist, 2014). Whitehurst et al. found observation scores had a higher correlation (.65) from Year 1 to Year 2 than value-added scores (.38).

The literature review described theoretical literature calling for differentiated models for evaluating teachers (Darling-Hammond, 1990; Marzano et al., 2011). In this study, teacher years' experience in general and years' experience at the school were significant influences on principal mean ratings. These findings suggest principal could be de facto differentiating by ratings new teachers, whether to teaching or to their schools, differently than veteran teachers.

Principal and School Characteristics

The overarching question, "Do principal and school characteristics influence principal ratings of teachers on the North Carolina teacher evaluation instrument?" was answered using variables demonstrating a statistically significant influence on principal mean ratings of teachers: principal years' experience as an administrator, teacher turnover rate, and percent beginning teachers. The models were statistically significant for

Standards 1 through 4 predicting a small amount of variability, but only the variable teacher turnover rate was significant at the .05 level for each model. The smaller than expected response rate reduced the statistical power of the research model and could explain the lack of statistical significance for models based on multiple criterion variables. The model for Standard 5 was not significant. Standard 5 addresses teacher reflection, and no indicators are considered observable. Standards 1 through 4 each contain some observable indicators.

Limitations

This study examined the influence of principal and school characteristics on principal mean ratings of teachers, but limitations included those associated with quantitative studies, regression design, sampling frames, the principal as unit of study, self-reported measures, and lower than expected response rate.

Quantitative design and regression. This study was limited by the overall regression design. This quantitative study did not address principal motivation and processes that might be better considered in a qualitative study. The study used multiple regression tests to answer the research questions which allowed the researcher only to examine predictive, not causal, relationships. Some relationships between criterion and predictor variables were not linear. This study only examined principal mean ratings of teachers and did not consider dispersion of ratings across standards or among teachers.

Exploratory study. This study was an exploratory study with five predictor variables in the original study design and three exploratory factors. For each of these eight variables, a separate series of regression tests were conducted for Standards 1 through 5 resulting in 40 separate series of tests. An additional 5 tests were conducted

exploring a model that included multiple predictor variables. Because this was an exploratory study, a significance level of .05 was used increasing the likelihood of Type I errors.

Principal as unit of study. This study was limited by the unit of study. Using the principal as the unit of study meant teacher evaluation data were collected in the aggregate. One reason the principal was chosen as the unit of study was because aggregate rating data are public information, but individual teacher ratings are protected data. The aggregate data allowed the researcher to examine the influence of the predictor variables on principal mean ratings of teachers, but did not allow the researcher to examine the influence of predictor variables on principal ratings of teachers as related to individual teacher factors such as content area and grade level. For example, this study did not distinguish among principal ratings of English language arts, mathematics, science, and nontested content area teachers. Case studies described in Chapter 2 in Cincinnati Public Schools, Washoe County School District, Vaughn Elementary School, and a Florida School district found different strengths of correlation between principal ratings of teachers and teacher value-added data across content areas and grade levels (Gallagher, 2004; Harris & Sass, 2009; Holtzapple, 2005; Kimball et al., 2004; Milanowski, 2004). Student classroom characteristics were not considered due to the principal as the unit of study. These characteristics, shown in previous studies to correlate to principal ratings, include student prior achievement levels and socioeconomic status (Jacob & Lefgren, 2008; Borman & Kimball, 2005).

Sampling frame and response rate. Another limitation of this study was the sampling frame and response rate. To participate in this study, principals had to be the

principal who assigned teachers their most recent summative evaluation ratings. This criterion meant principals who had assigned the summative ratings and then chose to leave or were reassigned were not included in the sampling frame. Among the principals meeting the criteria and randomly selected to participate in the study, a lower-than-expected response rate of only 17% of the principals emailed met the completion criteria to be included in the study.

Self-reported measures. Because this study used a survey, validity of the results depend on the accuracy of principal responses. Self-reported measures were used to capture information related to principal implicit person theory and dominant leadership orientation frames. A review of data indicated two principals responded inaccurately to the question related to their previous years' experience.

Generalizability of Study

The generalizability of this study is limited by the parameters of the study. This study only addressed principal ratings on the North Carolina teacher evaluation instrument. This instrument is not based on a more universal instrument such as Danielson's (2013) Framework for Teaching. Principals who were responded tended to have more experience than the population, were predominantly white, were from small to midsize districts, and more were from schools exceeding growth than would be expected based on the population. Principals did represent different grade spans, Title I versus non-Title I schools, and state board of education regions.

Recommendations for Further Study

This exploratory study indicated potential for further research. Recommendations include future studies designed to address temporal effects, rating dispersion, teacher factors, proficiency data, and qualitative data.

Temporal effects. Implicit person theory suggests that entity theorists are more likely to stereotype individuals and less likely to change ratings over time, and incremental theorists are more likely to change ratings over time particularly to reflect improvement (Molden et al., 2006). This study only examined one year's summative data and did not compare changes in teacher ratings over time. A quantitative study using hierarchical linear modeling to examine the change in teacher rating over time as the criterion variable and principal implicit person theory as the predictor variable should be conducted.

Rating dispersion. This study only examined principal mean ratings, but did not consider the dispersion of ratings by the principal. Implicit person theory research suggests entity theorists would assign more extreme ratings (Chiu et al., 1997). A quantitative study should be conducted examining the influence of principal leadership frames on dispersion of ratings across standards and among teachers using standard deviation, interquartile range, or other measure of statistical dispersion.

Principal and teacher factors. Principal previous type of experience as described by content area and grade level has been shown to influence the evaluation process. A quantitative study should be conducted to examine the influence of principal previous instructional experience on principal ratings of individual teachers comparing

principal ratings of teachers of similar content areas and subject areas to ratings of teachers from different content areas and subject areas than the principal.

Proficiency data. This study included only school growth status as determined by EVAAS value-added models. School proficiency composites based on the percent of students at Level 3 or higher were not included in this study. Whitehurst et al. (2014) found 38% of teachers with a mean incoming achievement level of students in the top quintile were rated in the top quintile, and 9% of teachers with a mean incoming achievement level of students in the bottom quintile were rated in the bottom quintile. Implicit person theory research suggests entity theorists under cognitive load are not likely to consider situational factors when making judgments of others (Levy et al., 1998; Levy et al., 2001; Chiu et al., 1997). A quantitative study should be could be conducted comparing the predictive power of school growth status to school proficiency composite on principal mean ratings of teachers.

Qualitative data. This quantitative study only captured self-reported principal characteristics and compared to raw data captured from public sources. The study was not designed to collect data on principal motivation or provide a voice to principals related to the implementation of the evaluation system. This study did not triangulate data related to principal reported behaviors, principal observed behaviors, and teacher perceptions of principal behavior. A qualitative study examining principal purposes in evaluating teachers should be conducted to capture principal voice and better capture the factors influencing principal ratings of teachers. This study addressed a largely unexplored area of research.

Implications of Research

Social cognitive theory (Bandura, 1986) suggests reciprocal interactions exist among personal characteristics, external factors, and motivations and actions. This study investigated an area mostly neglected in existing empirical literature—the role of principal and school characteristics in principal ratings of teachers—and should be considered exploratory in nature. Implications of this research include a new lens for exploring teacher evaluations, policy improvements, and professional development and resources for principals.

New lens for exploring teacher evaluations. This quantitative study used simple and multiple regression techniques to determine the influence of principal and school characteristics on principal ratings of teachers. This study is important because no studies were found analyzing the influence of principal characteristics on their ratings of teachers. Previous studies focused on finding the correct teacher factors to create a model explaining evaluation ratings. Varying strengths of correlation between principal ratings and value-added data have been found in studies, but no studies explicitly explored the relationships between principal and school characteristics and principal ratings of teachers. Researchers have examined perceptions of evaluation processes, described the evaluation process, and identified problems associated with the process, but they have not focused on a key player in the process, the principal.

Policy implications. The National Council on Teacher Quality (2013) reported 20 states connect tenure to evaluation ratings, 19 states design professional development around evaluation results, 8 states connect teacher evaluation to licensure advancement, and 6 states connect evaluation data to teacher salaries. With the implementation of more

rigorous evaluation practices and higher stakes connected to evaluations, data collected through this research could be useful to policy makers in amending evaluation processes, departments of education in developing professional development for principals related to teacher evaluation, and institutions of higher education in preparing future principals.

As a result of the Race to the Top grant, North Carolina now publishes the percent of teachers assigned each rating for each standard as part of each school's report card. Providing this information without fully understanding factors beyond teacher instructional performance offers a misleading snapshot of schools. A school's distribution of teacher ratings does not adequately capture the quality of teachers in the school. A better understanding of factors influencing principal ratings of teachers should help policy makers in developing and revising evaluation processes, informing communities about the quality of teachers in their schools, and adopting policies in some way connected to teacher ratings such as awarding tenure, offering multi-year contracts, and developing merit pay systems. The North Carolina Department of Public Instruction and local education authorities can benefit from a better understanding of factors influencing principal ratings of teachers in developing future professional development related to evaluation processes, particularly efforts aimed to improve interrater reliability. Higher education principal administration programs can benefit from a better understanding of the principal's role in the evaluation process to better prepare future principals.

