

Characteristics of High Quality Teaching	16
Clinical Supervision	17
Developing Quality Teachers	19
Linking Teacher Quality and Student Achievement	21
Governmental Influence.....	22
Virginia’s Uniform Performance Standards (2011)	25
The Challenge	32
Virginia’s Performance Standards	33
Performance Standard 1: Professional Knowledge (10%).....	34
Performance Standard 2: Instructional Planning (10%).....	38
Performance Standard 3: Instructional Delivery (10%).....	40
Performance Standard 4: Assessment of and for Student Learning....	44
Performance Standard 5: Learning Environment	49
Performance Standard 6: Professionalism	53
Performance Standard 7: Student Academic Progress	56
Summary of the Research Base	59
Concluding Thoughts	60
Chapter 3: Methodology.....	62
Research Questions.....	62
The Approach of Phenomenology	62

Four Defining Characteristics of Phenomenology	64
Context of the Study	65
The School	66
The Students	67
The Teachers	71
The Data	72
Phases of the Study	73
Phase 1: Recruitment	73
Phase 2: Interviews	76
Phase 3: Transcriptions	80
Phase 4: Coding	82
Phase 5: Analysis	84
Trustworthiness.....	84
Researcher Stance	88
Chapter 4: Findings.....	91
Research Question 1.....	91
Limited Understandings.....	91
Categories and Ratings.....	94
Summary for Research Question 1.....	95
Research Question 2.....	95

Mistrust.....	95
Student Classroom Placements.....	111
Manipulated Scores.....	115
Communication Breakdowns.....	121
Emotions.....	131
Summary for Research Question 2.....	137
Research Question 3.....	138
“It Doesn’t”.....	138
Useless.....	139
Blame Placing.....	143
The Evaluator.....	144
The Evaluation System.....	147
The School Division.....	148
Summary for Research Question 3.....	149
Chapter 5: Discussion and Conclusions.....	150
Introduction.....	150
Summary of Findings.....	150
Conclusions.....	151
Inadequate Research Base.....	152
Problematic Implementation.....	152

Misalignment for Specialists.....	153
Critique of the Study.....	154
Recruitment.....	154
Interview Protocol/Instrumentation.....	155
Overall Methodology.....	155
Implications of the Study.....	156
Implications of the Research Base.....	156
Implications for System Implementation.....	156
Implications for Specialists.....	158
Implications for the Evaluator.....	159
Epilogue.....	160
References.....	164
Appendices.....	185
Appendix A: Internal Review Board Approval.....	185
Appendix B: Interview Protocol.....	187

List of Figures

Figure 1. Word Cloud of Emotional Words Used by Participants.....137

List of Tables

Table 1. <i>A summary of selected features of positive learning environment for standard 5: learning environment.....</i>	53
Table 2. <i>Statistics for bells elementary school, schools in virginia, urban schools in virginia, and schools in virginia.....</i>	68
Table 3. <i>Participant demographics.....</i>	76
Table 4. <i>Ways interview questions aligned with research questions.....</i>	79
Table 5. <i>Participant statistics.....</i>	82

Chapter 1

Introduction

Learners deserve effective teachers. What student ever posed a question who did not deserve a response that inspired greatness? Teachers matter. And if teaching and learning matter it is critical that effective teachers be identified and supported (VDOE, Guidelines, 2011). If we are to evaluate teachers, then we must also identify evaluation systems strong enough to distinguish the practice of high quality teachers (Stronge, 2006). Schools are only as effective as the teachers, and one of the primary ways of ascertaining whether or not an effective teaching force exists is through evaluation.

Teaching is a complex job and the evaluation of such a complex job is only possible when the evaluation system takes into consideration the complexities of the job. Learning to become an effective teacher takes time and effort, and includes receiving support from other effective teachers whose guidance can help novice teachers learn about the complexities of meeting individual needs and improve their practices (Tucker, P. D., Stronge, J. H., & Gareis, C. R., 2002). Because a high quality teacher is critical to a high quality education, improving the evaluation of teachers is necessary to ensure that students have an effective teacher. The benefits of such an evaluation system are widely known (Johnson, 1999).

Rose's Vignette

“Teacher of the Year” is a title given to highly effective teachers. Rose (pseudonym) believed she had been recognized for her dedication and ability to inspire learning in her students. Through the use of songs, chants, and dance routines, her fifth graders made remarkable strides as evidenced by their annual assessments. For 20 years, Rose was validated for her teaching expertise through the evaluation process and was consistently asked to mentor student teachers, lead professional development activities within her division, and share her “secrets” at numerous educational conferences. Rose had every right to believe that her passion perfectly aligned with her career path. She was productive and believed that what she did every day in the classroom mattered in the lives of children. Rose believed she was an effective teacher.

After two decades in the classroom doing what she loved, her peers and her evaluator named Rose “Teacher of the Year,” a strong validation that she was an effective teacher. The following school year, Rose’s site administrator implemented the division’s new teacher evaluation system designed to meet the requirements of the federally mandated Race to the Top initiative (USDE, 2009). Rose had been through 20 annual evaluations as an elementary teacher. She saw no reason to worry about another new evaluation system with an already familiar site administrator. She was confident that her abilities and her work ethic would once

again yield superior scores based on her performance, as they had done so many times before.

But she was wrong. In fact, not only did she fail to receive superior scores, for the first time ever Rose was rated “inadequate” in multiple categories. She was devastated. The same evaluator who only a few months before had awarded her “Teacher of the Year” now deemed her a failure in several areas in which she had previously excelled. Rose’s shock could not be contained. When she asked for an explanation, the administrator said that the school division was in the midst of evolving toward a new student growth model and new baselines were being set for everyone, including veteran teachers (“Rose,” personal communication, March 5, 2015).

The fact that a new growth model based on student achievement was being set was irrelevant to Rose, who felt crushed to the core. She looked at the results of her latest evaluation with disbelief. She took it personally. The site administrator dismissed her feelings of inadequacy and continued to champion the division’s newly implemented system. Rose announced her resignation immediately, to the horror of her principal who offered to rectify the situation by adjusting some numbers. Rose was indignant at this offer and threatened to expose any manipulation of the evaluation scores simply to appease her or to diffuse the

situation. She turned in her written resignation the following morning. Rose ended her teaching career 30 days later. She now works in another field.

A system that once held her in high esteem in effect betrayed her as the result of a newly implemented evaluation system. Not only did Rose lose the career she loved, but the field of education lost something even greater: an experienced, effective teacher. The real losers were the children who would never experience her professional, loving touch. For 20 years, Rose labored with thousands of children from every background. She gave of her time and her talents to try to give each child what they deserved in a teacher. She spent her own money to get what they needed if the school division did not have the funds to do so. Rose's identity was that of a teacher. It defined her and gave her a purpose. The evaluation process caused Rose inner conflict which could not be denied. Within a matter of minutes, she knew it was a conflict with which she could not live.

It is sometimes said that "beauty is in the eye of the beholder." Discovering the most beautiful piece of artwork or the most riveting storyline can be highly subjective, dependent upon who is the judge. The evaluation of a person and their performance can be even more highly subjective. During the last few decades in the United States, evaluations for public school teachers have undergone many changes. The thought of evaluation itself may cause a person anxiety and worry, even if the person perceives that he/she is doing a good job. Yet some evaluation

practices have made teachers to feel that the evaluation process was in stark contrast to their performance in the classroom (Kane, T. J., Kerr, K. A., Pianta, R. C., & Measures of Effective Teaching Project, 2014.; Stiggins & Duke, 1988).

Teachers have many obligations throughout the day and they must constantly seek a balance. The site administrators, the stakeholders, the parents, and the students all have expectations about how teachers should do their jobs. Rose may have thought she understood both the expectations and the tenets of the evaluation system under which she worked but apparently, she did not.

Teachers have seen the evaluation process continue in a constant flux over many decades as both state and federal agencies have continued to try and develop better ways of evaluating the work of a highly effective teacher (Darling-Hammond, 2015; Darling-Hammond, 1997). Veteran teachers have watched system after system be introduced and they have seen those same systems abandoned for yet another new process (Seldin, 1999). There exists the possibility within evaluation that negative perceptions may lead to feelings of inner conflict, just as positive perceptions may lead to feelings of inner satisfaction. However, highly effective systems of evaluation have been shown to positively affect student achievement when used to support teachers and their practices (Kelly, 2012; Taylor & Tyler, 2011; Tucker & Stronge, 2005).

High Quality Teachers

No single characteristic exemplifies a highly qualified teacher, though efforts have been made to describe what highly qualified teachers do. The term highly qualified teachers (HQT) comes from the original language of the No Child Left Behind Act (NCLB, 2001). This piece of legislation designated funds to educational agencies for the purpose of improving the student achievement through the professional development of highly qualified teachers and principals. Though we may be able to formulate dispositions of teachers who are effective, the truth is that no single list of characteristics will work for all teachers. Each teacher is unique and each situation differs according to its clientele. Teaching is a multi-faceted and highly complex process (Pogodzinski, Umpstead, & Witt, 2015) and schools are complex environments. Successful teaching does not result in a standardized product and cannot be reduced to a single technique in order to arrive at a desired result (Darling-Hammond, 2015).

Teaching is rational work because it involves reasonable and sensible work with people; therefore, it involves relationships. Teaching is not work in which knowledge may be simply imparted in a neatly tied package. Teaching and learning are cooperative efforts between teachers and students. It is important for the teacher and the student to work together to understand the material and deal with confusion and misunderstandings as they arise.

Further, teaching involves context-specific tasks (Reynolds, 2011; Crump, Vaquero, & Milliken, 2008) in content-specific environments. It is founded upon the teacher's content knowledge and the students' developmental abilities. Because student achievement is the goal, the methods whereby teachers make knowledge available are very important (Marzano, 2004). Since this is the case, the evaluation of teachers involves the overarching goal of improving teaching practices.

Teacher Evaluation

Teacher evaluation is receiving attention in large part because evaluation systems are a requirement for states and divisions who receive funding under the federal *Race to the Top* initiative (USDE, 2009), or who were awarded flexibility waivers under the *No Child Left Behind Act of 2001* (NCLB, 2001). Teacher evaluation has been promoted as being one of the primary tools used to improve educational policies. Because Virginia accepted ARRA funds, the Virginia Department of Education had to adopt a new system that aligned with federal expectations. The new *Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers* (VDOE, 2011) effective July 1, 2011, were designed in an effort to bring clarity and specificity to evaluation of the commonwealth's teaching force and elevate student academic performance to the primary focus.

The United States is in a critical period in dealing with teacher evaluations because in almost every division in every state across the country, evaluation processes are undergoing changes, some quite radical (Kane, Kerr, Pianta, & Measures of Effective Teaching Project, 2014). It is imperative that our leadership focus on improving the quality of teaching and learning in order to avoid integrating changes that may possibly impacting students' learning (Green, 2011). History has shown that in the rush to make changes sometimes systems have been employed prior to the establishment of solid research-based foundations. Virginia's current evaluation system lacks clear ways of distinguishing between those teachers who are struggling and those who are exceptional (VDOE, 2011). Without a structurally sound, well thought-out evaluation system, as well as those who knew how to use it properly, the teacher evaluation process can contribute little to the improvement of teacher practices (Kane, et al, 2014).

Personal Significance

School was valued in my home and I was successful at all areas of school. My mother taught third grade. She had a master's degree in Home Economics which qualified her under the 1960s certification system. Everything about school motivated me to do whatever I could to receive the approval of my teachers. School expectations for me were always made clear and most of the time I fulfilled those expectations.

Originally, I did not want to become a teacher. With such strong family ties to the field of education, it is perplexing as to why I did not consider becoming a teacher sooner. I changed my college major from pre-law to elementary education and became a teacher. My first teaching job was in Tanzania, East Africa on a mission. Fourteen students were enrolled in nine different grade levels from the countries of Tanzania, Uganda, and the United States, including my own two boys. A school was constructed under my supervision for the purposes of educating these children and I became the Headmistress. I planned the schedule, designed the curriculum for nine grade levels, and organized the instruction to include many of the parents. I was responsible for judging the effectiveness of the school program. No one evaluated me and the job I had done.

After returning home, I secured a teaching position in a local elementary school. This school offered both traditional and multi-age classes. The first year I was assigned to a 3rd/4th grade self-contained class with 24 students. I thrived in an environment where my principal encouraged my growth celebrated my individual creativity. The following year I was moved to a multi-age 5th/6th grade class with a team teacher. It seemed only natural to pursue a master's degree in elementary education the summer after my first full year in the classroom.

Genesis of Research Ideas

Career moves ultimately brought me to Virginia where I taught fourth grade Virginia History, fourth and fifth grade language arts, and fifth grade mathematics. In my 23rd year, I found myself in a Virginia middle school, where I taught 6th and 7th grade mathematics. I felt as if I was a good teacher and I rarely thought about the evaluation process because I had experienced only success. Over the years, I had been nominated for several awards, such as Teacher of the Year in my school division.

In the 20th year of my teaching career, significant changes occurred with the evaluation system and new administration. It seemed to me that my professional knowledge was no longer respected by my primary administrator. Because of the mandates from the American Recovery and Reinvestment Act of 2009 (ARRA), student academic progress became 40% of my teacher evaluation criteria and my evaluator determined that test scores would serve as the only measure of progress; therefore data collection became a priority for school divisions. The changes in the evaluation process were not obvious at first, even with my place on the division-wide evaluation committee.

The evaluation system would change the way I felt about myself as a teacher. My teaching strategies were no longer validated after the new mandates took effect later that year. It was obvious that priorities had shifted to student

outcomes based on data from one source. Observations were inconsistent and the good comments I often received on formative evaluations disappeared. The new evaluation system did not highlight many of the characteristics I valued in good teaching practices. I tried to align my teaching strategies with the new evaluation system so I would be successful. Binders and binders of student work were collected for use with the evaluation system. These examples of student work were supposed to show student growth throughout the year. However, over the course of two decades I knew when students were progressing and when they were being successful. So if my evaluator needed to see work samples, he saw work samples, but I'm not sure the evaluator was able to rate my teaching by looking at student work for 90 seconds. It seemed like a waste of my time.

Data collection was made a priority by the administration. Observations began to bring worry and angst. Many of the changes affected my self-esteem and my identity as a teacher. Songs, dances, and art projects that were aligned with the SOLs for the purpose of reaching struggling students were frowned upon. I was no longer happy. Armed with questions concerning the newly implemented teacher evaluation system, I looked to my administrator for answers, only to discover that administrators were still in training and there were questions he/she could not answer. None of the past years' successes seemed to matter. Just like Rose, I left

the school I loved not long after. The pursuit of my doctoral degree led me to research in the area of teacher evaluation.

Study Rationale and Purpose

Because of my personal experiences with teacher evaluation across nine grade levels and 23 years, my personal bias toward teachers was guarded as I researched this topic. The purpose of this study was to explore what teachers understood and perceived about the teacher evaluation system at one elementary school in Virginia. In this critical examination of teachers' understandings and perceptions, some of the challenges with which teachers have been faced in regards to professional growth and self-efficacy were illuminated as teachers shared their descriptions of work and evaluation.

The teacher evaluation process is but one of many directives that have trickled down from federal recommendations to state policies to division regulations. Teacher evaluations are enacted in every school every year for all teachers (VDOE, 2011). It should not be forgotten that teacher evaluation happens in the lives of real people. Teachers put their energies, hopes, and dreams on the line daily with the purpose of enriching the lives of children (Quinn, 2014). The fact that teachers are indeed human will mean that they will vary in their abilities to facilitate student academic progress consistently from year to year and from classroom to classroom. Understanding the ways in which the evaluation process

has impacted teachers might lead us to a greater understanding of the ways in which it has impacted student progress, as well. Because of this possible connection, there was a need for research that examined the teacher evaluation system in Virginia. Further, there was also a need for the type of research that sought to understand perceptions about the evaluation process from those who experienced it firsthand - the teachers. This research gave these teachers a voice. Unfortunately, Rose is no longer teaching, but she can still have a voice. Listening to the voices of teachers sharing their perceptions may lead to significant changes in the evaluation process.

Description of the Study

The study sought answers to the research questions by interviewing elementary school teachers in one school in Virginia. The population was intentionally chosen because these teachers were evaluated by the same administrator who used the current evaluation system. Since the same administrator used the same evaluation system with all of the participants, there was more consistency and therefore, the data were more reliable. Furthermore, this particular school was not necessarily representative of schools found in Southwest Virginia, but it was representative of urban schools in Virginia, based on statistics gathered from online school report cards. For comparison, 10 elementary schools from 10 school divisions across the state of Virginia were chosen at random. These

10 elementary schools were situated in urban areas and were statistically very similar to BES. Statistics taken from the school report cards online were used to find averages and these averages were used as the “urban Virginia” statistics.

Moreover, the evaluator in this case was very interested in what the teachers understood and perceived with regards to the evaluation system and hoped to make use of the findings in order to better assist teachers and administrators in navigating the evaluation system and to strengthen their professional development in the future. Both the evaluator, as well as the appropriate staff members in the central office gave their approval for this study. All of the necessary documents were included with the initial request and I was told to proceed.

Research Questions

The study was guided by the following research questions:

- 1) What are teacher understandings of the teacher evaluation process?
- 2) What are teacher perceptions about their experiences with the teacher evaluation process?
- 3) In what ways do teachers believe that evaluation informs their practices?

Chapter 2

Review of Literature

Teacher supervision and evaluation in public schools in the United States has a long history as a field of study and this documented history contains many trends and changes (Glanz, 1994). Slow implementation and abrupt variations have made it difficult to create a definitive timeline of the history of evaluation because every state's implementation differs across contexts.

History of Teacher Evaluation

The earliest documented days of teacher evaluation focused on whether or not the teacher educated the students well enough to read the Bible. The foundation of the Old Deluder Law that was passed in Massachusetts in 1647 (Olivia & Pawlas, 2004) was that an educated citizen would be able to understand and follow the moral principles addressed in the Bible. Both clergy and lay people accepted the responsibility of inspecting schools. The purpose was to ensure that teachers were maintaining the strict moral and religious values of the community (Glanz & Sullivan, 2005). As the number of schools throughout the country began to flourish in the 1800s, responsibility for teacher evaluation shifted to the newly created positions of principals and headmasters (Olivia & Pawlas, 2004).

The “Science” of Evaluation

Max Weber, noted economist and sociologist, developed some characteristics of a bureaucratic organization in the early 1900s (Olivia & Pawlas, 2004). His organizational structure centered on a top-down management approach with strict rules and procedures. During the same time period, Frederick Taylor wrote the *Principles of Scientific Management* (Glanz, 1991). These principles were influential because they focused on the responsibilities of the manager in order to improve production and efficiency. It is very unlikely that either Weber or Taylor realized how influential their ideas would be on the development and utilization of evaluation systems for teachers. At the same time, the mainstream thought was that if one could simply study what a successful teacher did to get results, a simple checklist could be developed so that others could follow suit (Glanz & Sullivan, 2005).

Characteristics of Highly Qualified Teachers

John Dewey and other leaders and reformers in the 1920s began to recognize the importance of successfully including the teachers in their own evaluations. He and others believed that evaluation should not be something that was “done to them” (Glanz & Sullivan, 2005). Following this idea, more democratic-like suggestions for supervision began to emerge. Teacher evaluation processes began to focus on professional development as opposed to simply finding a way to

remove poor teachers from the system. Cooperative problem solving techniques were used as a part of professional development during this time (Pajak, 2000). However, Ellet and Teddlie (2003) reported that the evaluation of teachers was still more concerned with personal characteristics of the teachers than their ability to teach.

During the 1940s and 1950s, researchers began in-depth studies of those characteristics of successful teachers. Danielson and McGreal (2000) explained: “Educators of this era believed that teachers who possessed these traits were more likely to perform effectively, so these traits became the centerpiece items in the local teacher evaluation criteria” (p. 13). In the 1960s, Ellet and Teddlie (2003) reported that educational researchers began to focus on effective teaching practices that could possibly lead to a variety of student outcomes, thus laying the foundation for the work in “clinical supervision,” a term borrowed from the medical profession.

Clinical Supervision

In the 1970s, the work of two educational researchers, Goldhammer (1969) and Cogan (1972) became synonymous with teacher evaluation and supervision. They developed the model we now refer to as clinical supervision. This model was designed to encourage teachers and supervisors to work together in order to improve teaching and learning. Following on the heels of their work, Hunter

(1982) developed her clinical supervision model, although it was considered to be less collegial and formative than that of either Goldhammer or Cogan. Hunter enumerated seven essential elements of classroom instruction in her lesson design model and posited that when these elements were present and successfully implemented by the teacher, student learning could be maximized. School divisions and states across the nation began to adopt new teacher evaluation models derived from the research in clinical supervision and, in many cases, some form of these new models followed Hunter's lesson design. "Unfortunately, the adoption of these clinical supervision models promoted a very simplistic and summative orientation of evaluation" (Danielson & McGreal, 2000, p. 14).

In the 1990s, the clinical supervision model was still quite popular. However, many schools began to look at ways of collaborating with teachers in the development of different methods of evaluating its teachers. For instance, a) peer supervision, b) differentiated supervision, and c) action research were identified as alternative models that might become more successful evaluation techniques (Danielson & McGreal, 2000). With the 1996 publication of *Enhancing Professional Practice: A Framework for Teaching*, Charlotte Danielson became known for her framework focused on teacher performance domains, which also provided rich descriptions of what high quality teachers should be able to do. She included areas such as planning and preparation, the classroom environment,

instruction, and professionalism. Schools rallied around the exploration of this model as they had done previously with that of Hunter. Divisions and states went back to the drawing board with their teacher evaluation systems and tried to situate them around this new framework.

Developing Quality Teachers

Another publication in 1996 highlighted the importance of teacher quality in school improvement. *What Matters Most: Teaching for America's Future* issued a strong statement when it reported that students were unable to meet high standards without effective teachers (National Commission on Teaching and America's Future, 1996). The commission's conclusion was that teachers do make a difference, and if they do not, there should be accountability. Darling-Hammond (2000) conducted a study of policies for teacher education, licensing, and professional development that encompassed all 50 states and her findings indicated that there was, indeed a relationship between teacher quality and student achievement. Her focus on teacher quality provided the catalyst for current perceptions of how to design relevant and effective teacher evaluation systems.

Teacher evaluation systems continued to evolve, but were far from being stable. In the late 1990s, the Nevada school division of Washoe County developed a teacher evaluation pilot program because the complaint of educators was that the former evaluation system did little for teachers' professional development and was

believed to be something that was “done to them” rather than “done with them” (Sawyer, 2001). Teachers have little to do with the changes made to policies and procedures that affect them. Laine, Behrstock-Sherratt, and Lasagna (2011; 2010;) found that more than 70% of teachers say they have very little control over what happens in schools and are left out of any decision making within their own divisions. At the same time, 85% of teachers report that their principal is doing an “excellent” or a “pretty good job” (Markow, Macia, & Lee, 2013). One conclusion that might be drawn from such data is that teachers believe their views will be ignored even when good principals lead their schools.

That kind of thinking in school systems and policies has led teachers to feel isolated, frustrated, and undervalued (Lortie, 2002; Richardson & Placier, 2001). It was suggested by Callahan (1962) that our K-12 system was designed to isolate teachers and classrooms in order to promote efficiency. In the very first edition of his classic book *Schoolteacher*, Lortie (2002) observed over 40 years ago that practitioners historically have had little to say about schooling. Lortie wrote, “Teachers never did gain control of any area of practice where they were clearly in charge and most expert.” Instead, he observed, schooling has “been dominated by persons in other roles” (p. 12). Federal regulations and national control over policy making have caused teachers to feel left out of important decisions.

Linking Teacher Quality and Student Achievement

In the past several decades, studies continued to confirm that teacher quality was the most important indicator of student success and achievement (Darling-Hammond, 2000; Haycock, 2000; Kaplan & Owings, 2002; Sanders & Horn, 1994; Wright, Horn, & Sanders, 1997). Sanders and Rivers (1996) studied students with “effective” teachers and compared student achievement over a three-year period with students who did not have teachers classified as “effective.” Their findings indicated that even though students might have started at the same place, students who worked with more effective teachers showed more academic growth than their counterparts. Sanders and Rivers, however, never clearly defined what constituted an “effective” teacher.

A number of studies have found that the National Board of Professional Teaching Standards assessment process distinguishes teachers who are more effective in improving student achievement from others who do not achieve certification (Smith, Gordon, Colby & Wang, 2005; Cavalluzzo, 2004; Goldhaber & Anthony, 2005; Vandervoort, Amrein-Beardsley, & Berliner, 2004). According to this research, the frequent, skilled use of standards-based observation coupled with feedback for the teacher is significantly related to student achievement gains, as the process aids teachers in improving practice and effectiveness (Milanowski & Kimball, 2003). However, Wilcox (2000) found no substantive evidence that the

National Board for Professional Teaching Standards (2002) certification increased student achievement. Studies continue to confirm that teachers who are more highly qualified tend to produce students with greater achievement gains.

Governmental Influence

The educational influence of the federal government is not explicitly stated in the United States Constitution (Hamilton et al., 1818). Even so, the United States government has passed numerous pieces of legislation over the past 50 years that have significantly affected education in America. Most of this legislation focused on the link between teacher quality and student achievement in some form. Under President Lyndon B. Johnson, the Commissioner of Education designed the Elementary and Secondary Education Act (ESEA) of 1965 (U.S., 1965). This landmark piece of legislation was passed during a period that came to be known as the “War on Poverty.” The ESEA did not specifically address the correlation between teacher quality and student achievement, but instead centered on funding an educational system to meet the needs of all of the nation’s students.

During his tenure as President, George W. Bush proposed new legislation aimed at strengthening our nation’s schools. *No Child Left Behind* (2001) was a bipartisan bill passed by both houses of Congress and signed into law on January 8, 2002. The *No Child Left Behind Act* (2001) was a reauthorization of the ESEA (1965) whose purpose was “to close the achievement gap with accountability,

flexibility, and choice, so that no child is left behind.” NCLB linked teacher quality with increased student achievement when it specified that all students were to be taught by a “highly qualified” teacher defined with very specific indicators.

According to the United States Department of Education website (2007), the four pillars of NCLB were “stronger accountability for results, more freedom for states and communities, proven education methods, and more choices for parents.” This law required that public schools that received federal funding to administer an annual statewide standardized test to all students. Schools that received Title I funding through the ESEA (1965) were required to make Adequate Yearly Progress (AYP) in test scores each year. NCLB stated that “Every child deserves highly qualified teachers,” and it required states to “have a highly qualified teacher in every public school classroom by the end of the 2005-2006 school year”

(http://www.whitehouse.gov/infocus/education/teachers/quality_teachers.html).

Almost immediately teaching became highly politicized. Teachers in all divisions became accountable for student achievement as measured by high-stakes testing. States scrambled to either develop or revise academic standards in order to meet the rigorous federal requirements of NCLB and some even modified their curricula to ensure that all standards were being taught to all students. Along with rewriting standards and curricula came revising teacher evaluation systems for the purpose of ensuring that high quality teachers were educating students. By 2015,

criticism from all sides had accumulated so much that a bipartisan Congress stripped away all of the national features of NCLB (2001). Its replacement, the Every Student Succeeds Act (ESSA) turned the remnants over to the states.

One of the goals of the United States Department of Education's 2007 Strategic Plan was to improve student achievement. Objective three of this particular goal addressed the issue of "improving teacher quality." It states that "high-quality, effective teaching is one of the 27 most important contributors to improving student achievement" (U.S. Department of Education, 2007, p. 10). Schools across the nation were required to provide a "highly qualified" teacher in all core academic subject areas. A number of studies linked student achievement with teacher certification (Goldhaber & Brewer, 2000; Hawk, Coble, & Swanson, 1985; Walsh, 2001). Students taught by teachers who were certified in their content area performed better on measures of achievement than students who were taught by teachers working outside their content area (Fetler, 1999). Fetler found that teachers who had high scores on a math certification exam also had students with higher scores on math exams.

A pivotal occurrence fundamentally changed the way teachers are evaluated. Arne Duncan, Commissioner of Education (2009-2015) during the Obama Administration, designed another piece of key legislation for our nation's schools. Between 2009 and 2012, Congress adopted the legislation and appropriated

approximately \$5.05 billion for *The Race to the Top* (RTT) Initiative (U. S. Department of Ed., 2009) including approximately \$4.35 billion through the *American Recovery and Reinvestment Act of 2009*. One of the six core areas of RTT specifically focused on policies concerning teacher evaluations. There was growing concern about the need for evaluation systems that could possibly yield higher-quality information on how to improve teacher performance. There was also evidence that some of the evaluation policies promoted by RTT, such as multiple measures and multiple rating categories, could help to produce more valid and reliable estimates of teacher quality (Weisberg, Sexton, Mulhern, & Keeling, 2009; Jacob & Lefgren, 2008; Felder & Brent, 2004). If research links teacher quality with student achievement, then the job of effectively evaluating teachers becomes more important than ever. Given the financial investment made through RTT to promote specific types of teacher evaluation policies, it is reasonable to explore the ways that schools evaluate teachers.

Virginia's Uniform Performance Standards (2011)

In 2011, the coordinated work efforts of the Virginia Evaluation Work Group and the Center for Innovative Technology produced a document known as *Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers* (to be referred to as Guidelines). Various members of the education community were involved in the creation of this document whose goal it was to

better define teaching performance standards across the commonwealth. Along with teaching performance standards, specific examples were included that showed how teachers might meet each standard in their practices. The Virginia Teacher Evaluation Work Group was comprised of school board presidents, superintendents, leaders from teacher organizations, as well as those representing higher education. Along with consultants and the staff of the VDOE, the Guidelines were adopted on April 28, 2011.

One of the goals of the State Board of Education is to define performance standards and evaluation criteria for educators, including teachers, principals, and superintendents in order to provide guidance in implementing consistent teacher evaluations based on standards. These standards are found in the Board of Education's *Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers, Administrators, and Superintendents*. For example, in the section concerning the quality of classroom instruction and educational leadership, Section 22.1-253.13:5 of the *Code of Virginia* states (in part) that evaluations "shall be consistent with the performance objectives included in the *Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers, Administrators, and Superintendents*." Teacher evaluations should align with the curricula for the school and should be for the purpose of identifying individual needs for professional development.

Concerning the employment of teachers, a summation of Section 22.1-295 reads that a procedure will be developed by individual school boards so administrators may evaluate the effectiveness of its instructional personnel to use methodology, content knowledge, and classroom management to influence student academic progress. The *Code of Virginia* has two major goals. They are (a) to see that teacher evaluations are consistent with the performance standards and (b) to ensure that school boards address student academic progress in their evaluation of instructional personnel.

In order to meet these goals, seven uniform performance standards were developed and each one includes specific indicators as examples. Indicators are numbered under each standard, but they are not meant to be an exhaustive list. **“Further, all teachers are not expected to demonstrate each performance indicator”** (p. 4). The document contains details that show the purpose of the performance indicators is to “clarify” the expectations of teacher job performance and that **“Performance ratings are *NOT* made at the performance indicator level, but at the performance standard level”** (p. 4). An example of how a teacher might show proficiency of Performance Standard 1: Professional Knowledge is given in Performance Indicator 1.3: “Demonstrates an ability to link present content with past and future learning experiences, other subject areas, and real-world experiences and applications” (p. 5). An example of how a teacher might

show proficiency of Performance Standard 3: Instructional Delivery is given in Performance Indicator 3.4: “Differentiates instruction to meet the students’ needs” (p. 4).

The authors of the Guidelines offer a foundation of research on which the standards are based. In Virginia, the requirement states that teacher evaluations should be consistent with the performance standards documented. The goal of standards is to collect data that documents the performance of all instructional personnel, as well as to define expectations. Monitoring instruction in order to give meaningful feedback to teachers in the evaluation instrument is to be done by the administration. Feedback is usually given by means of a rating system based on a Performance Appraisal Rubric. The ratings range from Exemplary to Unacceptable and are specific in the way they describe the necessary behaviors to be shown in order to receive a certain rating. For example, in Performance Standard 1: Professional Knowledge, a teacher who received a rating of Exemplary (4) not only met the standard, but also demonstrated *extensive* knowledge of the content area, but enriched the curriculum. Teachers who receive a rating of Exemplary in any area are expected to be role models and leaders for other teachers. Ratings of Proficient (3) or Developing/Needs Improvement (2) may also be selected by the chosen by the administrator. A rating of Unacceptable (1) might mean that the teacher presented inaccurate or out-of-date content. It might also mean the teacher

inadequately addressed the developmental needs of the students in the classroom. Each Performance Standard includes a specific focus, sample performance indicators, and a performance appraisal rubric.

The *Uniform Performance Standards for Teachers* are purposefully aligned with the standards adopted by several professional organizations that define expectations for teacher performance and evaluation, specifically Interstate Teacher Assessment and Support Consortium (INTASC) and National Board for Professional Teaching Standards (NBPTS). INTASC does not include measures of student achievement in their evaluation of teacher quality.

The National Board for Professional Teaching Standards were created in 1987 for the purpose of “establishing high and rigorous standards for what accomplished teachers should know and be able to do” (NBPTS, 1987). NBPTS has informed standards for beginning teacher preparation and licensing standards developed by INTASC since 1992 and these standards have been adopted by more than 40 states as the basis for licensing teachers and approving programs. Both NBPTS and INTASC standards are responsive to a multicultural and multilingual student body that includes diverse approaches to learning except INTASC does not link student achievement to teacher quality.

The National Board Standards have also been important because they redefined teacher assessment. They have closely examined the practice of teaching

in relation to learning. One reason these standards promote productive learning through the evaluation process is because they are expressed in performance terms. They describe what teachers should know and be able to do rather than listing courses that teachers should take in order to be awarded a license or certificate. A multiple-methods approach to teacher evaluation is more appropriate because teaching is so complex (Felder & Brent, 2004). Local evaluations continue to vary widely across divisions and are often based on checklists of teacher behaviors that are not associated with effectiveness. Under the *Virginia Uniform Performance Standards*, each division was allowed to propose a teacher evaluation system that was designed based upon its own needs and context, provided that the system adhered closely to federal requirements.

Many states chose value-added measures (VAM) of determining student growth using test data. They employ complex algorithms in order to calculate how much teachers contribute to students' learning, holding constant factors such as demographics. The purpose of VAM in teacher evaluation is to determine the efficacy of the use of data from student test scores and to make key personnel decisions about classroom teachers, such as retention, dismissal, and compensation of teachers and principals. A number of states have adopted new or revamped teacher evaluation systems, based in part on value-added measures. Some states

require that up to 50% of the teacher evaluation must be based on data from student test scores.

When presented with this design, the Virginia Department of Education decided that it was not in the best interest of its schools to adopt this method. They chose the student growth model designed by them and not the value-added method in other states. Virginia's evaluation focuses heavily on student growth and teachers became directly responsible for student progress. The final category on Virginia's teacher evaluation system is entitled "Student Academic Progress" and the score from this category counts four times as much (40%) as the other six categories (10%).

Within the same time frame, many states adopted more rigorous college and career readiness standards. These new standards were intended to raise the expectations from "every student will earn a high school diploma" to the ambitious goal of "every student will be on-target for success in postsecondary education and training." These new assessments included constructed responses instead of exclusively multiple-choice style items and demanded higher-order thinking skills with up to a two-year increase in expected reading and writing skills. The newness of the assessments coupled with increased rigor has resulted in significantly lower numbers of students reaching the "proficient" levels.

According to the VDOE website, training on the performance standards-based evaluation model began with performance pay pilot schools in Phase I during the year 2011-2012. Teachers were allowed to voluntarily participate in the new evaluation process that would be mandatory for all teachers in Virginia beginning in 2012. Phase II training began July 1, 2012 when all teachers were evaluated under the Guidelines. Therefore, principals were still engaged in training on evaluating teachers under the new system at the same time all teachers were required to participate. Phase III training began in 2013, and in the second year, the evaluation was applicable for all teachers in Virginia.

The Challenge

The challenge for school leaders was that new teacher evaluation systems demanded the inclusion of student data at a time when scores on newly created assessments were dropping. The new and more rigorous college and career readiness standards meant that students and teachers were going to have to make adjustments with these new requirements. Scores dropped and both teachers and administrators began to get anxious about the changes. Also influencing the drop in scores was the fact that new Technology Enhanced Items were put in place on the standardized test. Students no longer had a fixed number of multiple-choice items. There were boxes in which to type answers, a click-and-drag feature, as well as the possibility that more than one answer could be correct. The fears that accompany

any new program were made worse by the inclusion of data that continued to show a significant decline, at least for the time being. Administrators were concerned that the new evaluation systems were eroding public trust and thus detrimental to building a community of collaboration among the teaching force. A critical issue for improving evaluation in general was to develop a system that could help to meaningfully assess the performance of teachers at every level from novice to expert.

Virginia's Performance Standards

Each of the Uniform Performance Standards is assigned a percentage in order to calculate a teacher's summative score. Of the first six, standards one through six each weigh 10% of the overall score, or 60% total. The final standard, student academic progress, weighs 40% of a teacher's summative score. Teachers are encouraged to use multiple measures to show students' academic progress each school year. The final score is calculated mathematically using the percentages given.

A brief overview of each standard in the Guidelines (2001) is important in understanding the current expectations of what teachers should be able to do. An explanation of each standard from the Guidelines (2011) will be accompanied by at least two example indicators (although there are many for each standard).

Following the example indicators, the rubric used to measure a teacher's level of

expertise relative to each of the Standards will be examined. An explanation of what it means to receive a rating of Proficient for each standard will follow and will include examples of teacher behaviors observable to an evaluator who believes that rating to be appropriate. Following the rating system, the research base underlying each standard will be described and examined. Since the Uniform Performance Standards are founded upon a research base, it will be beneficial to examine each piece of research and attempt to align the research with the results. Each of the seven sections will conclude with a critique to examine fit and alignment related to the research provided in the Guidelines.

Performance Standard 1: Professional Knowledge (10%). *The teacher demonstrates an understanding of the curriculum, subject content, and the developmental needs of students by providing relevant learning experiences.*

Two examples of possible performance indicators are (a) integrates key content elements and facilitates students' use of higher level thinking skills instruction and (b) communicates clearly and checks for understanding. A Proficient rating is the expected level of performance and indicates that the evaluator believes the teacher demonstrates an understanding of the curriculum, the content area, and the developmental needs of all students as the teacher is seen providing relevant learning experiences.

Four studies and a research review are referenced as the research foundation for this Standard. Teaching is an incredibly complex task that requires the teacher to have knowledge of content, pedagogy, curricula, learners, cultures, and the overall purpose of education. The authors of the Guidelines note that teachers with deep content knowledge are more likely to plan activities and lessons that help students internalize material at a much deeper level than teachers without deep content knowledge.

Various studies suggest that teachers with a strong background in their own content area are more likely to ask high-level questions of the students, allow for more student-based activities, and engage students in ways that encourage them to explore alternative solutions (Weiss & Miller, 2006; Wenglisky, 2000). However, a teacher's content knowledge alone will not translate into effective teaching practices. For instance, according to Hill, Rowan and Ball (2005), strong teacher knowledge in mathematics translates into student learning only if teachers are able to use their knowledge to plan developmentally appropriate activities in order to promote student growth. These researchers examined the content knowledge of elementary teachers with a quantitative survey in the content area of mathematics. They found that teachers' knowledge of mathematics significantly contributes to student mathematics learning while controlling other key student and teacher characteristics.

Rowan, Chiang, and Miller (1997) also examined the content knowledge of teachers with a quantitative survey. In their study, survey results were paired with the college majors of high school mathematics teachers. Their findings were that high school teachers who answered quiz problems correctly were more effective in helping their students achieve in the area of mathematics than teachers who answered fewer problems correctly. They also found that students whose teachers majored in mathematics at either the undergraduate or graduate level achieved more than those whose teachers did not major in mathematics, although the effect was small ($SD = .015$). Goldhaber and Brewer (1997) conducted two similar studies of high school mathematics teachers concerning their own college major and the possible effects on the achievement of high school mathematics students in that content area. In using quantitative methods to analyze their results, Goldhaber and Brewer (2000) found that students learned more from teachers who majored in mathematics than those whose teachers majored in anything other than mathematics.

The only study undergirding Performance Standard 1: Professional Knowledge that did not involve mathematics was a study conducted by Monk (1994), and it involved both mathematics and science. This study focused on content knowledge, as well as pedagogical knowledge of teachers and used the teachers' college coursework as a quantitative means of collecting data. Again,

high school teachers were the participants. The findings were that the amount of college-level mathematics or science courses taken by the teachers had a positive impact on student learning gains. The findings also showed that the effects of pedagogical coursework were more stable over time than the effects of subject matter preparation.

The research base for Performance Standard 1 contained studies from only one area, that of mathematics. High school teachers were used in three of the four instances and an emphasis was placed on mathematics in order to make the generalization that professional knowledge is positively correlated with student achievement. A research summary by Rice (2003) found that a teacher's coursework requirements in both pedagogy and content area had a positive impact on student achievement in middle and high school, but the focus was only on secondary grades in mathematics. The findings among all of the sources listed heavily emphasized mathematics. The research base generalized a direct correlation between content knowledge, student achievement, and teacher quality. Although a case by Hill, et al. (2005) was provided that found teacher content knowledge crucial in effective instruction, no research was ever reviewed in the area of literacy or social studies.

In summary, the analysis for the research base for this Standard revealed some confusing issues. They were (a) all of the studies quantitatively examined

high school mathematics teachers except one which included elementary teachers, (b) all of the foundational research cited was based on *Becoming a Nation of Readers* (Anderson, 1985), (c) after this time there was an abrupt shift by the IES to use only quantitative studies, (d) three of the four research studies were focused on high school teachers, and (e) there were no studies about the developmental needs of students. Meeting the developmental needs of all students is not equal to measuring student achievement.

Performance Standard 2: Instructional Planning (10%). *The teacher plans using the Virginia Standards of Learning, the school's curriculum, effective strategies, resources, and data to meet the needs of all students.*

Two examples of possible performance indicators are (a) planning for differentiated instruction and (b) aligning lesson objectives to the school's curriculum and student learning needs. A Proficient rating is the expected level of performance. However, an evaluator who gives a rating of Developing/Needs Improvement in this area indicates he/she believes the teacher is inconsistent in using the curriculum as a guide and is ineffective in using strategies, resources, and data in order to meet the needs of all students.

The research-based studies detailing the explanation of Standard 2: Instructional Planning uses the term expert in order to describe experienced teachers. The authors of the Guidelines take the position that having a teaching job

for a certain period of time makes the teacher an expert. The assumption is that this teacher is able to use a textbook loosely and veer from traditional plans when appropriate because of experience. Teachers, like students, have various life experiences. Many teachers do not possess this skill naturally. However, the goal of this Standard is to measure how well the teacher presented knowledge during the allotted time. It is assumed that every teacher will be able to allow the individual needs of the students to guide daily lesson plans and make sure every student achieves success a measure of success.

In the detailed explanation of Standard 2, it is assumed that teachers will be able to tell students what they are “expected to achieve.” Being very specific about the time in which students will have content made available to them is important. However, under this Standard, teachers are expected to set achievement goals for each student. Currently, teachers have less flexibility with time constraints placed upon them by school boards and site administrators, mainly because of high-stakes testing. High-stakes testing has been named by some as the cause of a shallow, narrow curriculum and the focus on testing to the detriment of student learning (Au, 2007). Teachers know their class times are fixed, but they must consider the needs of the students. Human behavior cannot be predicted, yet Standard 2 holds teachers accountable for teaching the Standards of Learning and the school curriculum while being responsible for the achievement of their students, even

though they may or may not have access to the resources they need to meet the needs of all students.

Standard 2 also details data-driven planning and the “proper” use of pacing guides (Stronge, 2010). It mentions that teachers should properly use student data daily in order to meet the needs of all students. They are to gather and analyze student data properly, make the appropriate application to good practice, and use this data to improve student results. With a fixed schedule, teachers may find it particularly difficult to study student data efficiently to be used for the proper assessment of student learning. A teacher’s performance, according to this Standard, will be based on whether or not he or she uses student data “properly” and it requires teachers to use data in order to make “good decisions” (Hess, 2009).

In summary, the analysis for the research base for this Standard revealed that at least some of the assumptions are not well defined. “Expert” is assumed to be relational with experience, but there is no guarantee of the accuracy of that statement. The data referenced may apply, in this case to student data, but it is unclear as to which categories of data are intended. “Properly” is another term that is unclear and most difficult to standardize.

Performance Standard 3: Instructional Delivery (10%). *The teacher effectively engages students in learning by using a variety of instructional strategies in order to meet individual learning needs.*

Two examples of possible performance indicators are (a) builds upon students' existing knowledge and skills and (b) consistently reinforces learning goals throughout the lesson. A Proficient rating is the expected level of performance. However, an evaluator who gives a rating of Unacceptable in this area indicates he/she believes the teacher's instruction does not adequately address the students' learning needs.

At first glance, it would seem that this Standard emphasizes instructional strategies. Student differences in both cognition and ability must be taken into account in order to reach each student at developmental levels. However, the detailed explanation of what a teacher should be able to know and to do in order to receive a Proficient rating in this category is lengthy. The research-based description of a typical student makes it clear that students come to the classroom in possession of a variety of abilities, interests, and background knowledge. Therefore, a one-size-fits-all approach to instruction would be ineffective. The work of Carlson, Lee, and Schroll (2004) were cited in the Guidelines because their findings indicated that a teacher's instructional strategies mattered. Walberg's (1984) study that used quantitative measures was the only study mentioned that positively correlates instruction with student achievement. He stated that the practice of reinforcement alone, which accounted for 1.17 standard deviations on educational outcomes, could be generalized across the board.