Recommendations for policy and practice. School context, teacher factors other than effectiveness, and principal factors influence principal ratings of teachers. If evaluation systems are to have stakes attached including contracts and merit pay or if

evaluation ratings or to be published as a measure of teachers or schools, then policies and processes should be in place ensuring that evaluation ratings are reliably measuring what is intended.

Recommendations for policy. In studies described in the literature review, teachers expressed concern with using one evaluation instrument for both formative and summative purposes (Bradshaw, 2002; Kimball, 2002; Piggot-Irvine, 2002). Bosetti (1994) described lack of alignment among purposes for evaluations as impeding the effectiveness of evaluation processes. To promote effective evaluation processes, evaluation instruments should be developed with a clear purpose, and instruments should clearly measure what is intended. Components of evaluation processes serving both formative and summative purposes should be clearly identified and aligned to both sets of purposes.

Evaluator attitudes toward evaluation instruments can influence their ratings (Tziner et al., 2002). Principals may consider some standards more clearly related to teacher effectiveness (Glickman et al., 2010; Sangora, 2012; Weisberg et al., 2009). Glickman et al. recommended low-inference standards for more effective evaluation processes. Some principals may perceive some standards as less relevant indicators of teacher quality. For example, third grade teachers demonstrating how they are preparing their students to graduate from high school, or teachers teaching lessons that integrate national and professional standards when they will be measured based on their effectiveness in teaching state standards. Evaluation processes should be simplified with "power standards" best capturing teacher qualities impacting student learning. North Carolina's standards can be complex and inferential. Support materials clearly defining

what each level of each indicator might look like should be developed for principals to better interpret different levels of performance.

States should provide timelines for implementation and resources for training that promote interrater reliability before the evaluation instrument is implemented statewide. Some organizations and schools districts including Memphis City Schools, Tennessee; Princie George County Schools, Mayland; and Partnerships to Uplift Communities Schools require a form of initial and/or ongoing certification process of evaluator rating accuracy (Jerald, 2012). Resources should be provided to allow this work to be ongoing with inside and outside observers comparing and discussing ratings. Statisticians and researchers at the state level should look at principal years' experience and teacher years' experience when examining the validity and reliability of principal ratings of teachers. Evaluation processes should be differentiated to allow principals time to coach and develop struggling teachers.

Recommendations for practice. Recommendations for practice focus on recommendations for professional development, audits, and professional development that includes reflection on audit results. The literature review included studies that found differing correlations among principal ratings and teacher effectiveness at different grade levels and content areas with some teachers perceiving principals as unable to provide adequate feedback in their content areas (Gallagher, 2004; Harris & Sass, 2009; Holtzapple, 2005; Kane et al., 2011; Kimball, 2004; Milanowski, 2004). Principals need training that will allow them to confidently and competently evaluate and provide feedback to all teachers in their schools. Principals do not rate teachers with same teacher effect equally (Borman & Kimball, 2005; Jacob & Lefgren, 2008), and they need

a better understanding of factors possibly influencing their ratings of teachers of the same effect differently. For example, principals, particularly those with an entity theory, should understand the tendency to not consider situational factors when evaluating others especially when experiencing cognitive load. The online evaluation system should provide the capacity to compare principal ratings of teachers to student achievement data.

Sartain et al. (2011) identified a tendency of some principals to rate all teachers higher when compared to ratings assigned by outside observers. A component of professional development for principals should include calibrating principal evaluation ratings with outside observer ratings. Because principal years' experience as an administrator influence principal ratings of teaches at a statistically significant level, ongoing professional development for principals should include reflections in heterogeneous groups that include veteran evaluators, novice evaluators, internal evaluators such as principals and assistant principals, and external evaluators such as central office staff on teacher ratings. To further improve calibration of ratings, resources clearly defining what each level of an indicator might look like in classrooms of different content areas and grade levels should be provided to evaluators. Four current practices are being used to provide ongoing calibration of ratings:

> Deep-dive training for groups of observers focused on specific dimensions of the observation instrument; one-on-one coaching provided by school system leaders or expert consultants; paired observations of live or video-recorded lessons; and group calibration sessions based on live or video-recorded lessons, sometimes using videoconferencing, to allow large groups to view, score, and discuss a live lesson together. (Jerald, 2012, p. 4)

Finally, processes and products should be developed that promote principal self-reflection on their ratings. Principals and superintendents should be provided with audit data each year describing the dispersion of ratings, comparing ratings to teacher

effectiveness data, and providing longitudinal trends in principal ratings and teacher effectiveness at the teacher level. Principal ratings of teachers tend to be more stable than teacher effectiveness because principals tend to avoid assigning teachers lower ratings than they have received in the past to protect relationships with these teachers (Kluger & DeNisi, 1996). Longitudinal reports for principals detailing the stability of their ratings of teachers over time and comparing these ratings to teacher effectiveness over time should be provided. Principals can rate teachers based on each teacher's strengths and weaknesses; ratings do not indicate differences in performance of different teachers (Cleveland et al., 1989). Dispersion statistics of principal ratings and comparisons of dispersions of principal ratings and teacher value-added data should be provided to principals. Finally, evaluator attitudes can influence the rating process (Tziner et al., 2002). Principal reflections should address the potential influence of their attitude toward the evaluation system or even toward certain indicators on their ratings of teachers.

Conclusion

The findings of this study suggest that experience matters. Principal years' experience as an administrator, teacher turnover rate, and the percent of beginning teachers in a school predict a small amount of variability in principal mean ratings of teachers. Principals with more years' experience tended to rate teachers higher. This relationship could be attributed to principal purpose in evaluation and attitude toward the evaluation instrument. Principals with more experience tend to adopt a more formative approach to teacher evaluation and are more concerned with principal-teacher relationships than with a more summative approach focused strictly on monitoring teacher performance (Piggot-Irvine, 2003; Rowe, 2000; Sartain et al., 2011). Principals

in schools with higher turnover rates and higher beginning teacher concentrations tended to rate teachers lower. Although not direct measures of teacher years' experience, higher ratios of beginning teachers and higher teacher turnover rates indicate schools with more teachers who are either new to the school or new to teaching.

This study was an exploratory study that has revealed opportunities for future research. Future studies should examine other predictor variables as well as consider other criterion variables that better capture the dispersion of principal ratings. There are other principal and school factors not tested such as principal teaching certification areas and school overall percent proficient that could predict principal mean ratings, but there are also other data to be examined describing principal ratings including statistics related to the dispersion of ratings and change in ratings over time. Finally, this study focused inherently on evaluators in the role of principal. Future research should consider the influence of other evaluator roles including assistant principal, central office staff, or instructional coach. A better understanding of these relationships should lead to improved policy and practices related to teacher evaluation.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago Public High Schools. *Journal of Labor Economics, 25*(1), 95-135. Doi: 0734-306X/2007/2501-0004

Altemeyer, B. (1981). *Right-wing authoritarianism.* Winnipeg: The University of Manitoba Press.

American Recovery and Reinvestment Act of 2009, Pub. L. 111-5, § 14005-6, Title XIV.

Aseltine, J. M., Faryniarz, J.O., & Rigazio-Digilio, A.J. (2006). *Supervision for learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

Ballou, D. (2008a). *Value-added assessment: Lessons from Tennessee.* Retrieved July 12, 2013, from http://dpi.state.nc.us/docs/superintendents/quarterly/2010-11/20100928/ballou-lessons.pdf

Ballou, D. (2008b). *Test scaling and value-added measurement.* Retrieved July 12, 2013, from http://www.wcer.wisc.edu/news/events/VAM%20Conference%20Final%20Papers/TestScaling_Dballou.pdf

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Blackwell, L. S., Trzesniewski, K.H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263.

Bolman, L. (2013). *Research using or influenced by Bolman and Deal's four frames.* Retrieved from http://www.leebolman.com/four_frame_research.htm

Bolman, L. & Deal, T. (1990). *Leadership orientations instrument.* Retrieved from http://www.leebolman.com/leadership_research.htm

Bolman, L. & Deal, T. (1991). Leadership and management effectiveness: A multi-frame, multisector analysis, *Human Resource Management*, 30, 509-534.

Bolman, L.G., & Deal, T.E. (2003). *Reframing organizations: Artistry, choice, and leadership* (3rd ed.). San Francisco: Jossey-Bass.

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & Kenzie, S. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48*(3), 587-05.

Borman, G. & Kimball, S. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close achievement gaps? *Elementary School Journal, 106*(1), 3-20.

Bosetti, L. (1994). Official policy and truncated practice: A need to reconceptualize the evaluation of teachers. *School Organization, 14*(1), 49-52.

Bradshaw, L. (2002). Local district implementation of state mandated teacher evaluation policies and procedures: The North Carolina case. *Journal of Personnel Evaluation in Education, 16*(2), 113-128.