Walberg's quantitative data collected on cues, engagement, and corrective feedback measured approximately one standard deviation. His study was the only research referenced that connected instruction and diagnostic-prescriptive methods with student learning.

This Standard requires effective teachers to design instruction that “motivates each student” and “communicates content in such a way that students are able to comprehend based on their individual prior learning and ability.” Students learn in a variety of ways and at various rates. With fixed class schedules, teachers are responsible for motivating student learning while properly using data they have gathered to design a student-driven lesson. Dunn, et al. (2009) indicated that at-risk students had increased achievement of one standard deviation when teachers accommodated for learning styles, while Lovelace (2005) suggested that instruction in general is positively related to student attitudes and achievement. Lovelace presented a meta-analysis of the extant research and the authors of the Guidelines undergirded the student engagement section using only his writings.

Schroeder, Scott, Tolson, Huane, and Lee (2007) found that connecting instruction to real-world problems is one of the most powerful practices a teacher can employ in order to increase student learning. Authors of the Guidelines referenced one study that took place in a mathematics classroom and indicated that teachers ask more than 99 % of all questions asked during a lesson (Craig & Cairo,

2005). The authors also used a case study by Stronge, Little, and Grant (2008) which found that effective teachers ask approximately seven times higher cognitive-level questions than those teachers considered to be ineffective. This is considered to be best practice in the educational community and the responsibility rests upon the teacher to get the desired achievement results.

The Guidelines draw attention to several specific instructional practices that the authors see as needful for effective teachers. They are (a) effective teachers stay involved with the lesson at all stages so adjustments can be made based on feedback from the students (Tursman, 1981), (b) effective teachers use research-based strategies to enhance the time students spend with teachers by making instruction student-centered (Johnson, 1997), and (c) effective teachers know that instructional strategies that engage the prior knowledge of students in an inquiry-based, hands-on format facilitate student learning (Covino & Iwanicki, 1996).

In summary, an analysis of the research base for this Standard revealed that the researchers admitted there is no single practice that is effective with all content areas and all grade levels. The authors based the idea of common effective practices on a study conducted by McDonald and Elias (1976). Not included were studies more current than 1976. The same rating system for this Standard is used to evaluate the experienced teacher, as well as the novice teacher. A rating of Proficient is the target rating for all teachers, regardless the years of their

experiences. This exposes one of the greatest weaknesses of the Standards, that of evaluating a novice teacher in the same ways as an experienced teacher. If the proficiency level of an experienced teacher is what is desired by the evaluator, what are the expectations for a novice teacher? Expectations are inappropriately high for a novice in this case and the supporting research base is weak and out-of-date.

Performance Standard 4: Assessment of and for Student Learning (10%). *The teacher systematically gathers, analyzes, and uses all relevant data to measure student academic progress, guide instructional content and delivery methods, and provide timely feedback to both students and parents throughout the school year.*

Two examples of possible performance indicators are (a) involves students in setting learning goals and monitoring their own progress and (b) uses assessment tools for both formative and summative purposes. Even though a Proficient rating is the expected level of performance, an evaluator who gives a rating of Exemplary in this area indicates he/she believes the teacher exceeds normal expectations and should serve as a role model for other teachers. In addition to systematically gathering and using student data to measure student achievement and to drive instruction, the evaluator is indicating that teachers who receive an exemplary rating are observed using a variety of assessments for the

purpose of instruction and teaching students ways to measure their own academic progress.

The research base for this Standard includes studies that indicate that teachers who regularly use formative assessments substantially affect student achievement. Black and Wiliam (1998) examined many empirical studies in their research review of this topic and found that formative assessments have positive effects on student achievement. Their studies also indicated that formative assessments were more effective for low-achievers than for other students. The authors of the Guidelines make the claim that formative assessments could reduce a possible achievement gap and raise overall achievement at the same time. There are no specific examples of these formative assessments offered.

The research foundation for Standard 4 specifically outlines several ways in which assessments are more likely to influence student learning. The claim is made that assessments are more likely to influence student learning if they are aligned with learning targets and are valid, reliable sources of gathering student data. The VDOE authors also believe that assessments do, in fact produce an accurate representation of student learning and state that when formative assessments are coupled with frequent informative feedback, student learning may be influenced. The research base for this Standard supports the use of formative assessments with very young students at a developmentally appropriate level so they will be able to

begin learning how to monitor their own learning. The researchers feel that assessments are more likely to influence student learning when the assessments emphasize testing processes and results. The authors of the Guidelines want assessment results communicated to the students in a “timely and effective manner.” They do not define the words “timely” or “effective” and state that a teacher’s evaluation rating in this particular standard is based upon whether or not the teacher documents student learning results properly and keeps complete records of student learning (Black & Wiliam, 1998).

Concerning the importance of feedback in learning, Zacharias (2007) quantitatively surveyed students and teachers in his study. Students felt that informative feedback made them aware of their mistakes and how to correct them, and were indicators of teacher expectations. On the other hand, teachers felt that providing feedback was an important part of instruction, but could prove to be arduous and painstaking. Concerning the use of multiple methods of learning, Guskey (2007) found that most teachers and site administrators believed portfolios were a valuable tool in displaying student achievement and described the usefulness of division, state, and national assessments as not very valuable. Therefore, teachers are many times instructed to use multiple sources of assessments when, at the end of the year, the major emphasis is placed on high-stakes assessment results. Tomlinson (2007) suggested that teachers must find a

proper fit for assessing each student since it is believed to be such an important mode of communication. In summary, an analysis for the research base for this Standard revealed some convoluted evidence. According to this research, teachers should align their instruction to standards, maximize instructional time, and work harder to cover more material in a given amount of instructional time. However, the research offered in the previous Standards indicated that students needed more time to develop and construct their own knowledge. It would be inconsistent for the authors to previously suggest that students need more time to develop and construct their own knowledge while rating teachers on “working harder to cover more material.” The depth of the curriculum may eventually suffer as the breadth of the curriculum continues to widen. Vogler (2002) stated that standardized tests motivate teachers to adopt better curricula and more effective pedagogical methods. In most cases, teachers do not have input in adopting the curricula, nor do they have voice in what is covered in standardized tests.

Concerning standardized testing, studies indicate that high-stakes assessments encourage teachers to narrow the curriculum and to place more focus on rote memorization and worksheets, leaving less time for higher-order skills to become part of daily lessons (Hamilton & Stecher, 2004; Jones & Egley, 2004). These same studies indicate that standardized assessments actually limit a teacher’s

creativity in lieu of prescribed approaches of instruction in order to cover every part of the curricula (Hamilton & Stecher, 2004; Jones & Egley, 2004).

No consideration is given for the minimal time within a school day that teachers have to efficiently document each student's progress, give applicable feedback, and communicate individualized feedback to both students and parents in a timely manner. All of the previously mentioned are meaningful goals for teachers to set in order to facilitate student achievement. The indicators for this Standard are used to rate a teacher on how well they assess student learning. For the evaluator, it may prove challenging to find strong evidence that a teacher does this consistently and effectively.

After pointing to studies that find both positive and negative aspects of using standardized testing to try and motivate teachers, the authors of the Guidelines suggest that teachers maintain a balance between state and national level assessments and classroom level assessments to optimize student learning. In order for a teacher to receive a rating of Proficient according to this Standard, a major portion of allotted class time would be needed for such assessments. The trade-off would be that class time would be taken for standardized assessment practice when students could be using this time for their own learning. It may prove challenging for teachers to find enough time to analyze all of the data these assessments would generate, including daily assessments for student-driven lessons, structured

assessments in order to meet the needs of all students, mandated benchmark assessments in each content area, as well as end-of-the-course standardized assessments. Teachers and evaluators have indicated that individualized assessments are a good idea, but could be “arduous and painstaking” (Zacharias, 2007).

Performance Standard 5: Learning Environment (10%). *The teacher uses resources, routines, and procedures to provide a respectful, positive, safe, student-centered environment that is conducive to learning.*

Two examples of possible performance indicators are (a) establishes clear expectations, with student input, for classroom rules and procedures early in the school year, and enforces them consistently and fairly and (b) respects students’ diversity, including language, culture, race, gender, and special needs. An evaluator who gives a rating of Proficient in this area indicates he/she believes the teacher uses routines and procedures to make the classroom respectful, safe, and student-centered, making this teacher’s classroom conducive to learning.

The foundational research for this Standard connects the learning environment with student achievement in several key ways. The researchers point to classroom management, classroom structure, a positive classroom climate, and classroom talk as the most important ways a teacher can set up a productive learning environment. One study examined a teacher’s organization in a first grade

classroom. The research team only observed students three times during one school year and observed letter-word reading. They found that “classroom organization significantly predicts spring word reading skills” (Cameron, Connor, Morrison, & Jewkes, 2008). It is unclear how they arrived at their findings.

Another part of the research base for this Standard was a professional development session delivered to the Kentucky Association of School Administrators by James Stronge in 2011. He referenced his own books ten times and the books of other authors four times. Therefore, most of the material was from either his own writings or information from an article that took the same stance as did he on keys to developing a successful learning environment (Emmer & Stough, 2001). All of these authors believed the “extant research is fairly clear that good classroom management has a positive influence on students’ motivational development” (Stronge, 2007; Emmer & Stough, 2001).

Another study took a different approach and examined rule clarity and teacher monitoring in a secondary school mathematics class to undergird their findings about how to provide a positive learning environment (Kunter, Baumert, & Koller, 2007). Again, Stronge’s work was cited when he found the top quartile teachers to be more organized than the other 75%. Stronge found these teachers communicated higher behavioral expectations to students than ineffective teachers, though he never clearly defined the word “ineffective” (Stronge, Ward, Tucker, &

Hindman, 2008). Cameron, et al. (2008) found a positive correlation between achievement in reading and an organized teacher. They found that practicing classroom routines in the beginning of the year makes a teacher's work easier every day and believed these routines are responsible for yielding gains in reading scores.

Wang, Haertel, and Walberg (1994) gathered over 11,000 statistics from a knowledge base about the questions "What helps students learn?" When the Guidelines were adopted in 2011, these statistics were 17 years old. The actual study was cited in the research base document for the Guidelines and came from Stronge's book (2010). His book was current, but again the information was taken from the same source. Another piece of evidence was a research synthesis by the same team of Wang, Haertel, and Walberg (1994). Hamre and Pianta (2005) examined the interactions between teachers and first grade at-risk students. They found positive effects at all grade levels.

A study of secondary teachers indicated that those who got to know their students on a personal basis believed students defined and achieved goals, but did not specify what those goals were or how they were achieved (Pressley, Raphael, Gallagher, & DiBella, 2004). A study by Allington and Johnston (2000) indicated that a population of middle school students developed positive attitudes toward the course when teachers interacted with students on a regular basis and believed that

exemplary teachers use authentic conversation to learn about students, though a working definition of “authentic” was not clearly defined.

The following table contains a list of attributes that the research base claims will help students to develop positive attitudes toward a course. The table separates the attributes into appropriate learning environments in which students might find themselves and relies on opportunities for the teachers to interact with students on a regular basis. Most, if not all of the attributes necessary are presented as the job of the teacher. See Table 1.

Table 1

*A Summary of Selected Features of Positive Learning Environment for Standard 5:
Learning Environment*

Positive Learning Environment Attributes	Features of Attributes
Classroom management and structure	Identifying and communicating desirable behavior, consistently applying rules and procedures, monitoring student behavior, taking preventive rather than reactive management actions, pacing class activities and transitioning between tasks smoothly, maximizing instructional time, keeping students on tasks, making learning meaningful
Positive classroom climate	Cooperation among teachers and students, common interest and values, pursuit of common goals, a clear academic focus, well-organized and well-planned lessons, explicit learning objectives, appropriate level of task difficult for students, appropriate instructional pace
Classroom talk	Respectful, supportive, and productive, modeled by teachers, practiced with students

(From The Research Base for the Uniform Performance Standards for Teachers, 2011, p. 21-22)

Performance Standard 6: Professionalism (10%). *The teacher maintains a commitment to professional ethics, communicates effectively, and takes responsibility for and participates in professional growth that results in enhanced*

student learning.

Two examples of possible performance indicators are (a) engages in activities outside the classroom intended for school and student enhancement and (b) builds positive and professional relationships with parents/guardians through frequent and effective communication concerning students' progress. Even though a Proficient rating is the expected level of performance, an evaluator who gives a rating of Developing/Needs Improvement indicates he/she sees very little interest in professional growth opportunities measured by attendance or only sees occasional application in the classroom.

Studies supporting this Standard may be placed into three categories. They are (a) teachers' behaviors, (b) teachers' growth, and (c) teachers' contribution to the learning community. Most of the research base for Professionalism included beliefs about teacher behaviors. One study indicated that effective teachers are warm, friendly, and caring (Walls, Nardi, von Minden, & Hoffman, 2002) even though Pajares (1992) indicated that beliefs about teaching are well formed by the time a student teacher enrolls in a teaching program. Pajares further suggests that these beliefs were self-perpetuating even in the face of contradiction. The authors of the Guidelines believe that teachers impact school success through their attitudes and they believe a teacher's enthusiasm and motivation connect teacher effectiveness with student achievement. However, there were no studies to back up

this claim. The Supreme Court decision in *Ambach v. Norwick* (1979) was referenced to suggest that teachers should be held to a higher standard than others and should serve as role models for students. Most of Standard 6: Professionalism deals with teacher behaviors and the consequences of bad behavior, that of dismissal.

One meta-analysis claimed to provide evidence that teachers who receive at least 49 hours of professional development were able to boost their students' achievement by 21 percentile points (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). They also went a step further when they suggested that this effect size is fairly consistent across content areas. No other evidence was provided and the comments about professional development were only three sentences in length.

According to the Guidelines, one empirical study (Little, 1993) and a book (Senge, 1990) provide the evidence that effective teachers will contribute to the learning community both individually, as well as collectively. The authors believe that effective teachers should be critics of educational policies, innovators of instructional techniques, and make internal changes that impact student learning. They also use a book entitled *The Fifth Discipline: The Art and Practice of the Learning Organization* (Senge, 1990) as the basis for indicating that teachers should network collectively in order to advance school improvement.

In summary, the analysis for the research base for Standard 6: Professionalism revealed that teacher behaviors such as attitude, enthusiasm, and motivation have a positive correlation with student achievement. Only one study of secondary science teachers was offered to suggest that professional development had an impact on student performance. The basis for this standard came from upper level science classes alone.

Performance Standard 7: Student Academic Progress (40%). *The work of the teacher results in acceptable, measurable, and appropriate student academic progress.*

If a teacher receives a Proficient rating in the previous six standards, it is most probably the case that the results of teaching (the focus of this standard) would be positive. However, this Standard is unique in that the Guidelines recommend that any student data collected under this category be revisited many times throughout the academic school year. A teacher who received a Proficient rating in this area would have collected several types of data throughout the year to compare and determine rates of student academic growth. Two examples of possible performance indicators are (a) sets acceptable, measurable, and appropriate achievement goals for student academic progress based on baseline data and (b) uses available performance outcome data to continually document and communicate student academic progress and develop interim learning targets. An

evaluator who gives a rating of Exemplary in this area indicates he/she believes the teacher's work has resulted in a high level of student progress among all populations of students.

Most of the evidence provided in the research base for this Standard centered on effective teachers helping all students to make progress. The authors of the Guidelines suggested that numerous studies conducted worldwide provided documented evidence that effective teachers had a significant impact on student achievement. They summarized the findings by referencing seven studies. Aaronson, Barrow, and Sander (2007) indicated that teachers with average effectiveness showed achievement gains with lower achieving students, but not with higher student achievers (Sanders & Rivers, 1996). This particular finding was taken from an empirical study that was 15 years old when the Guidelines were adopted. Sanders and Rivers (1996) found that teacher effects on student achievement are cumulative and residual. One study is cited as evidence that variations in teacher quality accounted for at least 7.5% of the total variation in measured achievement gains (Hanushek, Kain, O'Brien, & Rivkin, 2005) while another study claimed that teachers contributed 3%-10% of the variability in student gain score, when controlling for prior achievement and background data (Rowan, et al., 1997). Stronge, et al. (2008) suggested that teachers who were highly effective in producing student achievement gains in the top quartile tended

to produce top quartile residual gain scores in the content areas of reading, mathematics, science, and social studies. These data were gathered from one end-of-course content assessment. In the study, it was suggested that teachers who were ineffective (as defined by being in the bottom quartile) in one content area tended to be ineffective in all four of the above mentioned content areas.

Hattie (2003) found that effective teachers know how to monitor and assess student progress and to provide more relevant, useful feedback. Hattie's work also suggested that effective teachers know how to develop and test hypotheses about learning difficulties or instructional strategies. One experimental study by Fuchs, Deno, and Mirkin (1984) indicated that teachers who monitored student growth on a regular basis showed at least four outcomes. They are (a) greater student achievement, (b) more improvement in their instructional structure, (c) greater realism reflected in their pedagogical decisions, and (d) students who were more conscious of their own learning goals and progress.

In summary, the analysis for the research base for this Standard revealed that the authors believe effective teachers are able to document student achievement in various ways. However, there are no specific examples from the research to show that documenting student achievement leads to better teaching practices. Student achievement is the largest piece of a teacher's summative evaluation, 40% of the overall score. Further, this Standard assumes that teachers are able to set

“meaningful student achievement goals” for each of their students (VDOE, 2011).

One research team noted that teachers effected student learning by monitoring only when they modified instruction based on monitoring data (Stecker, Fuchs, & Fuchs, 2005), but monitoring alone was not shown to boost student achievement.

Summary of the Research Base

In looking at the research that provides a foundation for Guidelines, several issues emerged:

1. Several of the studies were quantitative, in the content of mathematics, mostly in high school level settings (Hill, et al., 2005; Rowan, et al., 1997; Goldhaber & Brewer, 1997; Goldhaber & Brewer, 2000).
2. The findings from some studies were inconsistent with the findings from other studies of the same type (Vogler, 2002; Black & Wiliam, 1998; Zacharias, 2007; Guskey, 2007).
3. The fact that student learning is more than student achievement is not considered in the Guidelines.
4. The findings of the research base go well beyond the data and are not generalizable.
5. The evaluators were in training the same year the system began to be implemented for teachers.

Concluding Thoughts

Evaluations have shifted over time and, most likely, teachers expect that they will shift again. Teacher practices are driven by beliefs (Pajares, 1992) and if teachers' experiences with evaluation systems are that the systems will continue to change regularly, then it would not be surprising to find that teachers do not have a clear understanding of the evaluation system under which they work. They may, in fact, be skeptical of the next evaluation system that will bring more change. For this reason, there may be tension among teachers with concern about the possibility of a failure to meet expectations. Since the evaluation system in Virginia is in place in some form across all school divisions, it is deserving of further examination.

In this short explanation, the research underlying the Guidelines is inconsistent. Among the studies cited, there is no research that examined teacher practices that changed over time. A study of the Guidelines should have included a review of research on teacher thinking, beliefs, and perceptions about the evaluation process. In addition, there are no provisions made in the Guidelines for specialists. The policy is currently that every teacher is to be evaluated by the same standards and there are no exceptions.

Teacher thinking is important because what teachers think determines what teachers do in the classroom (Richardson & Placier, 2001), but teacher thinking was not grounded in the Guidelines. Teacher perceptions are important because if

teachers perceive that the evaluation process is for their benefit, they will act differently than if they believe the system is of no use to them and serves as just a formality (Schon, 1983). If administrators are to evaluate high-quality teachers who perform their duties effectively, they must first discover what teachers understand about the evaluation process and what their perceptions are about their experiences with the evaluation process. If the goal of the current evaluation system is to identify effective teachers, then teachers should have a voice concerning ways the system is or is not accomplishing that goal. Understanding how teachers perceive the evaluation process will go a long way in ensuring that evaluation in Virginia is meeting that goal.

Chapter 3

Methodology

Research Questions

The research questions are (a) what are teacher understandings of the teacher evaluation process? (b) what are teacher perceptions about their experiences with the teacher evaluation process? and (c) in what ways do teachers believe that evaluation informs their practices? First, a rationale will be presented for the qualitative methodology that was used in this study. Then, a brief explanation of each of the phases of this study will be outlined. Finally, a description of the procedures and protocol that were used in collecting the data will be given. This section will conclude with how the data were represented.

The Approach of Phenomenology

One kind of research that proved beneficial for this study was that of phenomenology. Phenomenologists give careful, rich descriptions of the ordinary experiences of everyday life (Hycner, 1985; Moustakas, 1994). The purpose of this approach is to illuminate the specific and to identify phenomena through how they are perceived by the actors in a situation. Phenomenology translates into gathering deep information and perceptions through inductive, qualitative methods such as interviews, discussions and participant observation; and represents it from the perspective of the research participant(s). Epistemologically, these approaches are

based in a paradigm of personal knowledge and subjectivity and emphasize the importance of personal perspective and interpretation. Pure phenomenological research essentially seeks to describe rather than explain, and begins from a perspective void of hypotheses or preconceptions (Husserl, 1970). More recent researchers have refuted the possibility of beginning without biases. They have emphasized the importance of clarifying exactly how the interpretations and meanings have been placed on findings, as well as how researchers have been made visible in the frame of research as an interested and subjective actor rather than a detached and partial observer (Plummer, 1983; Stanley & Wise, 1993). Phenomenological methods are particularly effective at bringing to the surface the experiences and perceptions of individuals from their own perspectives, thus challenging normative assumptions.

When teachers were asked to describe things as they have individually experienced them their descriptions were the result of how they internalized the process of being evaluated. Their descriptions included perceptions, beliefs, memories, decisions, and actions. As teachers described their experiences with evaluation and what they understood about the process, they were actually describing what they have internalized in their subjective consciousness (Clark & Peterson, 1986). While school systems in the United States have drifted toward more standardization of evaluation, the reasons for doing so have been in

compliance with mandatory state and federal regulations. The interview questions were based on critical reflection about the phenomenon of teacher evaluation experiences and related knowledge. Since individual teachers internalized their own experiences and perceived situations differently, the fact that the evaluation process affected teachers differently was recognized.

Four defining characteristics of phenomenology. According to Giorgi (1989), the phenomenological approach has four defining characteristics. They are that the researcher (a) rigorously describes experiences, (b) uses the phenomenological reductions, (c) explores the intentional relationship between persons and situations, and (d) uses imaginative variation to disclose the structures of meaning in human experiences. This study is about the understandings and perceptions teachers had about the evaluation process and the ways they believed evaluation informed their practices. What teachers think affects what they do (Richardson & Placier, 2001). It was necessary to conduct interviews with individual teachers in order to understand the deeper perspectives that may only be captured in face-to-face conversations. Participants were asked to describe their personal perceptions about the evaluation process and how they believe the current evaluation system influences their practice.

Merton (1995) grounded his belief in the Thomas Theorem when he said that in studying the perceptions of people, one must allow them to define their own

situations. “If men define situations as real, they are real in their consequences.”

Access, role, reciprocity, trust, and rapport all build upon a mutual understanding and will lead to more accurate interpretations (Yeh & Inman, 2007).

Context of the Study

This study was conducted in one elementary school in Virginia. The rationale for using one school was that all of the participants were evaluated by the same administrator, presumably in similar ways. That decision removed some of the differences that might have been exposed had the research taken place with randomly chosen participants in different elementary schools across the state. This particular administrator worked with these teachers every day. When the participants described their experiences, they were very different, but the evaluator was the same evaluator.

As is the case everywhere, elementary schools in Virginia look very different in terms of populations, teachers, and administrators. Some schools have several administrators to oversee school responsibilities, while others do not. The same basic evaluation system is used for all teachers in Virginia; yet there are some differences in evaluation practices as a result of school division decisions about how to best implement the evaluation authored by the Virginia Department of Education (VDOE, 2011). Some administrators have more time to devote to the

evaluation of teachers while others have less. These differences affect the evaluation process in schools.

The school. One school was selected and it will be known as Bells Elementary School (BES). At BES there was one principal, one full-time dean of students, and one part-time assistant principal who retired from full-time administration two years previously. The principal was in his fourth year as principal. The school was located in a relatively small, urban city in Virginia with various sizes of houses, duplexes, apartments, and mobile homes.

The average number of students for elementary schools in Virginia is 527, so BES was comparatively small with 375 students. The student to teacher ratio was 14:1, which was lower than both the average elementary school in the United States (15.6) and the average elementary school in Virginia (16.4). The ratio of 14:1 was equal with other Southwest Virginia, urban Virginia, and average Virginia schools. The school campus had several separate buildings, so students entered and left buildings to travel to other places on campus. Areas common for all students were the cafeteria, library, playground, and gymnasium. Visitors were accompanied to the destinations by school staff members with keys for all areas.

The students. There were 375 students in the school, with an average of 72 students per grade level in grades K-2 and an average of 47 students per grade level in grades 3-5. There were 18 students in the preschool class. The

demographic breakdown of BES was very different from that of a typical elementary school in the state of Virginia. Males outnumbered females at BES, where the student body was 54% male and 46% female. The average school in this school division was 52% male, which meant that BES had a higher than average proportion of male students compared to other schools in the division. Most schools in the state of Virginia were 51.5% male and 48.5% female on average. See Table 2 for statistical comparisons.

Table 2

Statistics for Bells Elementary School, Schools in Southwest Virginia, Urban Schools in Virginia, and Schools in Virginia

	BES	Southwest Virginia	Urban Virginia	Virginia State
Free and Reduced Lunch	94%	29%	62%	32%
Teacher/Student Ratio	14:1	15:1	14:1	15:1
Socio-Cultural	African Am. 75%	African Am. 5%	African Am. 58%	African Am. 23%
	Native Am./Alaska 0%	Native Am./Alaska 0%	Native Am./Alaska 0%	Native Am./Alaska 0.3%
	Native 0%	Native 0%	Native 0%	Native 0.3%
	Asian 0%	Asian 2%	Asian 2%	Asian 6%
	Hawaiian/Pacific Islander 0%	Hawaiian/Pacific Islander 0%	Hawaiian/Pacific Islander 0%	Hawaiian/Pacific Islander 0.2%
	Hispanic 5%	Hispanic 2%	Hispanic 12%	Hispanic 12%
	Multiracial 1%	Multiracial 3%	Multiracial 6%	Multiracial 4%
	White 17%	White 88%	White 21%	White 53%
Spending Per Student	\$12,106	\$9,464	\$13,977	\$11,065

Two hundred eighty-five students or 75% of the student population at BES identified as African-American and made up the largest portion of the student body. A typical school in the school division was made up of 26% African-American students, so BES had a different ethnic distribution as compared to other schools in the division. On average, a typical elementary school in Southwest Virginia had an African-American population of 5%. The Hispanic population in BES was only 5% (as compared to a 2% Hispanic population in schools in

Southwest Virginia and a 12% Hispanic population in the average Virginia school). At BES, 17% of the population identified as White, compared to 88% of students in an average Southwest Virginia school. The average urban Virginia school had only a 58% African-American population and a 12% Hispanic population.

Each school in Virginia had a statistical report online for public viewing. This school report card included statistics in a number of areas as reported to the VDOE annually. Ten school divisions in the same region of Virginia were selected for a comparative analysis. From each of the ten divisions, one elementary school with a PK-5 population was randomly chosen. From the division's report cards for each of these schools, the same statistics were collected and means were found for each category. Elementary schools were selected from each of the following divisions in Virginia: Alleghany, Bath, Bedford, Grayson, Montgomery, Pulaski, Radford City, Wythe, Tazewell, and Scott. These averages make up the category "Southwest Virginia Average" found in Table 2.

The VDOE considers 23% of Virginia's public schools to be in urban areas. Since BES also falls into the category of urban, it was useful to make comparisons between BES and other urban schools that were similar in size and demographics. Eight urban school divisions in the Commonwealth of Virginia were selected. From each of these school divisions, one elementary school with a PK-5

population was randomly chosen. From each school's report card, the same statistics were collected and means were found for each category in the table. Elementary schools were selected from each of the following school divisions in Virginia: Alexandria City, Arlington County, Hampton City, Henrico County, Newport News City, Norfolk City, Richmond City, and Virginia Beach City. BES was representative of other urban schools in Virginia, but not representative of an average school in Southwest Virginia. The BES school population was significantly different demographically than most schools in Southwest Virginia. For a statistical comparison between BES, Southwest Virginia Average, Urban Virginia Average, and Virginia State Average, refer to Table 2.

BES had a very high poverty level reflected in the number of students who qualified for free and reduced lunch, compared with the rest of the state. Those students receiving free and reduced lunch at BES was 94% as compared to 62% in other urban Virginia schools. The percentage of students in the state of Virginia who qualified for free and reduced lunch was 32%. This statistic was drastically different from the Southwest Virginia average of 29% of students receiving free and reduced lunch. The median household income in this urban setting was \$38,265 per year.

There were differences between BES and the other average schools in Southwest Virginia, Urban Virginia, and the Virginia state average. BES was very

different in terms of extremes from divisions in Southwest Virginia. BES was in Southwest Virginia, but the typical school in Southwest Virginia was predominantly White. BES also stood in stark contrast to other urban schools in that BES had a lower population of Hispanic students and students who were Multiracial.

The average school in Southwest Virginia reported spending \$9,464 per student each year. BES was part of an urban school division that spent an average of \$12,106 per student each year. This urban school division spends 27.9% more per student than does the average school division in Southwest Virginia. BES was in Southwest Virginia, but was not truly representative of Southwest Virginia schools. It was more representative of the average urban Virginia school.

The teachers. There were 41 teachers in this school including a librarian, two physical education teachers, one music teacher, one art teacher, one ELL/Spanish teacher, and four special education teachers. There was a reading specialist, an instructional coach, a counselor, and a behavioral specialist. There were five teachers each in grades K-2, four teachers in 3rd and 5th grades, and four teachers in 4th grade. There was one preschool teacher and 13 instructional and teaching assistants across the seven grade levels. Teachers in grades K-2 taught all subjects, while teachers in grades 3-5 taught two content areas and were departmentalized for the remaining subjects. At BES, 7% of teachers were in their

first or second year of teaching, whereas the median for new teachers in all elementary schools in Virginia was 9%. The majority of the teachers at BES were White, followed by teachers who were African-American. There were no teachers at BES who identified as Hispanic. Every teacher at BES (41) was invited to participate in this study. Some participants had been teaching more than 30 years and some participants were in their first or second year of teaching.

The data. The data for this study were exclusively from transcriptions of one-hour interviews conducted with each individual teacher. At the time of the interview, each participant was apprised of the purposes of the study and was provided with a copy of the IRB. After all questions had been answered, participants signed the IRB letter and the interview began. Voice recorders with an external microphone were used so that the voices would be clearly understood during the transcription phase. The same interview questions, specifically designed to answer the research questions, were asked of all of the participants. These questions served as prompts to collect the same types of data and to elicit descriptions about their evaluation experiences. The interviews left participants opportunities to describe other situations as they pertained to the evaluation experiences if they so wished. The interview questions can be found in Appendix B.

I realized that these teachers were very busy in the last semester of the school year and every precaution was taken so as not to impose on their time. The decision was made to only conduct one interview each. During recruitment, the participants were told that they would only be asked to give one hour of their time and would only be asked for more time if clarifications were needed. None of the interviews lasted more than the allotted time.

Phases of the Study

The study went through several phases and answered the proposed research questions. The phases were (a) recruitment of the participants, (b) interviews, (c) transcriptions, (d) coding, and (e) analysis. These phases structured the study and allowed the researcher to focus on each phase as the study developed.

Phase 1: Recruitment. The principal of BES was contacted in order to discuss the design and goals of the study. The central office staff members that were in charge of approving research for the school division were contacted about the study. All documents requested by the central office staff were submitted immediately so a decision could be made. They requested a copy of the IRB approval letter, the proposed methodology, a copy of the research questions, and the proposed interview questions. The central office staff members took approximately three weeks to approve the study.

After approval of the study, I was invited to a faculty meeting where the site administrator introduced me to the faculty. Both the teachers and the principal were assured of anonymity. All participants were informed that they would receive a transcript of their interview as soon as the interview phase was complete. All of the teachers from preschool through fifth grade were invited to participate in this study. The principal introduced me to the faculty and asked them to consider participating in the study before leaving the meeting. After I presented the goals and design of the study, each faculty member was given a piece of paper on which to provide contact information. Participants were instructed to turn in a blank sheet of paper if they did not wish to participate. All of the teachers left their paper in the same place as they left so as not to divulge which teachers were participants. When the papers were sorted later that day, 14 teachers had initially volunteered for the study.

The school's email system would not have been in the participant's best interest because of the risk of a breach in confidentiality. Therefore, the participants offered a phone number and/or another email address to use for contacting them. It was explained to each participant that there would be no material incentive for their participation. The interviews were set up within two days of recruitment and all of the interviews were complete after two weeks. The date, time and place of the interviews were chosen by the participants.

Each of the participants was contacted using the information provided during recruitment. After leaving three voicemails for one of the initial participants and receiving no response, it was assessed that this participant no longer wished to be contacted. As such, of the 41 potential recruits, 13 of the 41 teachers volunteered to participate. The 13 participants set up individual interviews at their chosen locations and interviews began four days after the initial faculty meeting. The participants were given the choice of location in order to further protect their identity.

The rationale for offering the opportunity to every teacher at BES was to have a population that worked with the same evaluator and had similar expectations placed upon them by this evaluator. Originally, the decision was made that any teacher beyond the first year of experience would be asked to participate. Later, the decision was changed in order to invite first year teachers to participate so that I could include their perspectives concerning the evaluation system. The demographic information about the participants is displayed in Table 3.

Table 3

Participant Demographics

Name	Experience Range	Current Assignment
Betty	15-19	classroom
Bobbi	20+	classroom
Ellen	5-9	specialist
Joan	1-4	specialist
Kylie	1-4	specialist
Laverne	10-14	specialist
Paris	10-14	classroom
Penelope	10-14	specialist
Rumer	20+	specialist
Scarlett	5-9	specialist
Sarah	1-4	specialist
Summer	20+	specialist

Phase 2: Interviews. The participants chose the date, time, and location of the interviews. Interviews were conducted in relatively quiet places with few distractions and a comfortable level of privacy. Places included libraries, coffee

shops, conference rooms, restaurants, and classrooms. The interview phase of the study was completed 11 days after recruitment. When interviews were conducted on the school campus, I met the participants in an agreed upon place so that it was unnecessary to visit the office. Almost one half of the 13 participants chose to be interviewed in their own classrooms.

The interview questions were designed in order for the participants to have an opportunity to describe their understandings of and perceptions about the teacher evaluation process. Interviews were conducted using an interview protocol with pre-determined prompts for the purpose of obtaining rich descriptions. The identities of the participants remained anonymous and each participant selected a pseudonym prior to the interview. To assure I protected the confidentiality of all of the participants, I changed the pseudonyms the participants chose to the names of actors and actresses found in a People magazine. These pseudonyms were used exclusively in reference to the participants in all forms of data. The identities, pseudonyms and all digital recordings were kept in separate locked filing cabinets.

At the beginning of each interview, participants were reminded of the goals of the study and were given the opportunity to ask questions. After all questions were answered, consent forms were signed and the voice recorders were set up. Participants were encouraged to speak openly about their experiences with the

evaluation process and were assured that the evaluator would not know about their participation.

The goal of a brief interview phase was to minimize participant conversations as much as possible. I wanted each participant to hear the interview questions for the first time in the context of their meeting. Interviews that lasted longer than one hour may have become tiresome to the participant and may have resulted in weak descriptions. Therefore, none of the interviews lasted longer than the promised one-hour time frame. Two voice recorders were used during the interviews. The use of two recorders minimized the chances of technology failures and ensured clear sound quality.

The ways in which the participants responded to the interview questions, including voice inflections, expressions, and other emotional reactions were as important as the words they used when they described their experiences (Glassner & Corzine, 1982; Moustakas, 1994). As each participant described understandings and perceptions about individual experiences, he/she was encouraged to talk about those experiences in ways that made sense. I did not share any personal information and my thoughts were focused on eliciting the participants' descriptions about their understandings and perceptions.

When the comments of participants led to discussions other than the proposed interview protocol, I remained flexible and allowed the participants to

continue as it related to the research questions. In a few instances, it was necessary to redirect participants when the discussions veered significantly from protocol. After one hour, interviews were closed. Later, during the transcription phase, it was only necessary to contact one participant for clarification. There were several instances in which time remained after all of the interview questions had been answered. In those instances, participants continued to share descriptions about their experiences. For an advanced organizer of the interview questions used, see Table 4.

Table 4

Ways Interview Questions Aligned With Research Questions

Research Questions	Interview Questions
Demographic Questions	In what school(s) and in which systems have you taught? Tell me about your years of teaching experience. Describe the various age groups with which you have worked. Please describe your teaching assignments. Include ways you have worked with colleagues such as co-teachers, special education teachers, and any experiences you have had with inclusion, or with paraprofessionals.
Research Question 1	Tell me what you know about the current evaluation system in your school division and how it was constructed. Describe for me any ways you personally had input into

<p>What are teacher understandings of the teacher evaluation process?</p>	<p>the design of the evaluation system. Describe how the current evaluation system works to evaluate teachers. I want to know how often you meet with your evaluator and what types of feedback you receive about the strengths and weaknesses in your practice.</p>
<p>Research Question 2</p> <p>What are teacher perceptions about their experiences with the teacher evaluation?</p>	<p>Please describe for me how your evaluator uses the evaluation system in your school. Describe for me what you know about the data that are collected during the evaluation process. What happens to these data and how does your evaluator use them?</p>
<p>Research Question 3</p> <p>In what ways do teachers believe that evaluation informs their practices?</p>	<p>Please describe for me how the evaluation system currently informs your practice. Do you believe it is doing the job it was designed to do or do you believe it could work better?</p>

In order to be a more focused listener, no field notes were taken during interviews. Reflective notes were made in a research journal which served as a secondary data source during the coding phase. Prior to the interviews, participants were advised about the recording of these notes.

Phase 3: Transcriptions. Taped interviews were downloaded to three technological devices to ensure that they could not be lost. A transcriber was hired to facilitate the process. After I received the transcriptions, each interview was

played several times to check for correctness and to add emphases where applicable. Thirteen interviews were transcribed during a two-week period to quickly facilitate the member checking phase of the work. There existed the possibility of changes once the participants read the transcripts. Transcribing the interviews soon after taping allowed me to more clearly remember verbal cues or emotions that were exhibited by the participants. The research journal was referenced many times.

Each participant was notified of the opportunity to acquire a copy of their raw transcript. Accompanying remarks reminded participants that the goal was to check for clarity of thoughts and expressions. Participants were given the opportunity to provide additional remarks. Five of the 13 participants chose to read their transcript. Four requested that the transcripts be sent electronically while one participant requested that the transcript be mailed to a home address. All requests were honored and participants were given a deadline so that possible changes could be made and coding could begin.

After reading the transcript, one teacher needed reassurance that responses would remain anonymous and confidential. Another teacher thought of additional comments and promptly provided those to me. One of the participants decided to withdraw from the study because she found it difficult to trust the process when her words were put into print. A brief discussion took place with the participant

concerning the rules of confidentiality, but her decision to withdraw was final. All contact information, data, and transcripts related to this participant were immediately removed from the data. As a result, all subsequent data analysis and interpretations were based on the perspectives of 12 teachers. Table 5 illustrates the percentage of classroom teachers vs. specialist teachers at this particular school participating in this study.

Table 5

Participant Statistics (Based on a Total of 41 Teachers)

Classroom Teachers (29)	Specialist Teachers (12)
3/41 (7%) faculty percentage	9/41 (22%) faculty percentage
3/29 (10%) classroom teacher percentage	9/12 (75%) specialist teacher percentage

Phase 4: Coding

When member checking was complete and all necessary changes had been made to the transcripts, the coding phase began. Every statement from every transcript was eventually coded. During the first reading, key words were highlighted. Notes were made in the margin during the second and third readings, but no codes were assigned. Transcripts were read multiple times and similarities were noted. After at least five readings, possible codes were added in the margins

of the transcripts. The primary codes emerged from multiple readings, but the individual interviews were not compared to one another during the early stages of analysis.

During the initial coding phase, transcriptions yielded 13 separate codes from 13 interviews. Using Post-It notes, possible primary codes were written on large envelopes with possible related codes organized underneath. Over the course of several days, 13 codes were narrowed to six primary codes with 13 secondary codes. The interviews were read once again and the six primary codes were assigned colors. During this reading, words and phrases were marked with the colors of the six primary codes.

After all primary codes were clearly marked the secondary codes were flagged with assigned colors. All of the secondary codes were clearly marked during another reading and the data were organized into more succinct groups. The transcripts were read until all codes were organized and the data was saturated. The data were marked with the pseudonyms and copies were made. The data were separated and placed in groups of similarity. This process took several days to complete for the 12 participants. When all of the data were organized, the findings were written in narrative form.

Phase 5: Analysis. After many days of putting all of the codes into raw narrative form, the analysis phase began. Duplicate material was removed and prominent codes were placed in the beginning. All of the findings were arranged in a logical order in narrative form. Many hours were spent reading the data to understand what was said by the participants. The biggest ideas were extrapolated so that possible implications could be drawn. This phase took several months to complete. Further clarifications of this procedure are found in the section that follows. The findings, implications, and possible future ideas for research are found in chapter 5.

Trustworthiness

During the study, many conscious decisions were made in order to maintain a sense of trustworthiness in research. During the interviews, the questions asked were written and approved by a committee of researchers. Every effort was made to ask open-ended questions specifically designed to encourage participants to share their experiences without leading them to any preconceived conclusions on the part of the researcher. Questions were asked in the same order to all participants. There were no personal comments added by the researcher.

A transcriptionist was hired early in the transcription phase. This person served in the capacity of an official court reporter and was trained in transcribing literal statements exactly as they were stated. This step was the first in

phenomenologically analyzing the interview data. Large right margins remained in the transcripts so that units of meaning as described by Hycner (1985) could be determined later from non-verbal and para-linguistic communication.

After the transcriptions were completed, I listened to the interviews and reread the transcriptions many times. The data was approached with a sense of openness to whatever meanings emerged. My presuppositions about the school and the participants were bracketed so that I used the participant's world-views and understood the meanings of what they said rather than what I expected them to say. I relied completely on what the participants said they meant instead of attaching my own meaning to what they said and I listened for a sense of the whole (Giorgi, 1975). The many readings provided a context for specific units of meaning and themes (Hycner, 1985). Intonations, emphases, and pauses were noted in a researcher journal, along with specific issues that arose. General impressions were recorded immediately after the interviews and these notes restrained my perceptions and bracketed my interpretations. The idea was to maintain the integrity of the participants' meanings as much as possible.

At this point, I began a process of going over every word, phrase, sentence, and paragraph and elicited participants' meanings from non-verbal communications. Staying as open as possible to the meanings, I did not address the research questions specifically. The essence of what the participants said was

extracted using as many literal words as possible. The participants each expressed unique and coherent meanings about their experiences irrespective of the research questions to which I sought answers. These very unique meanings were recorded in the margins.

When this process was completed, I addressed the research questions to the units of general meaning (Hycner, 1985) to determine whether or not what the participants had said responded to the research questions. Units of relevant meaning were assigned numbers that denoted which research questions were being answered. Redundancies were eliminated as the process continued and I relied on the number of times a meaning was mentioned. The number of times meanings were mentioned was significant because it indicated how important a particular issue was to the participants. Later, the frequencies of these meanings were displayed in a word cloud and units of relevant meaning were clustered together.

At this point I renewed my effort to bracket presuppositions (Hycner, 1975) and tried to stay as true to the phenomenon as possible. I determined that the units of relevant meaning clustered together naturally around six common themes. My committee chair repeated the process and came up with the same basic clusters. From the themes I wrote a summary about each of the themes. At this point, it would have been ideal to have returned to the participants and discussed the clusters. Since member checking had already been conducted, the literal words and

phrases were found to have been correct. However, it had already been decided that there would be no second interview; therefore I was unable to discuss the clusters with the participants and I modified the themes around the research questions.

In the final phase, great care was taken in order to protect the integrity of the units of meaning. Since I knew that the descriptions were not the experiences (Hycner, 1985), the best that I could do was to seek understanding through the phrases and words the participants chose to use. All of the participants described their experiences after some time had elapsed. [E] was in the fourth year at this school site and the majority of the participants had been teaching at this school for all of those four years. A disadvantage in describing past experiences was that time had elapsed for the participants. On the other hand, an advantage was that it possibly allowed for much fuller verbal descriptions since the participants reflected back on their unique experiences both consciously and verbally. From this point, key phrases and words were highlighted and marked with the pseudonyms of participants. Sticky notes indicated locations of similar meanings throughout the data. Finally, units of meaning were cut apart and sorted into groups of meaning which answered the research questions. These were put into narrative form and the research questions were answered.

Researcher Stance

I am a 23-year veteran in the field of education. The first ten years of my career were in elementary classrooms in the United States and abroad, while the last 13 years were in middle school mathematics (although I do not hold a mathematics endorsement). Both my bachelor's and master's degrees were in elementary education (N-8). I taught with the belief that course content was secondary to teaching the whole child; any graduate courses that I selected focused on ways that students learn.

Since 2003, I have been Nationally Board Certified. This certification was valid for a ten-year period, at which time I was successful in the renewal process. The National Board for Professional Teaching Standards deemed me worthy of that certification by judging my portfolio submissions in the area of Early Adolescence/Mathematics (ages 9-15). Soon after renewal in 2013, I realized there were conflicts between my educational beliefs and those of my site administrator. The message that there were inadequacies in my practice caused me to be dissatisfied with that teaching job.