Brown, R. (2010). 18 states and District of Columbia are finalists for education grants. *The New York Times.* Retrieved from http://www.nytimes.com/2010/07/28/education/28education.html?_r=1&ref=education

Bryant, M., & Currin, D. (1995). Views of teacher evaluation from novice and expert evaluators. *Journal of Curriculum and Supervision, 10*(3), 250-261.

Chiu, C., Hong, Y., & Dweck, C. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality & Social Psychology, 73*(1), 19-30.

Chow, A., Wong, E., Yeung, A., & Mo, K. (2002). Teachers' perceptions of appraiser-appraisee relationships. *Journal of Personnel Evaluation in Education, 16*(2), 85-101.

Cleveland, J. N., Murphy, K.R. & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology*, *74*(1), 130-135.

Cogshall, J. (2012). Toward the effective teaching of new career- and college-ready standards: Making professional learning systemic. Washington, D.C.: The National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/TowardEffectiveTeaching.pdf

Cohen, J., Cohen, P. West, S. G., & Aiken, L. (2003). Applied multiple regression/correlation analysis in the behavioral sciences. Thousand Oaks, CA: Sage.

Coles, J. (2005). *Multiple perspectives in alternative school leadership development: A study of three first-year urban principals*. (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Full Text. (304997986).

Conley, S., Muncey, D., & You, S. (2005). Standards-based evaluation and teacher career satisfaction: A structural equation modeling analysis. *Journal of Personnel Evaluation in Education, 18*, 39-65.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement, 60*(6), 821-836.

Coopersmith, S. (1967). *The antecedents of self-esteem.* San Francisco: Freeman.

Cubberly, E. (1916). *Public school administration.* Boston: Houghton Mifflin.

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

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

Darling-Hammond, L. (1990). *Handbook of teacher evaluation: Assessing elementary and secondary school teachers.* London, UK: Sage Publications.

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

DeVellis, R. (2003) Scale development: Theory and applications. Thousand Oaks: Sage Publications.

Dixon, A. (2011). Focus on reform legislation in SREB states: Evaluation policies. Southern Regional Education Board. Retrieved from http://publications.sreb.org/2011/11S07_Focus_Teach_Eval.pdf

Duke, D. L., & Stiggins, R.J. (1990). Beyond minimum competence: Evaluation for professional development. In J. Millman & L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers* (pp. 133-146). Newbury Park, CA: Corwin Press, Inc.

Durocher, E. A. (1995). *Leadership orientations of school administrators: A survey of nationally recognized school leaders*. (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Full Text. (304190984).

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048. Doi: 10.1037/0003-066X.41.10.1040

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development.* Philadelphia: Psychology Press.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6,* 267-285.

Eady, C. K., & Zepeda, S. J. (2007). Evaluation, supervision, and staff development under mandated reform: The perceptions and practices of rural middle school principals. *The Rural Educator, 28*(2), 1-7.

Eckley, M. (1997). *The relationship between teacher empowerment and principal leadership styles.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses (9802625).

Elmore, R. F. (2004). *School reform from the inside out: Policy, practice, and performance.* Cambridge, Mass: Harvard Education Press.

Erdley, C. S. & Dweck, C. S. (1993). Children's implicit theories as predictors of their social judgments. *Child Development, 64,* 863-878.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160.

Fowler, F. J. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage.

Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education, 79*(4), 79-107.

Garson, G. D. (2006). *Structural equation modeling.* Retrieved from http://www2.chass.ncsu.edu/garson/pa765/structur.htm

Jerald, C. (2012). *Ensuring accurate feedback from observations: Perspectives from practice.* Seattle: Bill and Melinda Gates Foundation. Retrieved from http://www.gatesfoundation.org/college-readyeducation/Documents/ensuring-accuracy-summary.pdf

Gervey, B., Chiu, C., Hong, Y, & Dweck, C. S. (1999). Differential use of person information in decision-making about guilt vs. innocence: The role of implicit theories. *Personality and Social Psychology Bulletin, 25*, 17-27.

Gladwell, M. (2005). *Blink: The power of thinking without thinking.* New York: Little, Brown and Co.

Glickman, C., Gordon, S. P., & Ross-Gordon, J. (2010). *SuperVision for Successful Schools. SuperVision and Instructional Leadership: A Developmental Approach* (8th ed.). Boston: Allyn & Bacon.

Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers.* New York: Holt, Rinehart and Winston.

Goodall, S. H. R. (2008). *Perceptions of leadership and school climate through eyes of principals and teachers of Appalachian No Child Left Behind Blue Ribbon Schools, (NCLBBRS).* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (304448279)

Gutshall, C.A. (2013). Teachers' mindsets for students with and without disabilities. *Psychology in the Schools*, *51*(10), 1073-1083.

Hallinger, P. (1983). Principal instructional management rating scale. Pelham, NY: Leading Development Associates.

Harris, D. & Sass, T. (2009). *What makes for a good teacher and who can tell?* Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research. The Urban Institute. Retrieved from www.caldercenter.org.

Hassan, A. (2007). Human resource development and organizational values. *Journal of European Industrial Training, 31*(6), 435-448.

Henrikson, H. A. (2007). *The relationship between the use of multiple leadership frames and the meeting of principal succession challenges.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (304715067)

Heslin, P. A. (2003). The effect of prior judgments and implicit person theory on performance appraisals (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (NQ84768)

Heslin, P., Latham, G., & VandeWalle, D. (2005). The effect of implicit person theory on performance appraisals. *Journal of Applied Psychology, 90*(5), 842-856. Doi:10.1037/0021-9010.90.5.842

Holtzapple, E. (2005). Criterion-related validity evidence for a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(3), 207-219.

Hunter, M. (1984). Knowing, teaching and supervising. In Philip Hosford (ed.) *Using What We Know About Teaching.* Alexandria, VA: ASCD.

Ing, M. (2010). Using informal classroom observation to improve instruction. *Journal of Educational Administration, 48*(3), 337-358. Doi: 10.1108/09578231011041053

Jacob, B. A., & Lefgren, L. (2006). When principals rate teachers: The best and the worst stand out. *Education Next, 6*(2), 58-64.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics, 26*(1), 101-136.

Joint Committee on Standards for Educational Evaluation. (2014). Personnel evaluation standards. Retrieved from http://www.jcsee.org/personnel-evaluation-standards

Kane, T. & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Author. Retrieved from http:// www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Evaluating teacher effectiveness: Can classroom observations identify practices that raise achievement? *Education Next*, *11(*3). Retrieved April 1, 2012, from http://educationnext.org/evaluating-teacher-effectiveness/

Kerlinger, F. N. (1989). Liberalism and conservatism: The nature and structure of social attitudes. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kerrins, J. A., & Cushing, K. S. (2000). Taking a second look: Expert and novice differences when observing the same classroom teaching segment a second time. *Educational Assessment, Evaluation and Accountability, 14*(1), 5-24.

Kimball, S. M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education, 16*(4), 241-268.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education, 79*(4), 54-78.

King, B. W. (2006). *Relationship of principal's leadership behaviors to academic achievement and school improvement efforts*. (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Full Text. (305300875).

Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: Historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254-284.

Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research, 48*(1), 1-20. Doi:10.1080/00131880500498297

Larsen, M. (2009). Stressful, hectic, daunting: A critical policy study of the Ontario Teacher Performance Appraisal System. *Canadian Journal of Educational Administration and Policy,* (95), 1-44.

Leithwood, K. (1992). The move toward transformational leadership. *Education Leadership, 49*(5), 8-12.

Leithwood, K., Louis, K., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning. Learning from Leadership Project, the Wallace Foundation.* Retrieved from http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/How-Leadership-Influences-Student-Learning.pdf

Levy, S., Plaks, J., Hong, Y., Chiu, C., & Dweck, C. (2001). Static versus dynamic theories and the perception of groups: Different routes to different destinations. *Personality & Social Psychology Review (Lawrence Erlbaum Associates), 5*(2), 156-168.

Levy, S., Stroessner, S., & Dweck, C. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*(6), 1421-1436. Doi:10.1037/0022-3514.74.6.1421.

Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development, 70*(5), 1163-1181.

Lund, A., & Lund, M. (2012). Laerds Statistics. Retrieved from statistics.laerd.com

Martinez, M. C. (2006). *Sources of school leadership: Their influence on school capacity, teaching, and learning*. (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Full Text. (305291159).

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

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy 4(*4), 572-606.

Meade, R. E. (1992). *The California School Leadership Academy: Its effects on the leadership orientation of California elementary school principals*. (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Full Text. (304003751)

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

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79*(4), 33-53.

Molden, D. C., Plaks, J. E., & Dweck, C. S. (2006). "Meaningful" social inferences: Effects of implicit theories on inferential processes. Journal Of Experimental Social Psychology, 42(6), 738-752. Doi:10.1016/j.jesp.2005.11.005

Morgan, G. (2006). *Images of organization.* London: Sage.

Morsh, J. E. & Wilder, E. W. (1954). Identifying the effective instructor: A review of the quantitative studies 1900-1952. *Research Bulletin AFPTRC-TR-54-44.* Chanute Air Force Base, Illinois: Air Force Personnel and Training Research Center.