During my time in the elementary and middle school classrooms, I came to value qualities in teachers that I felt were characteristic of true professionals. Sacrifices were made every day by teachers in my building to meet the needs of students. These teachers cared deeply about the success of students and this was a

priority for me, as well. Special education teachers and paraprofessionals were assigned to my classroom and our teaching was always a collaborative effort for the benefit of students. My classrooms were social endeavors where teaching and learning was uniquely constructed in context. It was also a noisy place as students strove to assist one another in the organization and mastery of learning goals.

Evaluations made me extremely nervous when I was a new teacher, but over time, I came to appreciate the value of receiving feedback from knowledgeable evaluators and colleagues. In most situations where I worked, teachers assisted each other in finding new and better ways to reach students. Most evaluators were effective instructional leaders who assisted teachers in making changes that enhanced the learning environment. I came to believe that the evaluation process was a collaborative effort between teachers and evaluators.

When the newest evaluation system began in Virginia (2011), conflict arose that challenged my beliefs about teaching. A new administrator, along with a new evaluation system changed the way I looked at my role as a teacher. I no longer had the assistance of an evaluator to enhance my practice. The feedback I received was superficial, condemnatory, and final. For me, the evaluation process turned into something that was done *to* me instead of *for* me and I no longer saw it as a collaborative effort. Some of the accomplishments I valued, such as serving as an adjunct faculty member at a local college, were not valued by the administration.

In fact, I was told to make a choice as to which job I wanted when a rescheduled conference day interfered with an afternoon college class.

Feeling confused, I stepped away from the K-12 classroom in pursuit of my doctorate degree in curriculum and instruction with an emphasis on teacher education. For the past three years I have focused on coursework and future plans for research. My desire was to focus on the evaluation process to discern what went wrong so suddenly and why I quickly felt as if I no longer could teach well. This study was the result of my interest in illuminating possible problems educators face with Virginia's newest evaluation system.

Chapter 4

Findings

The purpose of the study was to explore what teachers understood and perceived about the teacher evaluation system at one elementary school in Virginia. Participants in this study included three classroom teachers and 9 specialist teachers. Specialist teachers worked in specialty areas such as special education, foreign language, art, music, counseling, physical education, and library.

This chapter has been organized around the original research questions. Each research question is given, followed by the answers which came from the data. A summary was provided at the end of each section.

Research Question 1. What are teacher understandings of the teacher evaluation process?

Limited understandings. Participants were asked to describe their knowledge of the evaluation system. This school division's system was the same system outlined in the Virginia Department of Education's Guidelines (VDOE, 2011). None of the participants were aware that the evaluation system was developed and implemented by the VDOE, even though they had access to this information online. None of the teachers described any presentation of the evaluation system either to the teaching staff or individual teachers.

They were also unsure of the number of observations that needed to be completed by [E] during the school year. Two participants believed that the administration team collectively completed all final evaluations. Members of the administration team conducted observations and gave teachers feedback; however, only the principal was the evaluator of record and the participants indicated that feedback given by other members of the administration team was not considered during final evaluation meetings.

None of the participants were able to describe what happened to the data collected through the evaluation process nor were they able to list individuals who had access to that data. Two participants admitted they had never thought about it before: “I have no idea...I really don’t know. I never see them (the evaluations) again.” Two participants said their cumulative folder was available online and one recently asked [E] what happened to the data from evaluations, but did not receive an answer. All of the participants believed that the evaluation data was filed online for central office personnel and that anyone involved in administration for the school division had access. Unanimously, the participants were only concerned about low evaluation scores that might follow them to another school division should they ever decide to relocate. Participants believed that central office personnel often delayed teacher/specialist contracts until final evaluations were completed and their major concern was focused on their continued employment.

“All I cared about was seeing good scores,” said Penelope. Six participants admitted confusion as they realized their contracts were signed before their final evaluation meetings were completed.

None of the participants said that they had ever referred to an evaluation after it was submitted online. For instance, they said they saw their scores, signed the evaluations, and never referred to the comments to make changes in other years. When asked why, Kylie responded that the process was so emotionally charged that she was glad when it was completed for another year. All of the participants unanimously thought the teacher evaluation data was housed in the central office and also believed that the data would only be used if the administration wanted to dismiss teachers. One participant believed there were items in cumulative folders that administrative personnel wanted to keep private, even from its owner. “I knew four friends once who were curious about their evaluation data and they dropped by central office to see it. They were told that they had to make an appointment to see the contents.”

Participants believed they learned about the evaluation process through trial and error. “To this day, I do not understand this like in my previous school system. I have never had anyone to show me what is going to be looked at when I’m evaluated.” Not understanding clear expectations about the evaluation process caused participants to feel insecure about the process. They had a fear about what

they did not know and wanted someone to teach them more about the process and answer their questions. The various answers that were offered in this section showed a limited understanding about the evaluation system.

Categories and ratings. Within Virginia's teacher evaluation system (Guidelines, 2011), teachers are rated in seven distinct categories. The first six categories are weighted 10% of the final evaluation score, while the last category is weighted 40%. The categories are (a) professional knowledge, (b) instructional planning, (c) instructional delivery, (d) assessment of and for student learning, (e) learning environment, and (f) professionalism. Of the eight participants who mentioned the category weights, none of them were able to correctly identify the seven categories, even though all of them articulated various ways that the student academic progress category affected their evaluations. One participant believed "safety" was the most important part of a teacher's job and believed it should be weighted heavier than other categories. Safety was not one of the seven categories in the Guidelines (VDOE, 2011) although it may be interpreted as a part of Standard 5: Learning Environment.

Within the evaluation system, teachers were rated using a standard rubric. The correct rubric designations used to rate teacher performances were (1) Unacceptable, (2) Developing/Needs Improvement, (3) Proficient, and (4) Exemplary. None of the teachers used the correct verbiage for the designations.

Instead, Betty used “meets the standard” instead of Proficient while Penelope used “emerging” instead of Developing/Needs Improvement. Other examples of words and phrases used by participants included “meets expectations” and “exceeds expectations.”

Summary for research question 1. The participants in this study had a limited understanding of the evaluation process in their school division. They were not able to describe the process consistently nor were they able to articulate how the process was used in their school. The participants offered various answers for questions about the number of observations, the tenets of the instrument, or answers that described how often they were required to meet with [E]. They were also unable to list the seven categories of the evaluation system and the designations used on the rubric to judge their performances. Even though they were extremely worried and concerned about the processes, none of the participants had a strong understanding about it.

Research Question 2. What are teacher perceptions about their experiences with the teacher evaluation process?

Mistrust. A feeling of mistrust was reported among the participants at this school. The teachers either used the exact words or they used words that conveyed the same sentiment (e.g., “I’ve said this whole thing was rigged from the beginning,” and “How can I be a 4 at the end of one year and a 2 at the beginning

of the next?”). They described ways practices that made them distrust both the evaluation process and [E].

Along with the feelings of mistrust about the process and [E], participants believed that no one within the school division wanted to hear from them about how the evaluation system affected them until I approached them about the interviews. They were very anxious to express their feelings to me in private. “When you are outspoken, you can be assured that you are probably not going to have a good evaluation from principals,” said Bobbi. Sarah indicated that she believed she would feel more comfortable later on when it came to “disagreeing with [E]” about the results of her evaluation.

Eight of the participants expressed the perception that there were known “favorites” (i.e., favored teachers) with [E]. To take it one step further, these participants also perceived that these “favorites” were African-American teachers who were given more privileges than White teachers. They offered only two examples of the reported favoritism, but they felt that African-American teachers were treated differently overall. The examples they reported involved leaving the building to smoke or eat lunch during times when teachers were required to be on campus. A White participant responded that she was unaware of how African-American teachers were “getting away with it.” A feeling of mistrust led the participants to believe that some teachers were being given special treatment.

Participants were cautious when they spoke about [E] and often paused when they described their professional relationships. They hesitated to express what they considered to be complaints. Bobbi hesitated to discuss [E] and almost decided not to continue for fear of retribution if she was discovered. “I’ll probably regret that I’ve done this...Oh, boy, I probably shouldn’t have said that. I’m going to regret saying this, aren’t I?” At this point during the interview, the voice recorders were temporarily paused until she decided that she wanted to continue. Suddenly there was a sense that Bobbi did not trust the researcher to keep her identity confidential.

The participants showed mistrust for the school division’s administration and believed their jobs were made more difficult than was necessary. They perceived that their own school division created a fear of failure with the obvious pressure they were under and felt that this pressure to succeed stifled the creativity of teachers. “They say, ‘You better do it this way or you will be marked down.’ We are always on edge. They look at one day and it could have just been that we were having a bad day.” These teachers believed that a certain amount of fear was a normal part of their job and they did not trust their school division to act in their best interests.

Participants believed that ratings and feedback were subjective and that consequently, there was a lack of consistency. Paris valued [E]’s opinions. She

said, “I’m the type of person that, whatever they fill out, that’s what I feel about myself. I put a lot of importance on what they say.” Paris seemed shocked to report that [E] had given the veteran teacher on her team an “emerging” score in one category and described this as “insane.” She did not trust that [E] would rate all teachers fairly.

The participants had mistrust for [E] because they believed [E] did not fully understand the responsibilities of their job and therefore, could not have thoroughly evaluated their practices. Ellen reported that [E] suggested differentiation for a small group and she felt that was impossible and “a waste of time.” [E] offered no reason for the differentiation and Ellen believed it was because there was an indicator in one category that mentioned differentiation. “As far as what my students can do,” she said, “[E] doesn’t know anything. He has no idea what I do.” Sarah said, “I feel like [E] tries the best he can to understand what it is that I do, but doesn’t.” Bobbi said she tried to “gently” educate him on good practices in elementary school, but added that she never embarrassed him in front of others. “There were just some things he was wrong on. I’ve done this a long time. I know what I’m doing.”

Part of the sense of mistrust these participants had stemmed from the beliefs that [E]’s limited knowledge about their responsibilities negatively affected their evaluation scores. “He has become more knowledgeable,” said Kylie, “but I think

his knowledge of what I do in here is still limited. It makes a difference to my evaluation score every year.” All of the participants felt that more training for [E] was necessary, but they wondered aloud how any evaluator could possibly be trained in every content area on every grade level. The nine specialist teachers mentioned that they mistrusted [E] because there seemed to be an effort on the part of [E] to “check off” each of the indicators for the categories and they believed in doing so, they were unable to show examples of these indicators in their teaching because their jobs were different from those of classroom teachers.

In reflecting on the four years of implementation, all of the participants believed that [E] decreased the frequency of observations, resulting in an inconsistent evaluation schedule. “Especially the last couple of years, [E] spent very little time in here,” said Summer. “If he spent more time in here, I would feel better about my evaluations being true,” said Laverne. Another participant had only one observation and believed that the low number of observations contributed to poor evaluation scores. “[E] doesn’t know what I do,” Ellen said. She believed that if there were more observations there would have been more opportunities to point out specific areas for improvement.

Mistrust developed as the participants perceived that [E] had limited knowledge about their job responsibilities. They mentioned that either there were no observations at all during the year or that there were infrequent observations

prior to final meetings with [E]. Only two participants were involved in any follow-up meetings where discussions took place. “I’d rather you see what the kid does all year than just the SOL score. I just don’t think that is fair to me,” said Sarah.

Participants believed that the impact of student test scores on teacher evaluation scores made teachers anxious and afraid to “be creative” since they also believed that the evaluation instrument was subjective. They mistrusted [E]’s ability to use the instrument in a fair manner. Five of the participants mentioned that they believed they could lose their jobs over a poor evaluation score. “Some of us say, ‘what more could we have done for these students when they don’t pass?’” asked Penelope. “If we did everything we could for them (students), then we should be proud.” Sarah believed that the school division based the worth of its teachers on standardized test scores. Several participants believed that [E] used the evaluation scores as a form of intimidation and they mistrusted [E]’s ability to use it for their benefit. “Now whether they used them for that reason or not, I don’t know, but the threat was always there and that needed to stop,” said Kylie.

Participants believed that the pressure to “teach to a test” was unfair to teachers and caused them to mistrust the system. All of them discussed changes that they had seen in the evaluation procedures and pointed to the student academic progress section as the category that needed the most change. After the new

evaluation system was ushered in by the VDOE in April 2011, the student academic progress category was assigned a weight of 40% of the score while the other six categories were assigned a weight of only 10%. “I’ve seen teachers fudge stuff on in-house tests because they wanted to keep their jobs,” said Scarlett. “People got scared and did things because they were scared.” Scarlett clarified that she was referring to teachers who changed the scores of in-house tests for primary grades that were used for comparison purposes only and not the state-wide standardized test results.

One teacher developed a sense of mistrust when she was reassigned in the middle of the school year. A mid-year testing simulation showed stagnant progress among her students so [E] scattered her students throughout other classrooms and then reassigned her to another area for the remainder of the school year. She responded that she was not endorsed in the new area and was afraid of any other changes that [E] might suddenly make. Participants believed that a student’s SOL score was the least significant measure of students’ progress each year and they believed that [E]’s high rate of expectations especially for special education students was unrealistic.

Because of the belief that [E] had changing, unrealistic expectations, participants felt that the rapport between [E] and the faculty was adversely affected. They mistrusted [E]’s suggestions and believed that their opinions were

not valued by [E]. Overall, participants believed that some classroom teachers and specialist teachers were given several extra assignments in addition to their regular teaching responsibilities and others were not. Some participants said that they felt overwhelmed with the extra responsibilities and lacked the time to adequately prepare for their regular classroom duties.

For example, the participants described several committees to which they were assigned, while at the same time certain teachers were not required by [E] to serve on committees. They said they were “flattered” that [E] trusted them in leadership roles, but felt even more pressure to perform in their regular teaching assignments. Extra responsibilities added to what were already stressful situations. Betty was frustrated when she was forced to regularly work long hours at school because of these extra assignments and teachers without extra assignments were able to leave as soon as school was dismissed. A feeling of mistrust permeated the level of professional rapport the participants believed they had with [E] and they perceived these feelings of mistrust affected the way they felt about [E] during the evaluation process.

The participants desired to have good rapport with [E], but they were confused about their relationship and mistrusted his motives. They described situations in which they felt “used.” Participants wondered whether or not [E] genuinely trusted them to serve in leadership capacities or knew they would accept

the extra assignments. Four participants believed that [E] did not “like” them or their students and provided quotes that publicly supported these feelings. “[E] doesn’t like my class,” said Summer. “He said in public that we were a ‘thorn in his side.’” Summer believed that she was part of a group that [E] desired to “get rid of.” Scarlett said, “He runs teachers away. Teachers have left because they felt they were picked on.” These teachers and specialists knew the rapport they had with [E] was not the professional rapport that they desired, but they continued to believe that it was an important part of their jobs.

Of the participants who believed that rapport needed improvement also believed that [E] was a “critical person.” Rumer believed that no amount of work was ever good enough for [E] and this led to feelings of mistrust and anxiety concerning final evaluations. “Because [E] was so critical, I have been extremely concerned about the evaluation process. It’s only [E]’s opinion and that scares me.” Paris also described [E] as a “critical person” after a teacher was reprimanded when she tried to engage in conversation with [E] during an observation. The situation occurred in front of the students, though she believed that they were probably unaware of any problem. “You don’t have to wonder what [E] is thinking,” Joan remarked. Because of the perception that rapport was less than desirable, the participants mistrusted the results of final evaluations and believed that these evaluations were ineffective in stimulating professional growth.

Participants believed that [E] “overlooked” them when preparing observation schedules each year. Sarah said, “But I don’t want to say anything and maybe get [E] in trouble.” Paris described an observation for which there had been no follow-up meeting and she believed [E] simply procrastinated until it was too late. Rumer believed that [E] ignored her requests for additional help in understanding the expectations of data collection. She said, “I sent email after email: ‘Please discuss this with me.’ No response. That’s just how [E] is as an administrator.” Since they did not trust [E], the participants believed that the evaluation system worked against them.

All of the participants described evaluation meetings when [E] appeared to hide negative evaluation comments from their view until the final copy was emailed to them for their signature. They said they believed the poor evaluation scores were often used by [E] specifically to motivate them throughout the school year; however they felt it created a sense of mistrust between them. Since these participants knew that standardized test scores were the only data that mattered in the student academic progress category, they remained confused about which items describing their teaching performance deserved immediate attention. “It really doesn’t matter now since it is over,” said Ellen. “I know what he really thinks about me and I’m worried about next year.”

Summer was concerned that there were no meetings convened for the purposes of answering evaluation questions and Rumer believed that if she spoke up, [E] would be frustrated enough to retaliate later. Ellen described it in this way: “I think [E] would feel that I was being confrontational and had the perception that [E] was not conducting school business in the correct fashion. Do I risk being the target of frustrations on me in the future? I don’t feel like I should.” Ellen said she did not trust [E] to answer her questions correctly and had not been able to get anyone else to answer her questions about evaluation.

The two teachers who were new to BES believed that they had been drawn into uncomfortable discussions with [E] about other colleagues and they said they did not trust [E] to keep conversations confidential. They reasoned that if [E] talked to them about other teachers then they could not trust him not to discuss their personal issues with others. These two specialist teachers believed that confidential information about evaluation scores were shared with them by [E] inappropriately. For instance, these specialist teachers were informed about classroom teachers with whom they worked who had been placed on plans of improvements (POI). Kylie said, “That’s none of my business, but [E] just mentioned it to me in passing.” “I wished that [E] had not done that,” said Sarah. They later noticed that [E] did not follow up with these colleagues to support their

progress and they never told the colleagues what [E] said. “How can he know if changes are being made if they have not been observed since then?”

The participants believed that [E] showed favoritism to some classroom teachers and specialist teachers. They mentioned the subject in casual descriptions as if the meaning of favoritism in these contexts were understood. Participants offered specific examples of favoritism within the school and ways that they believed favoritism affected their evaluations. “The most frustrating thing last year was that [E] scheduled an observation with me and then left the group session I was leading,” said Laverne. “Some weeks later, I learned that [E] had questioned some teachers in the inner circle and actually wrote up an evaluation on me based on what those teachers said! I don’t trust him to be fair with us,” she said.

Joan believed the “favorites” who were mentioned told [E] their opinions about faculty members and those opinions became [E]’s perceptions of them. Scarlett believed the “favorites” had less stressful evaluation meetings because of the friendly rapport that was noticeable between [E] and those teachers. Summer believed the idea of “favorites” at BES was a widely known fact. Rumer said the evaluation system was used to reprimand a veteran teacher who was not among the “favorites.” Those classroom teachers and specialist teachers who did not believe that they were part of the “favorites” did not trust [E] to make decisions that benefitted them.

Three participants cautiously admitted that they believed they were included in [E]'s group of "favorites" because they felt they received an inordinate amount of encouragement from [E], though they were unsure of why. They believed favoritism impacted their performances because they received only positive feedback from [E]. "I feel there are things I could improve in, but [E] just tells me I am doing great," said Sarah. "This is not what I hear from the others on my team." Kylie felt that [E] appreciated the extra things she did to help parents and others in the school and therefore she was allowed some leniency on some of her classroom responsibilities for which she was evaluated.

The participants mistrusted [E]'s use of the evaluation system because of their beliefs that the evaluation process had been often affected by what they termed as "gossip." They even believed that [E] listened to the opinions of the "favorites" when it came to scoring evaluations. "The principal shouldn't listen to gossip," Paris said. "Some people just want to get attention and [E] gives it to them," said Joan. The participants that mentioned "gossip" all shared personal examples of how conversations between the faculty and [E] affected their overall evaluations. "We were told, 'Well, that's what someone said.' That is so not fair. Almost every year, [E] brings something up in my final meeting totally out of the blue that I don't expect," said Rumer. Participants mistrusted [E]'s subjective use

of the evaluation system because they believed [E] listened to gossip and believed that they were never evaluated in a fair manner because of it.

Three participants reported that there had been at least one situation in which [E] had written up a full evaluation without an observation. They were also bothered that detailed scores such as 3.4 and 2.7 were given in their evaluations. In the Guidelines, the instructions for the use of the rubric explained the use of numbers 1 through 4 when rating performances of teachers; however, evaluators may choose their own use of numbers. These participants viewed this practice as less than credible, but said they declined to ask [E] about the unusual scores in follow up meetings.

Participants believed that an important aspect of the job of evaluators was to gain deep understandings of the different responsibilities of classroom teachers and specialist teachers. All of the participants believed that this was vitally important in order for [E] to be able to make suggestions. They were concerned that [E] had a limited knowledge of the responsibilities of teachers in ways that prevented sound discussions about improvements. Participants offered an example as a way to show that limited knowledge did impact their jobs. Betty said [E] rationed paper in the primary grades when [E] should have known that those students needed more because of the nature of their developmental age. They believed that [E] ignored the specifics about student development and cited “teacher equality” as the reason

for the rule. These participants said since their objections were unheard, they bought the materials they needed for their classrooms even though they saw materials they needed being stored in the office. This caused them to mistrust his ability to make fair decisions.

Five participants said they did not trust [E] enough to ask questions or make suggestions about school matters. They said they were “torn” between following [E]’s mandates and making their own judgments when it came to their students. “[E] just has limited knowledge about what I do and I don’t find many of [E]’s suggestions to be helpful. It’s a struggle.” These participants attributed a lack of professional development on their part to inconsistent observation schedules and [E]’s limited knowledge about their individual teaching responsibilities. They also gave different answers to the same interview questions; an indication that they are not receiving consistent information about evaluations. Ellen brought some of her own observation notes and pointed out that, although the observation was written up by [E] as a 41-minute session, in actuality the observation only lasted 15 minutes. “Because of what I do and who I work with, I don’t do anything for 41 minutes. I don’t know where [E] got that number,” she said.

Examples such as these caused participants to mistrust [E]’s use of the system and viewed the process as “unfair.” Participants indicated multiple times that they preferred to be notified when [E] planned on conducting an observation.

They wanted brief conversations concerning the goals of the observations prior to [E]'s arrival. "I can't have everything in one lesson so if I knew what [E] was looking for, I could make sure I provided that snapshot. But with [E], I never know." The participants believed that [E] chose inconvenient times to observe, such as the day before Thanksgiving break and the first day after returning from spring break and did not trust his motives for doing so. They believed this was an unfair practice.

Participants said that they did not trust [E]'s use of the evaluation system because they felt many of the strategies were counter-productive to their professional development. They described the ways in which [E] completed evaluation scores prior to final meetings so that their participation was minimal. "[E] had it all typed up and already decided before I got there," said Penelope. Bobbi challenged the scores on one particular evaluation and she was given more favorable scores that reflected her participation in the meeting. "I sat down with my data notebook and pointed some things out and scores were changed from emerging to proficient," she said. All of the participants wanted [E] to provide opportunities for real conversations during the evaluation process. They described situations when they discovered improvement [E] wanted them to make, but there were very few school days left in which to implement any changes so they felt the opportunity for improvement was lost.

All of the participants believed that there was an unrealistic emphasis placed on data collection procedures for both classroom teachers and specialist teachers and they did not trust this practice to evaluate their contribution to student learning in a fair manner. They believed the stress and pressure associated with data collection affected their students as well. Participants admitted that student data collection was the most misunderstood part of the evaluation process. They described [E] as “unorganized” and believed this to be one of the reasons [E] was unable to thoroughly explain the data collection process to everyone. “It’s very last minute and not very organized. I believe [E] is going to be staying up late, thinking about something and just puts something down.” Scarlett said. “The evaluation is so subjective and is used by [E] sometimes to threaten the jobs of struggling teachers,” said Summer, “just because of how their kids are doing on the SOL test. They are being threatened now.”

Student classroom placements. There was also mistrust about student placements because ten participants believed student placements were made by [E] in an inequitable manner. These participants believed that student classroom placements were unequal among the teachers and believed that this had a direct impact on their final evaluation scores. These teachers believed that teachers who had a class made up of mostly struggling readers had less favorable evaluation experiences since the student academic progress category weighted standardized

test results 40% more than other categories. They maintained that this one factor was at least one reason why some teachers scored higher on final evaluations than others. Participants believed that teachers who began the year with strong academic students ultimately scored higher on final evaluations which they believed were, in part due to unfair student classroom placements. Participants believed that more equitable classroom placements would ensure that academically strong students were placed in each class.

These teachers wanted changes made to the student academic progress category of the evaluation because they felt student progress should be a reflection of individual student growth from multiple sources. However, at BES, the data for this category came from one source; that of standardized tests. The teachers were concerned about the comparisons made between students in the same grade levels. The category did not concern the participants as much as ways the data was handled. When [E] compared the students BES, he compared the scores as if all students were academic equivalents and they thought that was unfair to the students and unfair to the teachers. Betty said, "I've seen teachers who are doing a really good job, but it doesn't look like it, because their colleagues had a higher group to begin with." These ten participants did not trust the system because it was set up this way and maintained that the unequal student classroom placements set them up for unfair evaluations of both teachers and students.

One of the classroom teachers believed that the classroom placements gave some teachers an unfair advantage during evaluations, despite the fact that this particular teacher recently won an award for high academic student progress. She believed that the work was made “easier” for her because of student classroom placements and added that academically strong students “naturally possessed” work ethics that allowed for her to be more creative in the classroom. The other nine participants also believed that the demographics of classroom assignments had negatively impacted their final evaluations, too. Two of the participants did not mention student classroom placements. Penelope put it this way: “The most challenging students are going to misbehave more when you are having a formal observation. It could turn out to be a negative situation for the teacher.”

Paris worried about job security because of some of the students who had been assigned to her classroom. “I had a rough bunch my first year of teaching and the other teachers on my team had absolute angels. Something was out of balance there.” She believed student classroom placements were manipulated by [E] and, whether this was true or not, all of the participants believed it to be true. “For example, if they know a teacher who probably won’t be writing students up, they put behavior problems with them,” said Bobbi. “It’s just not right.”

Bobbi said that one school year [E] separated classes by gender. She described the meeting she had with [E] where she convinced [E] to reconsider.

These participants believed that student classroom placements impacted their evaluations in at least three ways: behavior, school attendance, and academic progress. They did not trust [E] to make fair classroom placements and perceived that [E] did this to favor some teachers. “They are not distributed evenly,” said Penelope, “but yet we are evaluated on that.” Participants believed they had no control over student classroom placements and already expressed anxiety about the upcoming school year.

Bobbi spoke at length about the “pay-for-performance” incentive previously instituted in the school division. During the year of the incentive, two teachers, one each from lower and upper grades, received cash bonuses for the top evaluation scores at BES. Ironically, one of those recipients was a participant who believed that incentive bonuses were distributed in an unfair manner. “Do I think it was fair?” asked Bobbi. “I do not. I do not. Every year you have a different clientele.”

There were other areas of mistrust surrounding the final evaluation scores. Sarah expected to be fired or to be placed on a plan of improvement if she received either a 1 or a 2 in any category. Summer believed that administrators in the central office considered the final evaluation scores before the annual contracts were renewed. Then it was discovered that the annual contract had already been renewed even though there had been no final evaluation meeting to date. She was obviously confused. “I must be OK then, right?” she asked. Laverne believed that contracts

were also based upon whether or not [E] wanted teachers to return to BES. Three participants indicated that they still had no knowledge as to whether they had a job for the upcoming year. One said, “I’m not going to be in trouble if I don’t get a job for next year, but I do know one thing. My kids have made a lot of progress since we’ve been together.”

Manipulated scores. There were also feelings of mistrust surrounding the topic of evaluation scores which many believed to have been manipulated. Summer said she was told by [E] that her continuing contract depended upon her evaluation scores. Sarah believed her work was scrutinized more closely because she worked with students in one of the grades in which a heavier emphasis was placed on SOL testing and said that part made her job particularly stressful. Paris did say that she believed the SOLs helped teachers keep the focus on all students; however, she did say that she thought these practices were unfair. Kylie voiced her concern to her central office supervisor and was told, “Don’t worry about it. Your evaluation scores don’t correlate with the job you’re doing anyway.” The participants noted that they believed the scores were often “made up” when the indicators were not aligned with their job. This was especially true among the specialist teachers who described the frustrations they felt when they were required to use the same evaluation instrument as classroom teachers.

All of the participants agreed that student test scores affected their own final evaluation scores somewhat. They believed that recent changes in the category weights had changed the focus of teacher evaluations. Most participants believed that the emphases placed on test results only occurred at the end of each school year. Ten participants believed that their jobs were dependent upon evaluation scores, which, in turn, were dependent upon student test scores. All participants believed that the use of student test scores was an unfair way to judge teacher performance and that the system was set up for their failure.

According to the participants, the sense of mistrust surrounding the student academic progress category stemmed from the subjectivity of the evaluation instrument. Within the last three years, this school division offered “pay-for-performance” incentives to its teachers. According to Bobbi, bonuses were awarded to teachers whose evaluation scores were in the top 2% of the teachers from each school in the division. Checks in the amount of \$2,500 were awarded to teachers with the highest evaluation scores, one check to the highest primary teacher and one to the highest upper elementary teacher. One of the participants in this study reported that she had received a bonus check during this time. The bonuses were available for only one school year and [E] was the evaluator of record during that time.

Several of the participants believed that the subjectivity of the evaluation process often produced tensions among the faculty. They felt that a sense of competition was exacerbated by the bonuses that were offered. Participants believed that [E] manipulated the evaluation scores so that the “favored” teachers received bonuses. All of the participants believed that the subjectivity of the evaluation system prevented them from receiving unbiased evaluations. One participant transferred into BES because of a personality problem with a former evaluator which eventually translated into poor evaluation scores. Since arriving at BES, this participant has received only positive evaluations and believed the evaluation process worked. “I think it (the evaluation system) can do the job it is supposed to do if it’s done right,” she said.

Another participant believed the evaluation system was biased from the beginning. “When this first came up,” Rumer said, “my perception was that it would be very biased because everyone has a different opinion.” Most all of the participants believed the system was subjective since they were only observed once during a school year. “How can you give me a grade for the whole year when you only observed me for 30 minutes?” All of the participants believed the perceived subjectivity of the evaluation system should be changed, but only two participants offered any suggestions about what could be done to minimize the subjectivity. They believed the evaluation has been used in order to “get rid of people.”

Participants wanted more observations, more feedback, and more conversations with [E].

Participants noted their desire to please their evaluator. One participant rearranged the classroom late one evening because the belief was that it would raise final evaluation scores. One participant believed that [E] had based at least one evaluation on prior knowledge of her work because a final evaluation had been filed with no prior observation. Many believed the evaluation system was “a joke” and should not be taken seriously. At least one participant said she comforted a colleague who was upset about an evaluation score and told her, “Take it with a grain of salt. [E] observed you for 20 minutes this year.”

Summer believed that evaluation scores were subjective and were often manipulated. It was her belief that [E] prepared the evaluations the day before final meetings and gave no consideration to a complete body of work. “Exactly what [E] wants changes from year to year and I’m pretty sure that there should be more consistency.” Another participant believed that scores were manipulated by [E] as a way to show the growth of a teacher. “[E] told me, ‘I marked you as proficient, a 3, on this and that, and by the end of the year you will be a 4.’” Participants responded that they believed their evaluation scores were based upon how [E] felt that day. They said that since they were unable to figure out how [E] arrived at

their evaluation scores they believed it was a waste of time to try and change unacceptable scores after the meeting.

Another participant believed that a good relationship with the evaluator was the reason she received more favorable scores than other teachers. She said, “There are teachers that don’t have the relationship that I do with him and they don’t have good results.” Another participant believed that inconsistent evaluation scores had adversely affected her health. “You’re bad and then you’re great again. This has made me sick and affected my self-esteem as an educator. At the drop of a hat my career could be gone.” This participant was a veteran teacher with years of experience in multiple school systems.

Teachers not yet on continuing contracts¹ believed they were scrutinized more closely than teachers on continuing contracts and one participant described how low evaluation scores could be altered if teachers challenged the low scores. Every participant who believed the evaluation was subjective also believed that evaluations should not be left to the opinion of only one person. Most participants believed that a team of evaluators should collectively share information about the work of teachers. “You almost need an outside person to do this,” one participant said. Participants believed that the evaluator’s perception of them affected their

¹ Teachers employed after completing the probationary period of 3 years shall be entitled to continuing contracts during good behavior and competent service (*Code*, 2013).

evaluation scores and those that were “chummy” with [E] “got away with not doing their job.”

Nine of the 12 participants were specialist teachers and they uniformly described feelings of manipulation when asked about the evaluation scores. Standard 7: Student Academic Progress focuses mainly on Virginia’s Standards of Learning and the ways in which teachers show their part in student learning. For specialist teachers, the indicators for Standard 7 do not align with the responsibilities of their jobs so they unanimously believed that scores were manipulated. Paris recalled: “Once [E] pointed out that I had not shown any of the indicators listed, but I was given a 3 anyway. [E] should have kept trying until [E] saw them.” “[E] could look all year and I would never be seen doing that. It’s not part of what I do to help these students,” said Sarah.

Participants believed that the subjectivity of the evaluation instrument affected their final scores in categories such as how they dressed (in the Professionalism category) and the number of resources they used for collecting student data (in the student academic progress category). “One person could love what you are doing and another person could hate it.” One teacher described an unannounced observation that lasted three minutes. She received a full write-up and evaluation afterwards, even though [E] had only been in her classroom 3 minutes. “How was that even fair?” she asked. She became emotional when she

described the shock and disappointment she felt afterwards. “No, this is not valid. There is no evidence there to back up the scores that I was given.”

Summer believed that [E] manipulated the scores of strong teachers as well, and perceived this to be an attempt to make the numbers show growth for all teachers. There was confusion when they described the use of decimals in the evaluation rubric and they believed this practice to be an unfair use of the system. According to the Guidelines, rubrics were designed to use 1, 2, 3, or 4 when rating teacher performance. They were confused when scores such as 2.7 and 3.1 were used and believed these decimals ratings had at one time or another been the cause of low scores and unsuccessful evaluations because they felt they should have been given a score of 3 instead of a score of 2.8. None of the participants ever asked about the decimal system and perceived it to be one of the changes evaluators were allowed to make.

Communication breakdowns. Both teachers and specialists were conflicted over their job expectations because they said they knew how to do their jobs well, but felt that [E] did not adequately communicate his expectations well. They believed that [E]’s perceptions about teacher management of student behavior solidly impacted the evaluation process overall. Participants believed that receiving specific feedback on a consistent basis would have made it easier for them to have taken risks in the classroom with regards to instruction. Two of the classroom

teachers spoke about the consistent feedback given by one of their former evaluators that had resulted in what they believed to have been significant professional growth among the faculty. Participants believed that receiving this type of feedback from their former evaluator was the catalyst for many faculty members to seek National Board Certification and was a process they believed was beneficial for students.

Participants said they wanted [E] to be more visible in the school and that he should never have attempted to observe two teachers at once because of time constraints, though they said they would never have mentioned this to [E]. Ellen said, “[E] came over to me for about five minutes and then walked away. [E] was obviously observing the other teacher across the room, but there is no way that [E] could have been hearing us at the same time.” This perception resulted in the belief that final evaluation scores assigned to them by [E] were not credible.

Most participants voiced a widely held belief that [E] spent significantly less time conducting observations than did other administrators in the building, but [E] was the evaluator of record. Participants believed that the other members of the administrative team were “more fair” and communicated much more openly with them than did [E]. “I got a great evaluation from others. [One] administrator is a very seasoned administrator and I appreciated that very much,” said Rumer. Rumer wanted more conversations throughout the year instead of getting most of her

feedback at the end of the year when it was too late to make changes prior to the evaluation meeting.

Bobbi said frequent walk-throughs were encouraging and she wished that there were more of these types of informal observations. Half of the participants wished that suggestions made as a result of formal observations had been more succinct. Participants offered multiple answers when asked about the required number of observations for both teachers and specialists. Four participants believed that the number of observations they received annually depended upon [E]'s personal assessment of their practice. Although there were only 15 days remaining in the school year, three participants had not been observed at all. One of them, Sarah, said, "I feel like there is not a difference in how long a teacher has been here, but it is in how [E] perceives you are doing. If you have been having a rocky year, the observations are more frequent in a sense and longer. If the teacher is doing a great job, it's quicker."

All teachers voiced confusion about the number of observations teachers could expect. Six participants said they felt that the number of observations were inconsistent and subject to change. Betty believed the number of observations was greater for teachers in SOL testing grades. Kylie believed that teachers on continuing contracts were required to have at least one formal evaluation while teachers who were yet to be on continuing contracts were required to have three.

Participants believed that the number of observations should have been consistent for all teachers and specialists and they believed that the inconsistencies adversely affected their evaluation scores.

Five participants believed that teachers and specialists were required to have at least two formal observations per year. Three participants believed that teachers with continuing contracts received one formal observation while those without continuing contracts received two. Paris and Scarlett believed that all teachers and specialists were required to have three formal observations, but were puzzled when they were asked about their observations in the current year. “Now that you mentioned it,” Sarah said, “I haven’t had any observations during this school year. I guess he is just going off of what I have done in the past.” One participant believed that all teachers and specialist had between six and eight informal observations per year while another believed the number was between eight and 12. Both of these participants taught in grades that placed a much heavier emphasis on SOL testing than other grades. “I don’t think that it’s because [E] doesn’t want to come by, but a lack of time,” said Joan. “I think it has to do with our location in the building,” said Kylie. Participants believed that if the number of observations had been consistent, at least one of the unknowns would have been eliminated and evaluations would have been more productive.

Several participants believed that [E] did not have a clear picture of their responsibilities since they were observed during the same class more than once. One participant said she preferred to have suggestions from more than one class (e.g. reading). Kylie believed that the observations were an “after-thought” since [E] made what she termed random appointments. “That morning [E] came up to me and asked, ‘When are you teaching today?’ and I said, ‘Just now.’ It was the day before Thanksgiving and I didn’t think that was the time to be doing an observation.” Penelope believed that [E] did not conduct more observations because of other job constraints. There was a breakdown of communication as to the time and frequency of observations.

Summer believed that inconsistent observations did not support struggling teachers and perpetuated the problem of poor academic progress by their students. “There needed to be more timely feedback that helped them or got them on the road out. It’s a vicious cycle,” she said. She believed that struggling teachers either left for another school or continued to have their contracts renewed at BES. “[E] did not do what he needed to do to ‘keep them from coming back,’” she said. Participants believed that there was a correlation between ineffective primary teachers and low student pass rates in the upper grades. They believed that when a struggling primary teacher was not removed with a consistent use of the evaluation system, student academic progress suffered. The belief was that when those same

students continued to get behind each year, the upper elementary teachers became frustrated because they were not able to make enough progress to be successful. They also believed teachers in the upper grades felt that low pass rates were reflected in their evaluation scores.

Two participants were concerned about how the evaluation was implemented and cited inconsistent observations, along with minimal feedback as two of their biggest concerns. However, when the subject was approached at a later time, the same participants had some positive things to add. They believed evaluations helped them to become better teachers. Both of these individuals have spent less than five years in the classroom and both of them noted that they have received only positive feedback. One believed the evaluation system assisted in reflective practices and the other believed the evaluation system was working for her. Both participants were observed once during the current school year, but had not participated in a final meeting with [E] as of the day of the interviews. "I have to go to [E] and ask, but I want some more feedback," said Sarah. One participant believed that [E] has been lenient on her evaluations this year, but added that next year things will probably be more stressful. Penelope noted, "At least the evaluation laid out what [E] was looking for. If only I felt it was fair." Still another participant said that the feedback had been "good" so far, but that her attitude might have been different if her scores had been less than acceptable.

Participants gave different responses to questions about the amount of feedback they received from [E]. Four participants said that they received frequent feedback in both written and verbal forms. A few participants responded that they did not receive feedback until the last meeting of the school year. “How does that help me if I don’t hear anything is wrong until the last week of school? What can I change?” she said. One participant believed that the frequent amount of feedback she received was due to the school’s Title I designation. Another participant attributed the amount of feedback she received to the efficiency of the administrative team as a whole. Even though she said she liked the encouragement she received from other members of the administration team, she was quick to admit that, in the end, [E]’s opinion was the only one that mattered. One participant said that she received feedback almost weekly until preparations began for the Standards of Learning tests. “[E] was very specific in the comments on my evaluation and I always learned something about myself from every meeting.” Participants said they received both verbal feedback during private meetings, as well as written feedback from [E].

Besides the participants who reported that they received frequent feedback, the other participants believed there should be more. They described the amount of feedback they received as “almost never” and “the end of the year only.” All participants believed more feedback would have been beneficial to them. Paris

said, “I’m sure we could use more feedback, but maybe there’s not time for that.”

Participants believed the feedback they received from other members of the administration team were very helpful because they included conversations in the follow-up meetings. Their reactions were that all of the administration team should be involved in evaluations instead of only one person. They believed the most helpful types of feedback came in the form of conversations. All of the participants described general feedback as subjective. “It’s all in the way it’s presented. If it is done with care and tact, it can be a very useful tool,” said Betty.

Nine of the participants responded that the only feedback they received was during formal meetings. “New teachers need confidence; you know?” one said. Final evaluation meetings at BES were reserved until the last two weeks of school and participants said they were concerned about this practice. Not only did they believe that this was not beneficial to their professional development, but they also wanted to hear from [E] during the year so changes could be made. Some participants said that the only types of feedback they regularly received were the evaluations that were emailed to them. “He doesn’t meet with me. We don’t even meet,” said one of the specialists. “He just sends it to me electronically and I just sign.” Participants said that they did not try to avoid feedback. On the contrary, participants believed that more feedback on a regular basis would be very beneficial to their practices.

Participants were clear in that they desired more observations and more feedback in order to show greater improvement in their teaching practices. They believed that brief observations they experienced brought a lack of credibility to their evaluations. One said, “I don’t want to even know when I am going to be observed because I’m not putting on a dog and pony show.” Most participants expressed a desire for frequent and unannounced observations along with weekly walk-throughs as part of their evaluation plan. “I’m sure that there should be a strict schedule for observations, but [E] just doesn’t pull that off.” Participants believed that frequent observations coupled with clear feedback and expectations would result in less anxiety and more teacher growth.

Participants wanted changes made to the format of evaluation meetings and they wanted feedback to occur more often than once per year. They also wanted to see more two-way conversations included in evaluation meetings. “I look for a time in the meetings for me to speak up but we never have those kinds of conversations.” Many participants expressed anxiety over not knowing what [E] would bring up in the final meetings. They wanted professional development in support of their practices and believed that regular conversations should have included other administrators, paraprofessionals, colleagues, and even parents. “If [a mentor] was a part of the process when I am evaluated, that would be a lot more helpful and I think [E] would learn some things, too.” Some participants wanted

feedback in the form of a narrative and others want to know expectations prior to observations.

All of the participants were confused about the number of observations they could expect. One participant noted that [E] had completed a midterm evaluation without an observation. She wondered if this was acceptable. “I don’t know if I am having an end-of-the-year observation or not. I wish people would make this clearer.” A similar situation confused another participant who had received a good evaluation without an observation. She noted that the objectives listed in each category were items that [E] could not possibly have observed and noted her confusion as to how she could have been effectively rated with such a process. “[E] should have kept on with me until he saw all of those things, but I don’t know why I was able to pass off on it.” Situations such as these made participants feel as if the evaluations were useless to their professional development.

New teachers held different beliefs than veteran teachers about the number of observations they could expect. Some participants believed that more observations occurred in the first five years of teaching while other participants believed that only one observation occurred each year if that observation was successful. Even though they were confused about the number of observations, all of the participants believed more observations would have been beneficial. “I’ve only been observed two times this year and I’m pretty sure that I’m supposed to be

observed more than that, especially since I am not tenured.” Some perceived that the number of observations was determined by [E]. Two participants believed the inconsistencies were attributed to the forgetfulness of [E]. “My central office supervisor said not to worry about [E]’s forgetfulness. We only have 15 days of school left.” It seemed difficult for the participants to set goals when the expectations of evaluation were elusive.

Emotions. The participants were afraid that their identities would be discovered as a consequence of their involvement in the study. All participants had last-minute questions about their involvement and needed reassurance about complete anonymity. Three participants preferred fewer observations because of personal anxiety. They believed they would have better observations if [E] listened to their lessons from the hallway.

Participants were confused about the school’s annual focus that was selected by [E]. They reported that many hours of preparation went into emphasizing the focus and making it relevant to their students. During final evaluation meetings, participants noticed that their evaluation scores reflected how well [E] perceived that they prepared students in the chosen area. Since they described their frustrations over frequent changes, they were confused over how to proceed when another focus was chosen mid-year. Some participants went so far as to say that [E] weighted the school focus heavier than other items among the evaluation

categories, even though this practice was not consistent with the goals of the evaluation.

Six participants used the word “frustration” throughout the interviews. Five were specifically frustrated with [E] because of uncertainties surrounding the types of data they were expected to collect from students. Since data collection was a major component within the teacher evaluation system, all of the teachers and specialists who used the word “frustration” did so in the context of final evaluation meetings. One participant said, “This year I’ve not had my final evaluation, so I don’t know what kind of data he is looking for. It’s frustrating.” Others expressed frustration over unanswered questions about data collection. “[E] kept saying, ‘You should know how to do this. You should know what to do.’ It’s very frustrating. Extremely frustrating.”