National Center for Education Statistics. (2004). *Average number of public school teachers in public school districts who were dismissed in the previous year or did not have their contracts renewed based on poor performance, by years of teaching experience, and state: 2003–04.* Retrieved from http://nces.ed.gov/surveys/sass/tables/state_2004_40.asp on May 13, 2013

National Center for Education Statistics. (2009). *Characteristics of public, private, and Bureau of Indian Education elementary and secondary school principals in the United States: Results from the 2007–08 Schools and Staffing Survey.* Retrieved from http://nces.ed.gov/surveys/sass/tables/sass0708_2009323_p1s_07.asp on May 13, 2013.

National Center for Education Statistics. (2012). *Characteristics of public, private, and Bureau of Indian Education elementary and secondary school principals in the United States: Results from the 2011–12 Schools and Staffing Survey.* Retrieved from http://nces.ed.gov/surveys/sass/tables/sass0708_2009323_p1s_07.asp on May 13, 2013.

National Center for Education Statistics. (2012). *Number of teachers per public school district in 2011–12, average number of teachers per district in the previous year (2010–11) who were dismissed or did not have their contracts renewed for any reason, and as a result of poor performance, by tenure status of teachers and state: 2011–12.* Retrieved from http://nces.ed.gov/surveys/sass/tables/ sass1112_2013311_ d1s_008.asp on May 13, 2013

National Council on Teacher Quality. (2013). *State of the states 2013: Connect the dots: Using evaluations of teacher effectiveness to inform policy.* Retrieved from http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

National Governors Association. (2011). *Issue brief: Preparing principals to evaluate teachers.* Washington, D. C.: NGA Center for Best Practices.

No Child Left Behind Act. (2001), 20 U.S.C.§ 6301.

North Carolina Professional Teaching Standards Commission. (2007, June). *North Carolina professional teaching standards.* Retrieved July 6, 2010, from http://www.ncpublicschools.org/docs/profdev/standards/teachingstandards.pdf

North Carolina Public Schools. (2006). *4-Year Cohort Graduation Rate.* Retrieved from http://accrpt.ncpublicschools.org/app/2006/cgr

North Carolina Public Schools. (2008a). *Implementation plan.* Retrieved July 6, 2010, from fvbgmhttp://www.ncpublicschools.org/profdev/training/

North Carolina Public Schools. (2008b). *2008 cohort graduation rate released.* Retrieved November 3, 2010, from http://www.ncpublicschools.org/newsroom/news/2008-09/20080807-03

North Carolina Public Schools. (2010). *Public school employee evaluation.* Retrieved July 30, 2010, from http://www.ncpublicschools.org/fbs/personnel/evaluation/

North Carolina Public Schools. (September 7, 2011). *Briefing on educator effectiveness policies* (Prepared for the members of the Governor's Education Transformation Commission). Retrieved February 13, 2013, from http://ncees.ncdpi.wikispaces.net/file/view/Educator%20Effectiveness%20Policies%20Briefing%202011.pdf/253331532/Educator%20Effectiveness%20Policies%20Briefing%202011.pdf

North Carolina State Board Of Education. (2008). Policy adopting the North Carolina teacher evaluation rubric and process for teacher evaluation, TCP-C-004. Retrieved July 6, 2010 from http://www.ncpublicschools.org/docs/profdev/training/teacher/teacher-evaluation.pdf

Ovando, M., & Ramirez, A. (2007). Principals' instructional leadership within a teacher performance appraisal system: Enhancing students' academic success. *Journal of Personnel Evaluation in Education, 2007*(20), 85-110. Doi:10.1007/s11092-007-9048-1

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology. 46,* 598-609.

Painter, S. R. (2000). Principals' perceptions of barriers to teacher dismissal. *Educational Assessment, Evaluation and Accountability, 14*(3), 253-264.

Penix, G. E. (2009). *A study of teacher perceptions of the relationship between leadership styles of principals in high and low performing West Virginia elementary*

schools. (Doctoral Dissertation). Retrieved from Proquest Dissertations and Theses. (89183626).

Peterson, K. D. (2004). Research on school teacher evaluation. *NASSP Bulletin, 88*(639), 60-79.

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

Peterson, K. D., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation. *Educational Assessment, Evaluation and Accountability*, *14*(2), 135-153.

Piggot-Irvine, E. (2003). Key features of appraisal effectiveness. *International Journal of Educational Management, 17*(4), 170-178.

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

Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy, 4*(4), 492-519. Doi:10.1162/edfp.2009.4.4.492.

Rose, J. S. & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcomes. *Journal of Educational Research, 74,* 185-190.

Rowe, B. (2000). *The influence of teacher efficacy and readiness for self-directed learning on the implementation of a growth-oriented teacher performance appraisal process.* Paper presented at the meeting of the Annual Conference of the American Educational Research Association. New Orleans, LA. Retrieved from http://www.eric.ed.gov.proxygw.wrlc.org/PDFS/ED444942.pdf

Rowe, B. W. (2004). *A case study of the influence of teacher efficacy and readiness for self-directed learning on the implementation of a growth-oriented teacher performance appraisal process in on school district.* (Doctoral dissertation). Retrieved from Proquest database. (NR01753)

Sagona, R. (2012). *An examination of principal consistency in evaluating teachers.* (Doctoral dissertation). Retrieved from Proquest database. (3532959)

St. Maurice, H., & Cook, P. (2005). Towards standards for instructional supervision: A genealogy of standards. In S. P. Gordon (Eds.), *Standards for instructional supervision: Enhancing teaching and learning*). Larchmont, NY: Eye on Education, 1-14

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* University of Tennessee Value-Added Research and Assessment Center, Knoxville. Retrieved from http://www.cgp.upenn.edu/pdf/Sanders_Rivers-TVASS_teacher%20effects.pdf

Sanders, W. L., Wright, S. P., Rivers, J. C., & Leandro, J. G. (2009). A response to criticisms of SAS EVAAS. SAS White Paper. Cary, NC: SAS.

Sartain, L., Stoelinga, S., Brown, E., Luppescu, S., Matsko, K., Miller, F., Durwood, C., Jiang, J., & Glazer, D. (2011). Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation. Chicago: Consortium on Chicago School Research at the University of Chicago Urban Education Institute.

SAS. (n.d.). *School effectiveness.* Retrieved December 27, 2012, from http://www.sas.com/govedu/edu/k12/evaas/index.html

Sawyer, L. (2001). Revamping a teacher evaluation system. *Educational Leadership*, *58*(5).

Schaefer, D. F., & Dillman, D. A. (1998). Development of a standard e-mail methodology: Results from an experiment. *Public Opinion Quarterly, 62,* 378-397.

Sergiovanni, T. J. (2002). Managing a school: Theoretical and practical aspects. Tel Aviv: Open University.

Soper, D.S. (2013). A-priori sample size calculator [Software]. Available from http://www.danielsoper.com/statcalc

Suzuki, Y. (1994). A comparative study of the leadership orientation of frames of California Asian principals and other public school principals. (Doctoral Dissertation). Retried from Proquest Dissertations and Theses. (9426165).

Tziner, A., Murphy, K. & Cleveland, J. (2002). Does conscientiousness moderate the relationship between attitudes and beliefs regarding performance appraisal and rating behavior? *International Journal of Selection and Assessment*, 10, 218-224.

U.S. Census Bureau (n.d.). Ethnic groups. *American FactFinder Glossary.* Retrieved May 11, 2013, from http://factfinder2.census.gov/help/en/american_factfinder_help.htm#glossary/glossary.htm

U.S. Census Bureau. (2012). *What is race?* Retrieved May 11, 2013, from http://www.census.gov/population/race/

Weisberg, D., Sexton, S., Mulhern, J. and Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Brooklyn, N.Y.: The New Teacher Project. Retrieved January 20, 2012, from http://widgeteffect.org/downloads/TheWidgetEffect.pdf.

WestEd. (2010). Selected statistical models for incorporating student academic growth in estimating effect of teachers in tested grades and subjects: Technical comparison matrix. Retrieved from http://www.ncpublicschools.org/docs/effectiveness-model/evaas/selection/technical-comparrison.pdf

Whitehurst, G. J., Chingos, M. M., & Lindquist, K. M. (2014). Evaluating teachers with classroom observations: Lessons learned in four districts. Washington, D.C.: Brown Center on Education Policy at Brookings.

Williams, J. D. (2008). *The indicators of engaging leadership for private Christian schools in Texas.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (304602631)

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Berstein, H. T. (1984). Teacher evaluation. A study of effective practices. *Elementary School Journal*, *86*(1), 61-121.

Wright, S. P., White , J. T., Sanders, W. L., & Rivers, J. C. (2010). SAS EVAAS statistical models. Cary, NC: The SAS Institute.

Youngcourt, S., Leiva, P., & Jones, R. (2007). Perceived purposes of performance appraisal: Correlates of individual- and position-focused purposes on attitudinal outcomes. *Human Resource Development Quarterly, 18*(3), 315-343. Doi:10.1002/hrdq

Appendix 1: Exempt from IRB Review Request Form

THE GEORGE WASHINGTON UNIVERSITY
OFFICE OF HUMAN RESEARCH
INSTITUTIONAL REVIEW BOARD

EXEMPT FROM IRB REVIEW REQUEST FORM

Before completing this form, complete the Human Subject Research Determination worksheet to ensure that you are in fact required to submit your new study to the Office of Human Research. The OHR will only review studies deemed "human subject research."

Reporting Proposed Changes in Research: This exempt from IRB review determination only applies to this form/protocol, as currently proposed. Therefore, if there are any changes that increase the risks to subjects (e.g., methodology, data gathering instruments, type of information being accessed or disclosed, etc.) the changes must be submitted to the IRB/OHR for approval PRIOR TO implementation.

INVESTIGATOR AND TEAM CONTACT INFORMATION			
IRB# *(ADMIN USE ONLY--WILL BE ASSIGNED UPON SUBMISSION)*		VERSION DATE:	10/15/2013
TYPE OF HIPAA AUTHORIZATION REQUESTED:		Not applicable	
PROTOCOL TITLE AND SPONSOR:			
TITLE :The Influence of Principal and School Characteristics on Principal Ratings of Teachers using the North Carolina Teacher Evaluation Instrument			
SPONSOR :Unfunded			
PRINCIPAL INVESTIGATOR INFORMATION *(MUST BE GWU FACULTY)*			
LAST NAME:	CLAYTON	FIRST NAME:	JENNIFER Degree: Ph.D.
DEPARTMENT	EDUCATIONAL LEADERSHIP	SCHOOL:	GSEHD
CAMPUS ADDRESS:	1 OLD OYSTER POINT ROAD STE. 200 NEWPORT NEWS, VA 23602		
PHONE:	757-269-2203	EMAIL:	CLAYTONJ@GWU.EDU
PRINCIPAL CONTACT IF OTHER THAN PI: *(THIS MAY BE THE STUDENT/TRAINEE)*			
LAST NAME:	LEGGETT	FIRST NAME:	STACY

Recommendation:

- ☐ Study Registered as Exempt. Category: _____
- ☐ A HIPAA waiver of research subject authorization is justified for this study under 45 CFR 46 164.512 based on the following criteria:
  1. The proposed uses and disclosures of protected health information (PHI) involve no more than minimal risk to the privacy of individuals.
  2. The research could not practicably be conducted without the waiver.
  3. The research could not practicably be conducted without access to and use of the PHI.

  Please obtain permission from the privacy officer of the health care organization in which you will access protected health information before beginning your research.
- ☐ This research does **NOT** meet the regulatory/institutional requirements for exemption from IRB review. To conduct this research you must complete an IRB submission package for IRB review. For more information on completing a research submission, contact OHR at 202-994-2715.

Authorized Designee _____ Signature _____ Date _____

This Exempt Registration does not expire nor does it require renewal.

CAMPUS ADDRESS:	1103 W. MAIN STREET, WILLIAMSTON, NC 27892		
PHONE:	252-370-7422	EMAIL:	SRLEG@GWMAIL.GWU.EDU

Select the category that describes the proposed research activity:

The exemptions outlined below do not apply to ANY research involving prisoners. Research involving children may be exempt with specific restrictions. See below:

☐ 1. Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as research on instructional strategies; or research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

☒ 2. Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless:

The information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and

Any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

This category may not be applied to children, except in the observation of public behavior.

☐ 3. Research involving the use of educational tests (cognitive, diagnostic, aptitude, or achievement tests), survey procedures, interview procedures or observation of public behavior:

Of human subjects that are elected or appointed public officials or candidates for public office; or

Conducted under a Federal statute requiring that the confidentiality of the personally identifiable information be maintained throughout the research and thereafter.

☐ 4. Research involving the collection or study of pre-existing data sets, documents, records, or specimens, but only if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, either directly or through identifiers linked to subjects [i.e. through use of a key]. If research team does not receive, view or handle identifiable *original source data* at any point, study may be "not human subject research" (see link above to determine).

Date range of pre-existing data:

Research involving one of more of these existing data sets may require you to obtain, prior to using and/or disclosing identifiable health information from the existing data set, either HIPAA research subject authorization integrated into the consent form (see "HIPAA" section of Medical Consent Guidance) or a waiver of a research subject authorization granted by the GWU IRB.

☐ 5. Research/demonstration projects conducted by other federal departments designed to study or evaluate public programs, procedures for obtaining benefits or services under those programs, possible changes or alternatives to those programs, or possible changes in methods or levels of payment for benefits under those programs.

☐ 6. Taste and food quality and evaluation / consumer acceptance studies, as long as safe, normal foods are being consumed, and federal guidelines regarding acceptable levels of agricultural chemical or environmental contaminants are adhered to.

RESEARCH SUMMARY *(Please see Exempt Instructions to ensure all required information is included in application)*

Research Purpose	The purpose of this study is to evaluate the influence of principal and school characteristics on principal ratings of teachers using the North Carolina teacher evaluation instrument. Studies have found varying strengths of correlation between principal ratings and value-added data, but have not deeply explored principal characteristics influencing their ratings of teachers. This study will examine the influence of principal characteristics (years of experience, beliefs about the malleability of others' ability, and dominant leadership orientation) and school characteristics (grade span and aggregate value-added data) on principal ratings of teachers on Standards 1 through 5 of the North Carolina teacher evaluation instrument.
Study Population	The participants in this study will be a stratified, proportional sample of North Carolina school principals. Both males and females will be included, and no one will be excluded based on ethnicity or race. Because this study is comparing principal ratings of teachers in elementary, middle, and high schools, principals of buildings that serve students from a combination of these grade spans will be excluded. School value-added data will be included as a criterion variable, so schools without value-added data will not be included. In North Carolina, value-

	added data is currently available for teachers of students in Grade 4 through high school. Finally, new principals who did not assign teachers the most recent summative ratings will not be included in the study. There are currently 2,489 schools in North Carolina; 2,469 of these schools have a principal identified in North Carolina's Educational Directory and Demographical Information Exchange. Among these schools, there are 1,112 elementary schools, 432 middle schools, and 418 high schools. A sample of 399 principals will be randomly selected..
Subject Recruitment Methods	I am using the Educational Directory and Demographical Information Exchange database of North Carolina principals to filter and sort schools by grade span. I will use stratified, proportional techniques to identify 399 principals meeting the parameters described in "Study Population." I will send an email, which is attached, providing a brief description of the study including purpose and risks. The first page of the online survey will be a "Welcome and Informed Consent" page describing the purpose, benefits, and risks of the study. Reminder emails will be sent at the end of Week 1 and Week 2. The text of these emails are attached.
Methodology *(Step-by-step 1,2,3 description of study design)*	The objective of this study is to better understand the influence of principal and school characteristics on principal ratings of teachers using the North Carolina teacher evaluation instrument. I am going to administer an online survey to principals who have assigned teachers their most recent ratings on summative evaluations. The survey will take approximately 15 minutes to take. Principal emails will be collected from the database available publicly at the Educational Directory and Demographical Information Exchange. Responses will be collected using a secure, web-based survey program. The survey will collect the school's six-digit code. The six-digit school code will be used to match the school's evaluation rating distributions available in the Educator Effectiveness Database on www.ncpublicschools.org to principal survey responses. The following data will be collected by matching the school code to publicly available data: Title I status, school size, percent of teacher who are fully licensed, percent of classes taught by highly qualified teachers, percent of teachers with advanced degrees, percent of teachers with less than 3 years' experience, and teacher turnover. Prior research has not addressed these school characteristics as influencing principal ratings of teachers, but the data will be collected to provide a deeper understanding of school context. Once the data is matched, the six-digit school code will be deleted. The principal's total years' experience as a principal, administrator, and teacher will be collected. The total years' experience as a principal will be used in the regression model. School value-added composite, grade span, principal race, and principal gender will be collected as well as principal responses to 32 frequency rating items from Bolman and Deal's (1990) Leadership Orientation Instrument–Self, Part I, and 3 Likert scale items from Dweck, Chiu, and Hong's (1995) Kind of Person Instrument. Consent will be obtained online. Principals will be given an overview of the study, the purpose of the study, the procedures of the study, the potential risks and benefits, contact information if they have further questions, and a statement of consent with the option to stop completion of the survey at any time. They will consent by proceeding to the next page of the survey. The text of the "Welcome and Informed Consent" page is attached. Data will be collected over a three week period. Data will be matched during that three week period but no later than one week after collection is complete. Information containing the school code will only be accessed by the student investigator/research coordinator. Data will not be reported at the individual level. The student investigator/research coordinator will use SPSS on a single private computer to analyze data using multiple regression techniques. Descriptive statistics and findings from a series of multiple regression tests will be reported. Raw data will not be reported but must be shared with the Leadership Orientation Instrument's developers if requested. No sensitive information will be shared in the data file. Only data related to the Leadership Orientation Instrument will be shared, and this will only be shared upon request. All data will be collected from an online survey instrument and from publicly available data. No Protected Health Information (PHI) will be collected.