Participants indicated that they were frustrated with the timing of observations. Many of the participants said that with only 20 days left in the school year, they had experienced only one observation or they had not been observed at all. Participants who had been observed only once said that the observation was conducted during a review lesson for the Standards of Learning tests. They perceived a need for [E] to “get them done” and believed [E] was unable to watch them teach a typical lesson as a result. Since review lessons typically involved work sheets or games planned to aid student recall, participants believed that these

types of lessons could not have contained many important components of a good lesson.

Other participants said they were highly frustrated because their final meetings were conducted on the last day of the school year. “[E] didn’t get my evaluation done until the last day of the school year. By that time, it’s too late to change anything and it was a shock at how poorly I was rated on some things without any feedback.” She said she did not try to challenge any of the low scores she received on the final evaluation. Participants clarified that they wanted both observations and regular feedback, but felt that it was useless to receive feedback at the close of a school year. They felt as if these types of observations served no purpose other than to check off the results.

Participants were frustrated about an annual school focus that in their opinion changed much too frequently. Each year, [E] chose a school focus such as reading or writing. Each grade level was required to participate in the school focus and all teachers and specialists were to prepare lessons that supported the focus throughout the school year. Their frustrations were made manifest when [E] changed the focus mid-year. For example, a prior school focus was the comprehension of non-fictional texts. In November, the school focus was changed to writing and any preparations made for the previous focus area were no longer

applicable. Participants said changes like these occurred frequently and described their feelings as very frustrating.

Other sources of frustration involved the high expectations for standardized tests. Participants said that [E] expected all non-general education students to pass all of the end-of-the-year tests. “If my students could pass the tests every year, they wouldn’t be placed in my class. I’m lucky if one of my students passes a test.” Specialists who worked with non-general education students described the many ways they tried to assist student progress while they modified work on their level. They perceived that [E] expected all non-general education students to pass at the same rates as all general education students. Low pass rates impacted the specialists’ evaluation scores specifically in the student academic progress section. These frustrations caused participants to feel nervous when they met with the evaluator one-on-one.

Participants believed that teachers should have a larger voice in their own professional development and some said they were hurt that no one seemed to want to hear their thoughts on the matter. They believed that school officials should have wanted to hear more about specific accomplishments made throughout the year. “At first, I self-evaluated before the final meeting,” said one participant, “but in the end, [E] didn’t want to hear about it so I stopped wasting my time.” Most participants said they enjoyed talking about their class and wished more

conversation had been included as a part of the evaluation process. They said they tried to talk more about their classes but felt rushed in the evaluation meetings.

There were certainly frustrations among the participants, but there were many more emotions and feelings also. Although the interview questions did not specifically ask participants to describe their emotions and feelings, they were nevertheless prevalent in the responses. Emotions were noted in every context and were pervasive throughout the data. Therefore, it became impossible to separate emotional responses into distinct codes. Because the data were so heavily laced with emotions, it was necessary to double code the data in order to single out emotional language.

Overall the language of the participants about the evaluation process was predominantly negative in nature. Since the negative language was prevalent throughout the data, I decided to create a Word Cloud for each of the participants, followed by an aggregate Word Cloud to represent the entire group. A Word Cloud is a computer generated graphic (<https://tagul.com>) that displays the most frequently used words from interviews. My thought was that these displays could strengthen the trustworthiness of the study by showing each participant's most frequently used words.

As I began the process of generating the most frequently used words from the data, a pattern arose. Each of the specialist teachers described their experiences

using words common to their area. When the Word Clouds were generated, the identities of the participants were easy to discern. Therefore, in order to further protect the identities of the participants, 12 Word Clouds were aggregated and one Word Cloud was generated as a substitute for the whole. See Figure 1.

Figure 1

Word Cloud of Emotional Words Used by the Participants



Summary for research question 2. When the teachers were asked about their experiences with the teacher evaluation process, their perceptions were shaded with mistrust and emotions. The participants did not trust [E] to make student classroom placements in an equitable fashion and they believed this had a direct correlation to their evaluation scores. The teachers also did not place trust in

their evaluation scores because they described ways in which the scores were manipulated by [E]. There were communication breakdowns about job expectations, the frequency of observations, and how feedback would be given. Finally, many different emotions were prevalent throughout the data. A word cloud containing the most commonly used emotional phrases may be found in Appendix C.

Research Question 3. In what ways do teachers believe that evaluation informs their practices?

“It doesn’t.” All of the participants believed the evaluation process was “something that must be done” or a “formality we have to get through.” They believed that it could have been beneficial for everyone if it had been used correctly, although they had many different reasons why this evaluation system could never be used “fairly” for everyone. The classroom teachers (3) said that data collection was so time consuming that it took their focus off of other things they felt they should be spending time on such as lesson plans and collaboration. The specialist teachers mentioned numerous times about the ways they felt the evaluation instrument and their jobs were misaligned.

Even though they spoke about many topics during their one-hour interviews, both suggestions and emotions found in the data indicated that these teachers believed this evaluation instrument failed to inform their practices in the ways it

was designed to. Their feelings were that they were being evaluated on indicators that, many times were not naturally found in their classrooms, such as the differentiation of small groups for one of the specialists. The participants expressed their concern that they were being forced to standardize a process that was very personal for each of them. The specialist teachers all agreed that the evaluation process did not inform their practices at all. “As far as the evaluation, it makes certain I am doing what I am supposed to be doing, but it is not effective in helping me to improve my practice,” said Betty.

Useless. Only one participant believed that the evaluation system helped teachers to improve their practices through professional development and even this classroom teacher offered many suggestions she would like made to the evaluation process in the coming years. She was unable to give specific ways in which the process informed her practice. The remainder of the participants believed that the system was “useless” and described the system as “unable to be used by all educators.” Bobbi referred to the system as a “joke” and something that “had to be done.”

Participants who referred to the evaluation as subjective later described it as “fair” when they discussed a *positive* evaluation experience. Many were concerned that they had been evaluated by using a set of objectives that were not applicable. For example, Ellen described one of the standards that failed to align with her

specific responsibilities. Performance Standard 2: Instructional Planning had an indicator that read, “Plans for differentiated instruction.” As a specialist, she felt that it was impossible to differentiate in her situation but was told that she needed to differentiate or she would be “marked down.” “It just seems that sometimes evaluators are trying to put a square peg into a round hole and it frustrates me,” she said. “So I gave [E] the differentiation for my observation and I never did it again.”

The majority of the participants believed that the evaluation process was not an accurate measure of the jobs they did and did not feel that it was helpful to them. Many of them responded that they believed self-evaluations to be an effective tool in helping them to set goals for their practices. However, after spending hours on the self-evaluation process, they were told by [E] that those were unnecessary so they stopped the practice even though they felt it was beneficial. “I told [E] what I had done and that I thought it would be helpful to me,” said Bobbi, “and [E] just said, ‘We’ve got it right here. No need for that.’” Bobbi said she never self-evaluated again even though she believed it was one of the most beneficial parts of evaluation.

Most participants believed the evaluation system was just a formality. They believed that the emphasis on evaluation scores did not yield credibility with the faculty and the only year they believed it was different was the year bonuses were awarded. “It was required that [E] know what all teachers were doing back then

because central office would be paying closer attention. That gave it credibility.” Teachers believed that [E] invested more time into observations and suggestions during those evaluation meetings because both the central office staff and the superintendent would most likely be interested in which teachers received bonus checks and specifically what they had accomplished. Most participants believed that [E] spent very little preparation time on evaluations and believed the system was ineffective and was often used to “silence teachers.” Many believed their final meetings were purposely held during the last week of school so that they would be brief.

Participants believed that it was impossible for any system to effectively evaluate every teacher. For instance, all of the specialists expressed frustration because the evaluation seemed to use indicators that did not relate to their responsibilities. Specialist teachers were unanimous in the belief that Virginia’s evaluation system was better aligned with the responsibilities of a classroom teacher. One specialist worked with a member of the central office staff and created a more specific evaluation for her specialization. She currently operates under an evaluation system that contains indicators specific to her area of expertise.

Penelope pointed out that [E] suggested changes for her to make in her classroom. However, after she carefully considered the changes, she believed they

would not have benefitted her students because of their level of maturity. Penelope believed that [E]'s limited knowledge of her job expectations, coupled with the administration's insistence that all teachers and specialists be evaluated with the same instrument, were reasons for these suggestions. "I received a horrible evaluation because [E] did not know what I did. I knew these things were not relevant." Participants believed that there were many teachers in the building for whom the evaluation instrument did not easily apply and they believed this inability to generalize across the faculty ranks was a key reason that both teachers and specialists were frustrated with the system.

Participants believed the data collection part of this evaluation system changed what they considered to be the important things on which to focus. They believed that collecting data was both time-consuming and stressful, but knew that it was an important part of their final scores. The teachers said that the format of data collection was inconsistent. "We just figured it out on our own and then found out at the end of the year that we were doing it all wrong." Participants expressed the desire to have more training on the types and methods of data collection. Several of the participants preferred the portfolio method that was once used to show student progress prior to 2011. "I believed that portfolios showed more of me as a teacher," said Bobbi. All of the participants believed that the student academic

progress category was added by their local school division simply because the school division wanted the data for their own use.

The participants believed that the evaluation contained portions that either failed to apply to their particular jobs or believed that the evaluation itself was unable to assess their practices in a clear fashion. One participant said: “I just don’t think it has any bearing on what I do.” Two participants believed that the current evaluation system was not being used properly to either remove or support struggling teachers. Both of these participants described inconsistencies in the ways [E] dealt with the faculty. One said, “All I am asking is just to be evaluated fairly. There are not enough checks and balances with this evaluation and I have already left one school because of this.” Participants frequently used the words inconsistent and subjective when discussing the evaluation system.

Blame placing. Because these participants at this elementary school had limited knowledge of the evaluation system and had a sense of mistrust in dealing with the process, they placed the blame on other factors. They were (a) the evaluator, (b) the evaluation system, and (c) the school division. None of the teachers ever mentioned personal responsibility when placing blame and they did not have suggestions for rectifying the situation. They were frustrated and helpless to know what to do about the problems they saw with the evaluation system.

The evaluator. All of the participants mentioned some level of confusion when it came to job expectations and they placed the blame on [E]. Most of them believed that they were unable to ask [E] for clarification because [E] became frustrated when teachers questioned motives. They decided it was not worth the risk of “making waves.” Most participants felt [E] had unrealistic expectations of teachers and that these unrealistic expectations changed regularly during the school year. “We are never really sure about what is going on. It’s an awful feeling.”

Participants believed that [E] could have made more beneficial suggestions had [E] had a deeper knowledge base of content areas and grade levels. They believed that thorough observations were one of the keys to successful teacher evaluations and said they felt that they would have developed more professional skills if they had been observed more thoroughly by [E]. If they received a low score in one particular area, especially in an area they felt should have been rated much higher, they initially blamed [E]. Either they felt [E] was “unfair” or lacked the knowledge about their responsibilities to rate them properly.

Several participants cited [E]’s leadership qualities as some of the factors that adversely affected their evaluation scores and hindered their professional development. Several participants believed that [E]’s procrastination in situations such as scheduling observations and answering questions led to an overall sense of confusion within the faculty. They wanted answers to their questions sooner rather

than later so that they felt prepared for all of their responsibilities. They believed that [E]'s procrastination hindered them from being prepared. One such area of procrastination was concerning the approval of Smart Goals.

Each year, all teachers are to choose Smart Goals as personal targets for growth. Usually teachers are told from which areas the Smart Goals much come, but teachers are allowed to choose goals that best suit what they would like to accomplish. Rumer said, "[E] has to approve it (Smart Goal) but in the middle of the year last year [E] told me he didn't like mine and that his supervisor said I couldn't use it. That put me way behind. Of course, it didn't turn out well for me."

Several participants believed that both infrequent observations and last-minute follow-up meetings left them anxious and confused about the comments provided on their evaluation form. When Penelope challenged a low score she believed was unfair, [E] told her, "Oh, I was just up late last night. I'll fix it." Later she found that the low score was never adjusted in her favor. Penelope replied, "I just don't think [E] is putting the energy that he should be putting into it. It's a long evaluation."

Participants believed that an evaluator was to be a leader in finding "solutions" for problems that arose, but they did not believe that [E] was this type leader. Summer said, "To me, if you are going to be a leader, you should be able to lead people in finding solutions, or at least in helping to find someone who will

have the answers. That doesn't happen here." Most participants believed that they were left to their own interpretation of the evaluation's indicators, even though they voiced confusion. They said they approached their mentors and assistant administrators for more thorough explanations, but admitted that they were given incorrect information at times. All participants believed that unclear expectations by [E] made them feel insecure in their teaching responsibilities. One said it in this way: "We would like just a little more feedback on if we are doing this correctly, but [E] just says, 'You are a teacher. You know what to do.' But there is never any suggestion on how to fix it."

Participants felt the leadership was inconsistent and the expectations for teachers and specialists changed often. They said that each school year began with a specific focus in mind. Participants believed this was a good practice, but were frustrated when [E] returned from a conference and changed the focus completely. Many teachers described the work and preparation they had already made for the school focus, but [E] was not concerned about their efforts. "We were supposedly under-performing as a school and [E] went to a national conference. Instead of returning with specific things we could do as a faculty, it was mandated that all specialists have centers in their classrooms." One of the specialists was told that a new requirement mandated students to receive grades that began in mid-year. This participant said she failed to secure a meeting with [E] to answer questions about

this requirement but the failure of [E] to elaborate on the basis of the grades left her confused and highly frustrated. She said [E] was always busy and never found the time to clarify this requirement, even though she asked numerous times.

The evaluation system. Participants offered various explanations for why they felt the current evaluation system was “unfair.” 1) They believed a subjective evaluation hurt the self-esteem of teachers. 2) Specialists who worked with struggling students believed it was unfair to use student data because it was already a known fact these students functioned below grade level. 3) They believed that weighting the final category 40% was unfair because it reflected a false sense of what they had accomplished during the year and relied too heavily on student data. “The test scores are making us all crazy.”

All of the participants believed that assigning 40% to the student academic progress category was unfair and wanted changes made to this aspect of the evaluation. Instead, they believed this category should focus on a broad picture of each student’s annual progress, derived from multiple methods. Each participant expressed dissatisfaction with the evaluation system because of its strong ties with standardized tests. One said, “Students don’t come with the same experiences and the same background. It’s not fair to evaluate me on them.” Throughout the data, participants blamed the evaluation instrument when they had negative experiences with evaluation.

The school division. Participants believed that the school division was to blame for the fact that they had limited understandings about the evaluation system. They said that the school division implemented the current evaluation system before they provided thorough training to both teachers and evaluators. Joan believed the school division would have provided training during her first year of teaching at BES. When training was not offered, she believed her mentor or [E] would have provided the necessary training because she understood the evaluation to be an important part of her job. Again, training was not offered. Since proper training was not offered, Joan felt helpless as to how to prepare for evaluation.

Paris believed that the evaluation system had not been explained thoroughly to her, as well, and she said for several years she had been “guessing” what to do. The believed this breakdown to be the fault of the school division. A few of the specialist teachers said that their supervisor eventually answered all of their questions but they believed their veteran colleagues who had been through the process were most helpful. Betty believed that new teachers lacked the proper training to be successful in the evaluation process from the beginning and believed this issue was possibly one of the biggest reasons why new teachers leave the school division for other places. She felt that training should have been provided throughout the school year to all teachers and believed that [E] should have

periodically provided thorough explanations about the interpretation of the indicators. When participants expressed confusion over the expectations of the evaluation instrument, they believed that it was due to a lack of proper training by the school division. Two of the veteran teachers, Betty and Rumer were concerned about first-year teachers and their lack of training on the evaluation process.

Summary for research question 3. When teachers were asked for ways that the evaluation process informed their practices, they responded that it did not. Some said that it was a “joke” while others said it was “something we have to get through.” Participants had an overall belief that the system was useless in informing their practices. Sometimes teachers reported that they incorporated items into their lesson plans for an observation experience and then never used that strategy again once the observation had ended. Whenever these teachers were unsure of answers concerning the evaluation process, they chose to either blame the evaluator, the evaluation system, or the school division.

In the following chapter, a discussion of the significance of the findings and some concluding remarks will be made. Future research possibilities will be offered that extend the opportunities for studies in this area to change the ways that teachers and evaluators understand and use the evaluation system in Virginia.

Chapter 5

Discussion and Conclusions

Introduction

The purpose of this study was to engage one group of teachers in an elementary school in Virginia in a study of their knowledge about their teacher evaluation process, understand their perceptions about their experiences related to this process, and hear about ways teachers believed the evaluation process impacted their practices. This final chapter first summarizes the findings from this study. Next, conclusions are drawn with a critique of the study's recruitment, methodology, and interview protocol. Then, implications are made for the research base, specialists and the evaluator with suggestions for future research. Finally, the chapter ends with reflections in the epilogue.

Summary of Findings

The research questions were (a) what are teacher understandings of the teacher evaluation process? (b) what are teacher perceptions about their experiences with the teacher evaluation process? and (c) in what ways do teachers believe that evaluation informs their practices? In sum, the teachers in this elementary school had a limited understanding about Virginia's teacher evaluation system. Both classroom teachers and specialist teachers expressed anxiety and fear about the process, yet were unable to name the essential parts of the instrument

used. The teachers were suspicious of the system's ability to fairly and accurately evaluate their practices. They had the perception that this particular evaluation system was completely rigged against them. They thought the principal made unfair student classroom placements that favored some teachers and they thought that evaluation scores were manipulated. Because of these perceptions, these teachers failed to see the evaluation process as credible or valuable to their practices. The work of Milanowski and Kimball (2003) indicates the importance of regular feedback to a teacher's professional growth, but there was a significant breakdown in communication between the teachers and the evaluator in this school which fostered a sense of mistrust, rendering the process useless and making collaboration unlikely (Goldhammer, 1969; Cogan, 1972). The teachers found themselves in a self-described "no-win situation" and there was plenty of blame to go around. In frustration, they described a system over which they had no control (Lortie, 1975).

Conclusions

Three major conclusions were drawn based on the data collected and analyzed in this study. The conclusions focused on the empirical foundation for evaluation system and its validity for all teachers, the implementation processes, and the system's particular misalignment with the work of specialist teachers.

Inadequate Research Base

One conclusion is that the Guidelines (VDOE, 2011) were based on a foundation of research that is both inadequate and inappropriate for these elementary teachers. The research base did not provide the content or construct validity to support the evaluation system, especially for elementary and specialist teachers. The studies were quantitative in nature and the majority involved mathematics teachers in middle and high school settings. Middle and high schools look very different than elementary (Miller, 2015; Del Prete, 2010) and, given the unique responsibilities of elementary teachers (Millman & Darling-Hammond, 1990), the research that was used to implement this evaluation system did not take into account those unique responsibilities.

Problematic Implementation

Another conclusion is that this evaluation system was poorly implemented for the teachers in this elementary school as evidenced by the participants' perspectives. The evaluator observed and evaluated each of the 41 teachers in the school and used the same evaluation instrument for all of them. Observations were random and sporadic and left teachers with a sense of fear surrounding their jobs. Fear and inadequacy have been known to affect people negatively (Richardson & Placier, 2001; Sawyer, 2001) and these teachers were no exception. Under this system, they were unable to collaboratively formulate a plan with the evaluator

(Cogan, 1972; Goldhammer, 1969) and communicate in ways that would benefit their practices. The lack of feedback made the teachers unsure if expectations were met since they were often informed of their results when it was too late to make any meaningful changes. Darling-Hammond (2015) concluded that evaluation may create a punitive environment if it is not effectively implemented and these teachers were a prime example. Five years after the Virginia Department of Education pushed the evaluation system out to the schools (VDOE, 2011), not only had these teachers not realized any benefits, but they indicated that the system had prevented them from making any real changes.

Misalignment for Specialists

The third conclusion is that this evaluation system specifically failed to align with the jobs of specialists in this study. Since specialist teachers have unique responsibilities, the indicators could not possibly assess both classroom teachers and specialist teachers. In Standard 7: Student Academic Progress (Guidelines, 2011), teachers are required to collect student data which focuses on Virginia's Standards of Learning. While the teachers were supposed to develop their own tools for assessing progress in Music, Library, and similar specialist courses, they received little guidance assessment tools, the development of SMART goals (VDOE, 2011), and other strategic parts of the evaluation instrument.

The Virginia evaluation system was designed for use with all teachers in all content areas and settings. However, this study found that the evaluation instrument falls short in that endeavor because the indicators were created with classroom teachers in mind. There are no allowances made for a specialist teacher's specific content and most of the indicators would be observed infrequently by the evaluator, if ever. As the evaluator sought to use the same evaluation instrument with all 41 teachers in the school, the degree of misalignment became clear and its use only served to frustrate and confuse the specialist teachers.

Critique of the Study

As with any research project, this study did have limitations. Three major components of this investigation posed limitations: recruitment, interview protocol/instrumentation, and overall methodology.

Recruitment

All of the 41 teachers and specialists at BES were invited to participate in this study. While I was pleased that 12 of the 41 classroom teachers and specialist teachers ultimately participated, I originally thought that more classroom teachers would volunteer. Only after moving through the phases of the study was it apparent that the majority of the participants were specialists. Part of the methodology should have included plans to uncover participant demographics

prior to the interviews so that the chances of a more balanced population would have increased. The misalignment of the evaluation instrument was heavily prevalent throughout the data because nine of the 12 were specialist teachers.

Interview protocol/instrumentation

When the interview questions were designed, the most natural population was classroom teachers since my background is in elementary education. I failed to allow for the possibility that the majority of participants would be specialist teachers. Since I was unaware of the misalignment between the evaluation indicators and the jobs of specialist teachers, there was no consideration for this opportunity in the design of the questions. In future studies, plans will be made to modify the instrumentation based on nuances posed by the people who agree to participate in the research.

Overall methodology

During the phase of data analysis, there were other avenues of this phenomenon that I feel would have been beneficial to have explored. Since the methodology called for the use of identical questions only, I stayed on script and did not follow their comments to their full conclusions. Further, the participants were asked for only one hour of their time. Once the study began, I did not feel that I could break protocol and ask for more time or follow up interviews. Future

studies will make plans for multiple interviews based on the specific foci of the study.

Implications of the Study

Implications of the research base

The foundational studies of the research base provided by the VDOE lacked qualitative projects in elementary settings and the findings were generalized to the elementary context. The studies primarily focused on classroom teachers in middle or high school mathematics (one study was in science) and provided very little basis for generalizing to the elementary school. Since the teachers in the study felt the evaluation system was useless to them, the implications are that more research is needed concerning evaluation at the elementary site since there are challenges unique to elementary teachers. The implication is that elementary teachers have unique responsibilities and may need multiple ways to show their effectiveness. Future research that examines the construct validity of the research base and focuses on a system that aligns with the features of the work of everyone that is to be assessed is needed.