Research Specific Risks	There are no physical risks associated with this study, but there is a minimal risk of loss of privacy to participants. These risks are precipated by the collection of the school code and the use of an online survey instrument. The specific procedures for maintaining confidentiality and privacy are outlined below in that section.
Benefits *(to subject and society)*	Prior evaluation research has not deeply addressed the relationship between principal characteristics and principal ratings of teachers. This study will benefit the subject and society by providing an analysis of the factors influencing principal ratings of teachers. The findings of this study could be useful to policy makers in amending evaluation processes, departments of education in developing professional development for principals related to teacher evaluation, and institutions of higher education in preparing future principals. If subjects choose to remain principals, they could potentially benefit from any increased understanding of the evaluation process.
Data Analysis and Justification of Sample Size	The sample size for this study was obtained using G*Power 3.7.1 for multiple regression. Significance criterion was set at $\alpha = .05$, the desired power for F test was set at .95, and a moderate effect size of .15 was selected. The number of test predictors was set at 5 and a priori type of power analysis was chosen. The G*Power calculator recommended a sample size of 138 with an actual power of .9508. A second calculation for sample size was run using A-Priori Sample Size Calculator for Multiple Regression, a product of Statistics Calculators. This program also recommended a sample size of 138. This study will use 34.6% as the estimated response rate based on a meta-analysis of response rates for 56 online surveys reported in 39 studies that yielded a mean response rate of 34.6% (SD = 15.7%) (Cook, Heath, & Thomson, 2000). Using 34.6% as the estimated response rate for this study and a desired response from at least 138 principals gave an initial sample size of 399 (138/34.6%). Because the sample will be a stratified, proportional sample with three strata, the proportion to be selected from each level was found by dividing the possible schools at each level by the total number of possible schools: elementary, 1,112/1,962=56.7%; middle. 432/1,962=22.0%; and high school, 418/1,962=21.3%. These percentages were multiplied by the desired sample size of 399 to find the number of schools to sample at each level: 226 elementary, 88 middle, and 85 high schools.
Confidentiality and Privacy *(Include plan for data storage, deindentification, and destruction)*	The online survey program uses Secure Sockets Layer and Transport Layer Security encryption, has firewalls in place, holds a transparent privacy policy, and limits employee access to data. Randomly selecting principals from across the state, rather than from a single region or district, will reduce the risks of violations of confidentiality or privacy. Deleting the school code when data is matched with publicly available school data will reduce the risk of loss of privacy. The matched data will be stored on the student investigator/research coordinator's computer and will be password protected. The matched data will not include any direct identifiers, codes, or links to the subject's identification. Only the principal investigator, methodologist, and student investigator/research coordinator will have access to the matched data unless requested by the instrument developer. The instrument developer would not receive data that includes any direct identifiers, codes, or links to the subject's identification. Data including direct identifiers, codes, or links to the subject's identification will not be saved.
Use of results/findings *(plan for dissemination of information)*	No individual results will be communicated from this data. Data will be presented at conferences and submitted to publications for dissemination, but all data will be aggregated/summarized for presentation.

INVESTIGATIVE TEAM SIGNATURES: My signature indicates that I will respect and protect the rights and welfare of individuals enrolled in this research project. I will also carry out my responsibilities as Principal Investigator as is outlined in Federal-wide Assurance of Protection for Human Subjects, for which GWU is registered with OHRP/DHHS, and as detailed in GWU HRPP policies & procedures. I will be guided by the principles contained in the Belmont Report and The Code of Federal Regulations governing research with human subjects (45 CFR 46). I have queried all members of the research team to determine if they have an economic interest in this study as defined by GWU policies. ***These signatures must be originals and are required for submission.***

Principal Investigator (Print/Type)	Signature	Date
Dr. Jennifer Clayton	Jennifer R Clayton	

Sub-Investigator (Print/Type)		
Sub-Investigator (Print/Type)		
Student Investigator/Research Coordinator (Print/Type) Stacy Leggett		
DEPARTMENT CHAIR/DEAN SIGNATURE: My signature indicates that this project has been reviewed by the appropriate departmental parties, who have judged that 1) there is a scholarly and a scientific justification for the protocol, that the study is feasible, and that the proposed methods are scientifically valid, 2) that the department has made the space and time commitment necessary to carry out the project, 3) that the financial implications of the research have been considered and deemed acceptable to the department and 4) that all ethical principles have been appropriately addressed.		
Medical, Alan G. Wasserman, MD or Gary Simon, MD, PhD **Non-Medical**, Name of Dept. Chair Sharon Dannels	Signature Sharon Anderson Dannels	
Department Affiliation/Campus Location Department Of Educational Leadership, Hampton Roads		
Phone 757-269-2213	Fax 757-269-4992	Email SDANNELS@GWU.EDU

Please submit to OHR, 2030 M St. NW Suite 301 with all materials identified in the IRB Submission Checklist

Appendix 2: Welcome and Informed Consent

Welcome and Informed Consent

THE INFLUENCE OF PRINCIPAL AND SCHOOL CHARACTERISTICS ON PRINCIPAL RATINGS OF TEACHERS USING THE NORTH CAROLINA EVALUATION INSTRUMENT

You are invited to participate in a research study under the direction of Dr. Jennifer Clayton of the Graduate School of Education & Human Development, The George Washington University (GWU). The principal contact for this study is Stacy Leggett, a doctoral candidate of GWU.

1. Why is this study being done?
You are being asked to participate in this study because you assigned teachers in your school their most recent summative ratings on the North Carolina evaluation instrument. Please read this form & ask me any questions that will help you decide if you want to be in the study. I can be reached at srleg@gwmail.gwu.edu. Taking part in the study is completely voluntary and even if you decide you want to participate, you can quit at any time. You must be at least 18 years old to take part in this study.

The purpose of this study is to understand the influence of principal and school characteristics on the teacher evaluation process. A total of 399 participants will be invited to participate in this online survey.

2. What is involved in this study?
If you choose to take part in this study, you will participate in one 15 minute online survey. The following activities are specifically research related: acknowledging consent and completing the online survey. The total amount of time you will spend in connection with this study is approximately 15 minutes. You will be asked questions related to your leadership orientation, beliefs about the malleability of others' abilities, your school value-added composite, school code, years' experience, and demographics.

3. What are the risks of participating in this study?
There are no foreseen risks in this study, however, you may stop your participation in this study at any time. Your employment status will not, in any way, be affected should you choose not to participate or if you decide to withdraw from the study at any time.

4. Are there benefits to taking part in this study?
You will not benefit directly from your participation in the study. The benefit to the field of educational leadership and teacher evaluation that might result from this study is a more accurate understanding of the factors influencing principal ratings of teachers.

5. Will I receive payment for being in this study?
You will not receive any type of payment for participating in this study.

6. How will my privacy be protected?
The results of this research study will only be reported in summary or aggregate form. Individual results will not be reported, and the people who participated in this study will not be named or identified. GWU will not release any information about your research involvement without your written permission, unless required by law.

You will be asked to provide your school code in order to match your responses to publicly available school data. Once your responses are matched to the publicly available data, your school code will be deleted and no one, including the research team, will be able to match your responses to you or your school.

7. Problems or questions
You may contact me at 252-370-7422 or srleg@gwmail.gwu.edu or Dr. Clayton at 757-269-2203 or claytonj@gwu.edu if you have any further questions. The Office of Human Research of George Washington University, at (202) 994-2715, can provide further information about your rights as a research participant. If you think you have been harmed in this study, you can report this to the Dr. Clayton, the Principal Investigator of this study. Your records for the study may be reviewed by departments of the University responsible for overseeing research safety and compliance.

By continuing with this survey, you certify that you have read and understood the informed consent form above and agree to participate.

Appendix 3: Survey

School Code and Value-Added Composite

1. What is your 6-digit school code?

Your school code will be used to match your responses to publicly available school data. Once your data is matched, your school code will be deleted to protect your privacy.

2. Based on EVAAS data, what was your school's value-added composite?

Leadership Orientation-Self (Bolman & Deal, 1990)

You are asked to indicate how often each of the items below is true of you in implementing the North Carolina teacher evaluation instrument.

Please use the following scale in answering each item.

1 Never
2 Occasionally
3 Sometimes
4 Often
5 Always

So, you would answer '1' for an item that is never true of you in carrying out the teacher evaluation process, '2' for one that is occasionally true of you in carrying out the teacher evaluation process, '3' for one that is sometimes true of you in carrying out the teacher evaluation process, and so on.

Be discriminating! Your results will be more helpful if you think about each item and distinguish the things that you really do all the time from the things that you do seldom or never.

3. Think very clearly and logically.

1 Never	2 Occasionally	3 Sometimes	4 Often	5 Always
○	○	○	○	○

4. Show high levels of support and concern for others.

1 Never	2 Occasionally	3 Sometimes	4 Often	5 Always
○	○	○	○	○

5. Have exceptional ability to mobilize people and resources to get things done.

1 Never	2 Occasionally	3 Sometimes	4 Often	5 Always
○	○	○	○	○

6. Inspire others to do their best.

1 Never	2 Occasionally	3 Sometimes	4 Often	5 Always
○	○	○	○	○

7. Strongly emphasize careful planning and clear time lines.

1 Never	2 Occasionally	3 Sometimes	4 Often	5 Always
○	○	○	○	○

8. Build trust through open and collaborative relationships.

1 Never	2 Occasionally	3 Sometimes	4 Often	5 Always
○	○	○	○	○

9. Am a very skillful and shrewd negotiator.

1 Never	2 Occasionally	3 Sometimes	4 Often	5 Always
○	○	○	○	○

10. Am highly charismatic.

1 Never	2 Occasionally	3 Sometimes	4 Often	5 Always
○	○	○	○	○

23. Set specific, measurable goals and hold people accountable for results.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

24. Listen well and am unusually receptive to other people's ideas and input.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

25. Am politically very sensitive and skillful.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

26. See beyond current realities to generate exciting new opportunities.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

27. Have extraordinary attention to detail.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

28. Give personal recognition for work well done.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

29. Develop alliances to build a strong base of support.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

30. Generate loyalty and enthusiasm.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

31. Strongly believe in clear structure and a chain of command.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

32. Am a highly participative manager.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

33. Succeed in the face of conflict and opposition.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

34. Serve as an influential model of organizational aspirations and values.

1 Never ○ 2 Occasionally ○ 3 Sometimes ○ 4 Often ○ 5 Always ○

Kind of Person Theory (Dweck, Chiu, & Hong, 1995)

Using the scale below, please indicate the extent to which you agree or disagree with each of the following statements.

38. The kind of person someone is something very basic about them and it can't be changed very much.

Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○

36. People can do things differently, but the important parts of who they are can't really be changed.

Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○

37. Everyone is a certain kind of person and there is not much that can be done to really change that.

Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree
○	○	○	○	○	○

Demographic and School Data

38. What is your school grade span?

- ◯ elementary school
- ◯ middle school
- ◯ high school

39. How many years' experience do you have as a principal?

☐

40. How many years' experience do you have as an educational administrator?

☐

41. How many years' experience did you have as a classroom teacher?

☐

42. What is your gender?

- ◯ Female
- ◯ Male

43. What is your race?

- ☐ White
- ☐ Black of African American
- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Native Hawaiian or other Pacific Islander

44. What is your ethnicity?

- ◯ Hispanic or Latino
- ◯ not Hispanic or Latino

Appendix 4: Initial Email to Principal Participants

Dear <insert name>:

I am a doctoral student at The George Washington University in the Educational Leadership and Policy Studies program working with Dr. Jennifer Clayton, a professor in the Graduate School of Education and Human Development. I am requesting your participation in the study *The Influence of Principal and School Characteristics on Principal Ratings of Teachers Using the North Carolina Teacher Evaluation Instrument*. Your contact information was obtained through the Educational Directory and Demographical Information Exchange (EDDIE). You were selected to participate in this study because you are the principal who assigned the teachers in your school their most recent summative evaluation ratings using the North Carolina teacher evaluation instrument.

Your participation in this study will assist us in gaining a better understanding of the evaluation process. The survey will take about 15 minutes to complete. Your responses will be confidential and will be shared publicly only in summary or aggregate form. No individual responses or names will be shared. All responses will be stored confidentially.

The first page of the survey is a Welcome and Informed Consent page that provides further information on this study. By proceeding to the second page, you provide your consent to participate in the study.

Finally, one of the questions will request your school's value-added composite. If you do not know your value-added composite, it can be located in EVAAS on your School Composite report located under the heading School Reports.

For your convenience, you may click the link below to go directly to the survey:
https://www.surveymonkey.com/s/QHSYM57

The survey must be completed by [insert due date here].

You may contact me at 252-370-7422 or srleg@gwmail.gwu.edu or Dr. Clayton at 757-269-2203 or claytonj@gwu.edu if you have any further questions.

Thank you for your assistance.

Stacy Leggett
srleg@gwmail.gwu.edu
252-370-7422

Appendix 5: Second Email to Principal Participants

Dear <insert name>:

I am a doctoral student at The George Washington University in the Educational Leadership and Policy Studies program working with Dr. Jennifer Clayton, a professor in the Graduate School of Education and Human Development. I am requesting your participation in the study The Influence of Principal and School Characteristics on Principal Ratings of Teachers Using the North Carolina Teacher Evaluation Instrument. Your contact information was obtained through the Educational Directory and Demographical Information Exchange (EDDIE). You were selected to participate in this study because you are the principal who assigned the teachers in your school their most recent summative evaluation ratings using the North Carolina teacher evaluation instrument.

Your participation in this study will assist us in gaining a better understanding of the evaluation process. The survey will take about 15 minutes to complete. Your responses will be confidential and will be shared publicly only in summary or aggregate form. No individual responses or names will be shared. All responses will be stored confidentially.

The first page of the survey is a Welcome and Informed Consent page that provides further information on this study. By proceeding to the second page, you provide your consent to participate in the study.

Finally, one of the questions will request your school's value-added composite. If you do not know your value-added composite, it can be located in EVAAS on your School Composite report located under the heading School Reports.

For your convenience, you may click the link below to go directly to the survey:

You may contact me at 252-370-7422 or srleg@gwmail.gwu.edu or Dr. Clayton at 757-269-2203 or claytonj@gwu.edu if you have any further questions.

Thank you for your assistance.

Stacy Leggett
srleg@gwmail.gwu.edu
252-370-7422

Appendix 6: Third Email to Principal Participants

Dear <insert name>:

This is the final reminder that the survey for the study *The Influence of Principal and School Characteristics on Principal ratings of Teachers Using the North Carolina Teacher Evaluation Instrument* is due on [insert due date here]. If you have completed this survey, thank you for your participation. If you have not, the survey will take about 15 minutes to complete and your participation in this study will assist us in gaining a better understanding of the North Carolina teacher evaluation process.

One of the questions will request your school's value-added composite. If you do not know your value-added composite, it can be located in EVAAS on your School Composite report located under the heading School Reports.

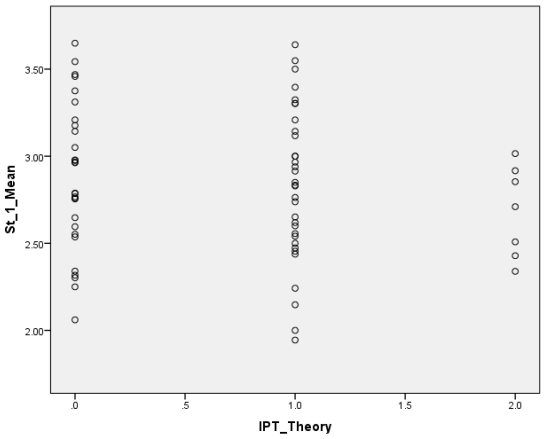
For your convenience, you may click the link below to go directly to the survey:
https://www.surveymonkey.com/s/QHSYM57

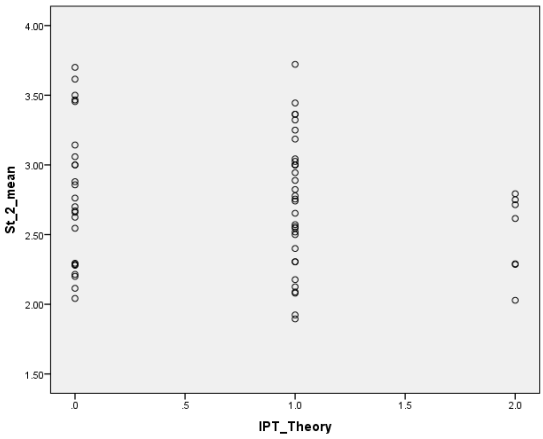
You may contact me at 252-370-7422 or srleg@gwmail.gwu.edu or Dr. Clayton at 757-269-2203 or claytonj@gwu.edu if you have any further questions.

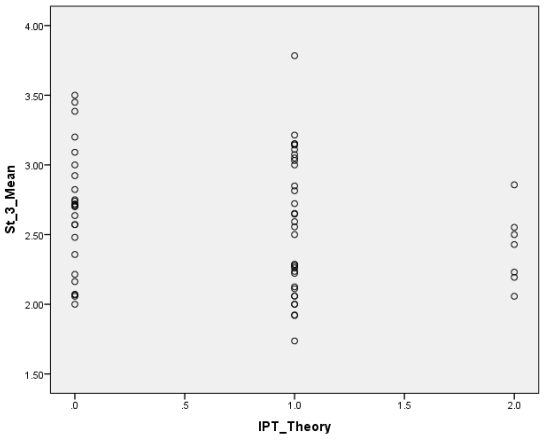
Thank you for your assistance.

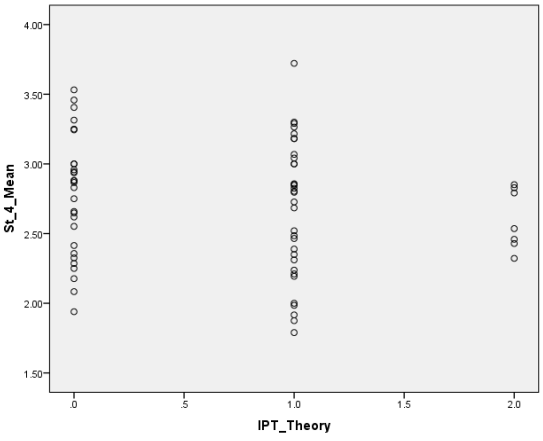
Stacy Leggett
srleg@gwmail.gwu.edu
252-370-7422

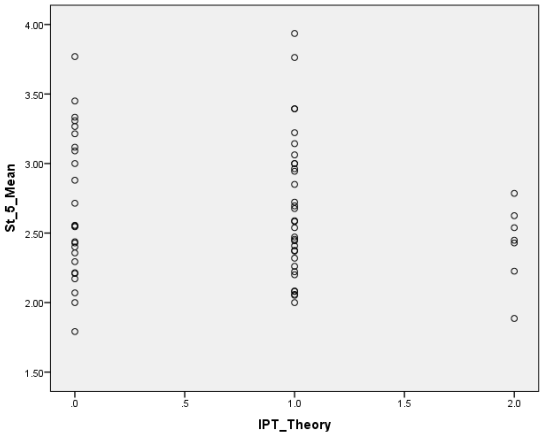
Appendix 7: Scatterplots Graphing Principal Mean Ratings of Teachers against Principal Implicit Person Theory



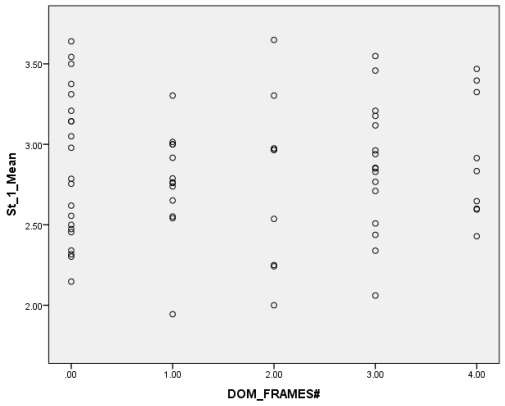


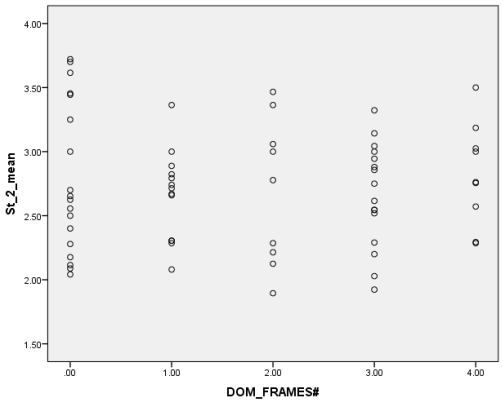


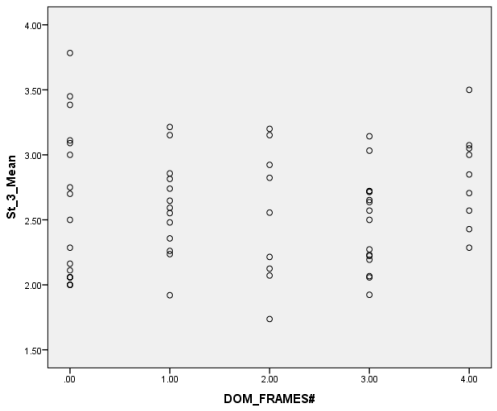


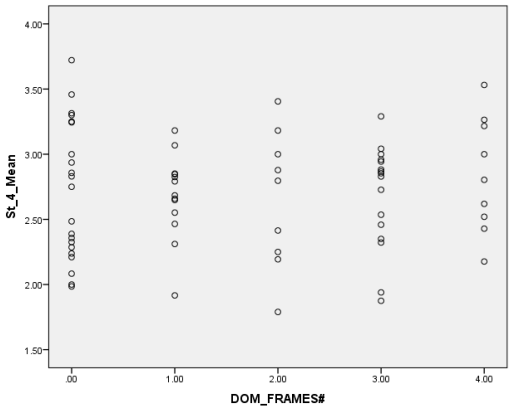


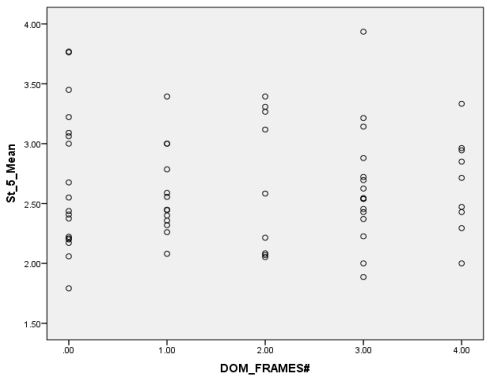
Appendix 8: Scatterplots Graphing Principal Mean Ratings of Teachers against Principal Number of Dominant Leadership Frames



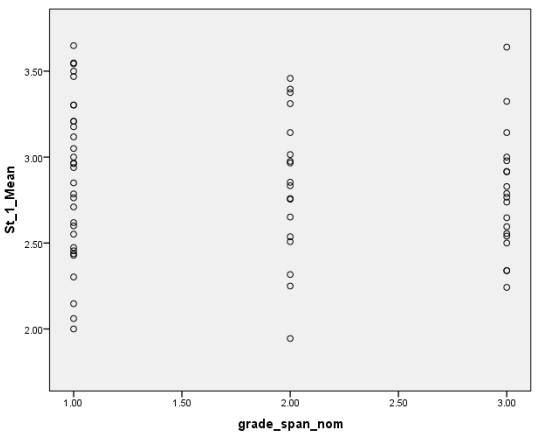


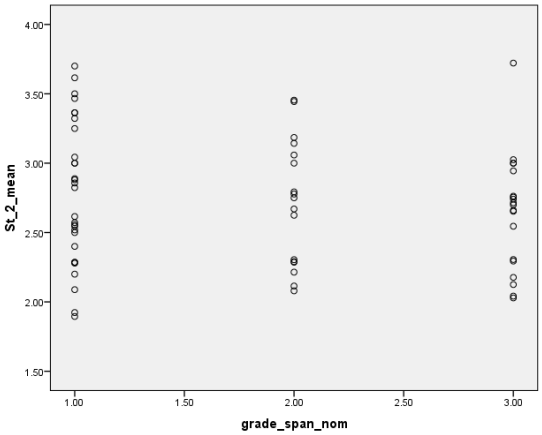


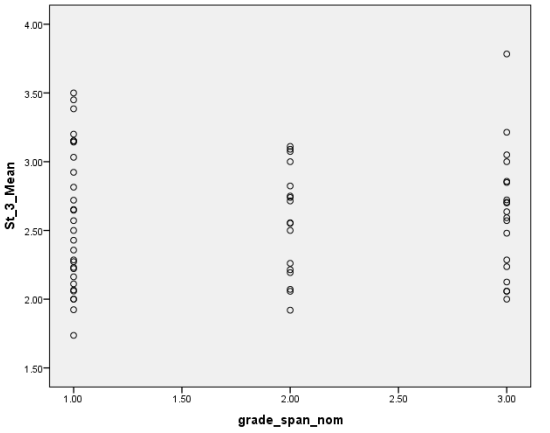


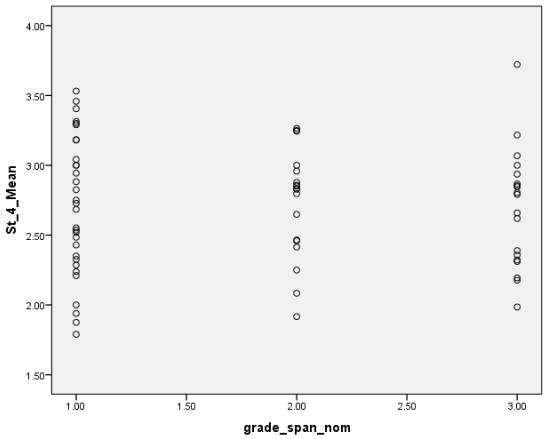


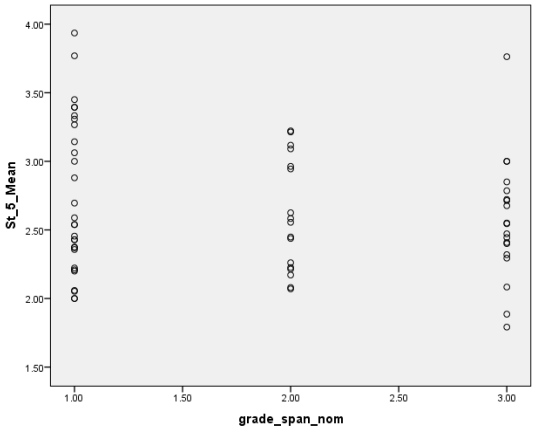
Appendix 9: Scatterplots Graphing Principal Mean Ratings against School Grade Span











Note: School grade span was treated as nominal variable with "elementary school" = 1, "middle school" = 2, and "high school" = 3.

Appendix 10: Scatterplots Graphing Principal Mean Ratings against Title I Status