Implications for system implementation

The findings of this study suggested that these teachers needed to know the evaluation tool and the evaluation process in order to better communicate with the evaluator. The findings also suggested that this lack of communication plagued the

system's implementation so that benefits were not realized. The teachers needed effective feedback and consistent implementation from the evaluator (Milanowski & Kimball, 2003) so that plans could be made for more beneficial observations. When observations are consistent and equitable to all, teachers may be given empowered roles in the process.

The one-shot annual evaluation that was the practice in this elementary school was not effective in changing the practices of these teachers. This one evaluator made an effort to effectively evaluate all of the 41 teachers on staff and the findings from this study suggested that negative consequences occurred as a result. The implication is that the evaluator may have needed relief from some of the daily school responsibilities so that more time could have been devoted to the evaluation process.

The teachers in this study indicated that they desired more feedback on a consistent basis and they expressed the need for more frequent observations. A single evaluator who conducted three 30 minute observations in a school with 41 teachers and who spent two additional hours filling out forms and discussing the results with each teacher would need approximately 7 ½ weeks to devote only to teacher evaluation with no other responsibilities. Assistance was needed from new or existing personnel. The implication is that more evaluators were needed to assist with the work of evaluation in a more objective manner. How would additional

administrators be trained? Would the teachers view multiple evaluators as fair and equitable? Is it possible to create a fair and balanced system that allows consideration for every teacher's schedule and responsibilities?

The findings from this study also suggested that the current teacher evaluation system does not provide information or opportunity for improvement. Based on the teachers' experiences, this process does not fulfill its purpose to identify effective teachers, support struggling teachers, and provide professional development for all. In the research base, there were no studies that examined teacher practices that changed over time; yet the work of Fenstermacher and Berliner (1985) indicates that more professional development is needed for novice teachers to change their thinking and thus, to change actions. Their work implies that an efficient evaluation is one that takes into consideration a teacher's experience. Future research needs to focus on teacher learning over time in order to support teachers at any level of experience.

Implications for specialists

This study illuminated the misalignment of the evaluation instrument and the specific responsibilities of specialist teachers in this school. Evaluators in Virginia's schools are required to use a one-size-fits-all approach to teacher evaluation (VDOE, 2011) and the findings from this study indicated this approach was not working. School divisions and administrators need to be able to tailor the

evaluation process based on teachers' roles and responsibilities. It may be that elementary teachers and specifically, elementary specialist teachers require an evaluation instrument tailed to meet the needs of their situations in an effort to more fairly evaluate their effectiveness. No provisions were made in the Guidelines (VDOE, 2011) for specialist teachers and under the current policy they are evaluated using the same instrument as that of classroom teachers. The specialist teachers in this elementary school volunteered to participate in this study primarily to voice their concerns about a misaligned evaluation process and set of expectations. Future research needs to be conducted with elementary classroom teachers and specialist teachers to discover ways that may create a more aligned and fair process.

Implications for the evaluator

Finally, when the study was concluded I had many questions about and for the evaluator in this school. I am certain that he did not consciously choose to be a part of the angst, fear, and worry that these teachers displayed. There was obviously a reason why he made the decisions that he made and why those decisions impacted the evaluation practices in this elementary school. The methodology in this study did not include the evaluator as a participant. Future studies should include all participants in the teacher evaluation process. Further, it would be beneficial to study a group of evaluators and study how they conduct

teacher evaluation in their schools. Perhaps a study whose purpose it was to interview both teachers as well as evaluators would be the most critical.

Epilogue

At the onset, I shared the story of a teacher named Rose who left the profession after an unsuccessful evaluation experience. She just walked out the door and gave up. Rose told me that each year she always tried to do something new and different to challenge herself professionally and she had been living by this philosophy for years. Was it a coincidence that a new evaluation system judged Rose unfairly? Was evaluation so intensely personal for her that she could not accept the criticism that she received from her evaluator? As I thought about Rose's situation, I was reminded that I left, too. I did not leave my job over one administrator's opinion. Part of me knew that it was time to move back into academia and become a researcher. But another part of me knew that something was changing about the ways teachers were being evaluated and my feelings were that it would lead to no good end.

Looking back on the day I decided to leave the classroom, I realize how different I am now as I was then. The more questions I answer, the more questions I ask. But I view my chosen profession much differently from this side. Evaluators conduct evaluations in various ways and teachers view experiences through their own lens. There are other important events happening inside schools other than the

principal's opinions of which teachers dress like appropriately (Standard 6: Professionalism) and which teachers most know their content (Standard 1: Professional Knowledge).

In my own situation as a classroom teacher, I do not believe I would have been able to explain much about the evaluation system that either so I have adjusted my view of the teachers who participated in the study. There are breakdowns in communication in schools across the state. Being an educator is a very busy and complex process and we often consider evaluation as separate from the actual process of education. Improvements may certainly be made by everyone concerned, but if we continue to ignore this elephant in the room and allow it to discourage our best teachers, are our children really being served to the best of our abilities?

The teachers at BES said that evaluation was just "something I have to do." They clearly saw evaluation as something that was done *to* them instead of *for* them. They should have been a part of the development of the evaluation process, but they indicated that they were not. If it was impossible for them to have been a part of the development, they most likely could have had an active part in offering input during the system's creation. They could have been updated frequently on what was developing so that they would have had opportunities for greater understanding of the purposes and goals of the new process. Just as schools

promote that every child should receive what he or she needs to succeed, the same could be said for every teacher. Teachers need to receive the necessary support from administrators to successfully thrive as unique and valuable in an environment of ever-changing rules and methods.

This study raised a lot of questions. Did my colleagues thoroughly understand the new evaluation system and I was the only teacher who did not? Maybe if I had understood the system I would have made the adjustments and worked differently. Was my former evaluator properly trained to evaluate my practice? Maybe if my evaluator had understood some of the emotions that I personally had about the evaluation process he might have handled my situation differently. Was it the evaluation policies that were misaligned with my practice instead of vice versa? Has it been best practices to adopt this one-size-fits-all teacher evaluation from the beginning? Is it even possible to conduct evaluation practices as we know we should and create something unique that works for everyone?

More than likely, the evaluation instrument will continue to morph and change foci. Administrators will come and go, as will students. One thing will remain: stakeholders will want teachers to be accountable for the work that they do with America's children. Knowledge is power and having the knowledge to at least know what they are up against is probably the greatest single thing teachers can do

for themselves. The remainder of my career as a researcher and as a teacher educator will be spent searching for the answers to these questions and helping teachers one at a time navigate the evaluation process.

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.
- Allington, R. L., & Johnston, P. H. (2000). *What do we know about effective fourth-grade teachers and their classrooms?* Albany, NY: The National Research Center on English Learning & Achievement, State University of New York.
- Ambach v. Norwich, 441 U.S. 68, 78-79 (1979). Taken from Thomas, S. B., Cambron-McCabe, N. H., & McCarthy, M. M. (2009). *Public school law: Teachers' and students' rights* (6th ed.). Boston, MA: Pearson.
- Anderson, R. C., National Academy of Education. Commission on Reading, & National Institute of Education (U.S.). (1985). *Becoming a nation of readers: The report of the commission on reading*. Pittsburgh, PA: National Academy of Education.
- Au, W. (2007). High stakes testing and curricular control: A qualitative meta-synthesis. *Educational Researcher, 36*(5), 258–267.
doi:10.3102/0013189X07306523

- Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–73.
doi:10.1080/0969595980050102
- Callahan, R. E. (1962). *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools*. Chicago: University of Chicago Press.
- Cameron, C. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2008). Effects of classroom organization on letter-word reading in first grade. *Journal of School Psychology*, 46(2), 173–192. doi: 10.1016/j.jsp.2007.03.002
- Carlson, E., Lee, H., & Schroll, K. (2004). Identifying attributes of high quality special education teachers. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 27(4), 350–359. doi:10.1177/08884064040402700403
- Cavalluzzo, L. C. (2004). *Is national board certification an effective signal of teacher quality?* CNA Corporation.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York, NY: Macmillan.
- Code of Virginia. (2013). Sec. 22.1-304. *Superintendent's Memo #134-13*.
Retrieved from

www.doe.virginia.gov/administrators/superintendents_memos/2013/134-13.html Nov. 4, 2016.

Cogan, M. (1972). *Clinical supervision*. Boston, MA: Houghton Mifflin.

Covino, E. A., & Iwanicki, E. (1996). Experienced teachers: Their constructs of effective teaching. *Journal of Personnel Evaluation in Education*, 10(4), 325–363.

Crabtree, B. F., & DiCicco-Bloom, B. (2006). The qualitative research interview. *Medical Education*, 40(4), 314–318.

Craig, J., & Cairo, L. (2005). *Assessing the relationship between questioning and understanding to improve learning and thinking (QUILT) and student achievement in mathematics: A pilot study*. Charleston, WV: Appalachia Educational Laboratory at Evantia.

Creswell, J. W. (2011). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. (4th ed.). Pearson.

Crump, M. J. C., Vaquero, J. M. M., & Milliken, B. (2008). Context-specific learning and control: The roles of awareness, task relevance, and relative salience. *Consciousness and Cognition*, 17(1), 22-36. doi: 10.1016/j.concog.2007.01.004

- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (2015). *Getting teacher evaluation right: What really matters for effectiveness and improvement?* New York, NY: Teachers College, Columbia University.
- Darling-Hammond, L. (Ed.). (2000). *Studies of excellence in teacher education*. New York, NY: American Association of Colleges for Teacher Education.
- Darling-Hammond, L., & Ball, D. L. (1998). *Teaching for high standards: What policymakers need to know and be able to do*. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania, co-published with National Commission on Teaching and America's Future.
- Darling-Hammond, L. (1997). Toward what end? The evaluation of student learning for the improvement of teaching. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.
- Del Prete, T. (2010). *Improving the odds: Developing powerful teaching practice and a culture of learning in urban high schools*. New York: Teachers College Press.

- Dunn, R., Honigsfeld, A., Doolan, L. S., Bostrom, L., Russo, K., Schiering, M. S., ... Tenedero, H. (2009). Impact of learning-style instructional strategies on students' achievement and attitudes: Perceptions of educators in diverse institutions. *Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, 82(3), 135–140. doi: 10.3200/TCHS.82.3.135-140
- Ellett, C. D., & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17(1), 101–128.
- Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, 36(2), 103–112. doi: 10.1207/S15326985EP3602_5
- Felder, R. M., & Brent, R. (2004). How to evaluate teaching. *Chemical Engineering Education*, 38(3), 200–215.
- Fenstermacher, G. D., & Berliner, D. C. (1985). Determining the value of staff development. *The Elementary School Journal*, 85(3), 281-314. doi: 10.1086/461407
- Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, 7(9). Retrieved from <http://epaa.asu.edu/ojs/article/view/544>

- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal*, 21(2), 449–460. doi:10.3102/00028312021002449
- Giorgi, A. (1989). One type of analysis of descriptive data: Procedures involved in following a scientific phenomenological method. *Methods*, 1(3), 39–61.
- Glanz, J. (1991). *Bureaucracy and professionalism: The evolution of public school supervision*. Rutherford, NJ: Associated University Press.
- Glanz, J. (1994). History of educational supervision: Proposal and prospects. Paper presentation, Annual Meeting of COPIS, Chicago, IL. Mar. 18, 1994. ERIC ED369178. Retrieved from <http://files.eric.ed.gov/fulltext/ED369178.pdf>
- Glanz, J., & Sullivan, S. (2005). *Supervision that improves teaching: Strategies and techniques* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Glassner, B., & Corzine, J. (1982). Library research as fieldwork: A strategy for qualitative content analysis. *Sociology and Social Research*, 66(3), 305.
- Goldhaber, D., & Anthony, E. (2005). *Can teacher quality be effectively assessed? National board certification as a signal of effective teaching*. NJ: Urban Institute.
- Goldhaber, D. D., & Brewer, D. J. (1997). Evaluating the effects of teacher degree level on educational performance. In W. J. Fowler (Ed.), *Developments in*

- school finance, 1996* (pp. 197–210). Washington, DC: National Center for Educational Statistics, U.S. Department of Education.
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129–146.
doi:10.3102/01623737022002129
- Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York, NY: Holt, Rinehart & Winston.
- Green, J. (2011). *Education, professionalism and the quest for accountability: Hitting the target but missing the point*. New York, NY: Routledge.
doi:P10.4324/9780203832561
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82. doi:10.1177/1525822X05279903.
- Guskey, T. R. (2007). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice, 26*(1), 19–27. doi:10.1111/j.1745-3992.2007.00085.x
- Hamilton, L., & Stecher, B. (2004). Responding effectively to test-based accountability. *Phi Delta Kappan, 85*(8), 578–583.

- Hamilton, A., Madison, J., & Jay, J. (1818). *The federalist, on the new constitution, written in the year 1788, by mr. Hamilton, mr. Madison, and mr. jay: With an appendix, containing the letters of pacificus and helvidius, on the proclamation of neutrality of 1793; also the original articles of confederation, and the constitution of the united states, with the amendments made thereto* (A new ed.). Washington: J. Gideon, june.
- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*(5), 949–967.
- Hanushek, E. A., Kain, J. F., O’Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality*. Cambridge, MA: National Bureau of Economic Research. Retrieved September 6, 2015 from <http://www.nber.org/papers/w11154.pdf>
- Hattie, J. (2003). *Teachers make a difference: What is the research evidence?* Retrieved September 12, 2014 from http://www.leadspage.govt.nz/leadership/pdf/john_hattie.pdf
- Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: Does it matter? *Journal of Teacher Education, 36*(3), 13–15.

- Haycock, K. (2000). *Thinking K-16 report: Good teaching matters*. [Online]. Washington, DC: The Education Trust. Available from <http://www.edtrust.org>
- Hess, F. M. (2009). The new stupid. *Educational Leadership*, 66(4) 12–17.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371–406.
- Hunter, M. (1982). *Mastery teaching*. Thousand Oaks, CA: Corwin Press.
- Husserl, E. (1970). *The Crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy*. Evanston, IL: Northwestern University Press.
- Hycner, R. (1985). Some guidelines for the phenomenological analysis of interview data. *Human Studies*, 8(1) 279–303.
- Interstate Teacher Assessment and Support Consortium. (1992). *Model standards for beginning teacher licensing and development: A resource for state dialogue*. Washington, DC: Council for Chief State School Officers.
- 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.

- Johnson, B. L. (1997). An organizational analysis of multiple perspectives of effective teaching: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 69–87. doi:10.1023/A:1007951321381
- Jones, B. D., & Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high stakes testing. *Education Policy Analysis Archives*, 12(39).
- Kane, T. J., Kerr, K. A., Pianta, R. C., & Measures of Effective Teaching Project. (2014). *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (1st ed.). San Francisco, CA: Jossey-Bass, a Wiley Brand.
- Kaplan, L., & Owings, W. (2002). The politics of teacher quality: Implications for principals. *National association of secondary school principals bulletin*, 86(633), 22–41.
- Kelly, S. (2012). *Assessing teacher quality: Understanding teacher effects on instruction and achievement*. New York: Teachers College Press.
- Kunter, M., Baumert, J., & Koller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17(5), 494–509. doi: 10.1016/j.learninstruc.2007.09.002

- Laine, S. W. M., Behrstock-Sherratt, E., & Lasagna, M. (2011; 2010;). *Improving teacher quality: A guide for education leaders* (1st ed.). San Francisco: Jossey-Bass.
- Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*, 15(2), 129–151.
- Lortie, D. C. (1975). *Schoolteacher: A sociological study* (2nd ed.). Chicago, IL: University of Chicago Press.
- Lovelace, M. K. (2005). Meta-analysis of experimental research based on the Dunn and Dunn model. *Journal of Educational Research*, 98(3), 176–183.
- Markow, D., Macia, L., & Lee, H. (2013). *The MetLife survey of the American teacher: Challenges for school leadership*. New York, NY: MetLife.
Retrieved from <https://www.metlife.com/assets/cao/foundation/MetLife-Teacher-Survey-2012.pdf>
- Marzano, R. (2004). *Building background knowledge for academic achievement: Research on what works in schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McDonald, F. J., & Elias, P. (1976). Executive summary report: Beginning teacher evaluation study, Phase II (PR-76-18). Princeton, NJ: Educational Testing Service.

Merton, R. (1995). The Thomas theorem and the Matthew effect. *Social forces*, 79(2), 379–422.

Milanowski, A., & Kimball, S. M. (2003). *The framework-based teacher performance assessment systems in Cincinnati and Washoe* (No. TC-03-07). Madison, WI: Consortium for Policy Research in Education.

Miller, W. B. (2015). Middle schools occupy unique space in education system. *Daily Gleaner*, Nov. 2, 2015.

Millman, J., & Darling-Hammond, L. (1990; 1989). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Sage Publications.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125–145.

Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.

National Board for Professional Teaching Standards. (1987; 2002). *What teachers should know and be able to do*. Arlington, VA: Author.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Department of Education.

- National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. New York, NY: NCTAF.
- Oliva, P., & Pawlas, G. (2004). *Supervision for today's schools*. NJ: Wiley Jossey-Bass Education
- Oplatka, I. (2011). Darling-hammond, L., Meyerson, D., la Pointe, M., & Orr, M. T. (2010). Preparing principals for a changing world: Lessons from effective school leadership programs. San Francisco: Jossey-bass, *Leadership and Policy in Schools*, 10(2), 246-249. doi: 10.1080/15700763.2011.557520
- Pajak, E. (2000). *Approaches to clinical supervision: Alternatives for improving instruction* (2nd ed.). Norwood, MA: Christopher Gordon.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of educational research*, 62(3), 307–332.
- Plummer, K. (1983). *Documents of life: An introduction to the problems and literature of a humanistic method*. London: Unwin Hyman.
- Pogodzinski, B., Umpstead, R., & Witt, J. (2015). Teacher evaluation reform implementation and labor relations. *Journal of Education Policy*, 30(4), 540–561.
- Pressley, M., Raphael, L., Gallagher, J. D., & DiBella, J. (2004). Providence-St. Mel School: How a school that works for African American students works.

- Journal of Educational Psychology*, 96(2), 216–235. doi:10.1037/0022-0663.96.2.216
- Quinn, R. E. (2014). *The best teacher in you: How to accelerate learning and change lives*. San Francisco: Berrett-Koehler Publishers.
- Raju Pamu, M. (2010). Early career teachers' quit intentions: Implications for teacher education. *International Journal of Educational Management*, 24(6), 478-491. doi:10.1108/09513541011067665
- Reynolds, S. (2011). Gaps in the system. *Complicity: An International Journal of Complexity and Education*, 8(1), 28–31.
- Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.
- Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.). Washington, DC: American Educational Research Association.
- Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on student achievement. *Sociology of Education*, 70(4), 256–284.
- Sanders, W. L., & Horn, S. (1994). The Tennessee value-added assessment system (TVASS): Mixed- model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299–311.

- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Sawyer, L. (2001). Revamping a teacher evaluation system. *Educational Leadership, 58*(5), 44–47.
- Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.
- Schroeder, C. M., Scott, T. P., Tolson, H., Huane, T., & Lee, Y. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44*(10), 1436–1460. doi:10.1002/tea.20212
- Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York, NY: Crown Business Publishing.
- Seldin, P. (1999). *Changing practice in evaluating teaching*. University of Michigan: Wiley.
- Smith, T., Gordon, B., Colby, S., & Wang, J. (2005). *An examination of the relationship of the depth of student learning and National Board certification status*. Office for Research on Teaching, Appalachian State University.

- Stanley, L., & Wise, S. (1993). *Breaking out again: Feminist ontology and epistemology*. London: Routledge.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795–819. doi:10.1002/pits.20113
- Stronge, J. (2010). *Evaluating what good teachers do: Eight research-based standards for assessing teacher excellence*. Larchmont, NY: Eye On Education.
- Stronge, J. H., Little, C., & Grant, L. W. (2008). *Gifted teachers for gifted students: An international comparative study of national award-winning teachers*. Asia-Pacific Conference on Gifted. Singapore.
- Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2008). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education, 20*(3), 165–184. doi:10.1007/s11092-008-9053-z
- Stronge, J. (2007). *Qualities of effective teachers* (2nd ed.). Alexandria, VA: Association of Supervision and Curriculum Development.
- Stronge, J. H. (2006; 2005). *Evaluating teaching: A guide to current thinking and best practice* (2nd ed.). Thousand Oaks, Calif: Corwin Press.

- Stiggins, R., & Duke, D. (1988) *The case for commitment to teacher growth: Research on teacher evaluation*. New York, NY: State University of New York Press.
- Taylor, E. S., Tyler, J. H., & Research, National Bureau of Economic. (2011). Effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers. SI: National Bureau of Economic Research.
- Tomlinson, C. A. (2007). Learning to love assessment. *Educational Leadership*, 65(4), 8–13.
- Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tucker, P. D., Stronge, J. H., & Gareis, C. R. (2002). *Handbook on teacher portfolios for evaluation and professional development*. Larchmont, NY: Eye on Education.
- Tursman, C. (1981). *Good Teachers: What to Look For. Education USA Special Report*. Arlington, VA: National School Public Relations Association.
- United States. Congress (107th, 1st session. (2001). *No child left behind act of 2001: Conference report to accompany H.R. 1*. Washington: U.S. G.P.O.

- United States Department of Education. (2009). *American recovery and reinvestment act: Executive summary*. Washington, DC: Author.
- United States Department of Education. (2009). *Race to the top program: Executive summary*. Washington, DC: Author.
- United States Department of Education. (2007). *Strategic Plan: Executive summary*. Washington, DC: Author.
- United States. Office of Education. (1965). Title I, elementary and secondary education act of 1965: Report. *Title I, Elementary and Secondary Education Act of 1965*.
- Vandevort, L. G., Amrein-Beardsley, A., & Berliner, D. (2004, Sept. 8). *National Board certified teachers and their students' achievement*. Educational Policy Analysis Archives [Online]. Available from <http://epaa.asu.edu/epaa/v12n46>
- Virginia State Department of Education. (2011). *Guidelines for uniform performance standards and evaluation criteria for teachers*. Retrieved July 27, 2015 from http://www.doe.virginia.gov/teaching/performance_evaluation/
- Virginia State Department of Education. (2011). *Research base for uniform performance standards and evaluation criteria for teachers*. Retrieved July 27, 2015 from http://www.doe.virginia.gov/teaching/performance_evaluation/

- Vogler, K. E. (2002). The impact of high-stakes, state-mandated student performance assessment on teachers' instructional practices. *Education, 123*(1), 39–56.
- Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19–27.
- Walls, R. T., Nardi, A. H., von Minden, A. M., & Hoffman, N. (2002). The characteristics of effective and ineffective teachers. *Teacher Education Quarterly, 29*(1), 39–48.
- Walsh, K. (2001). *Teacher certification revisited: Stumbling for quality*. Baltimore: Abell Foundation. Retrieved October 15, 2015 from <http://www.abellfoundation.org>
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1994). Synthesis of research: What helps students learn? *Educational Leadership, 51*(4), 74–79.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Retrieved September 16, 2015 from <http://tntp.org/publications/view/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness>

- Weiss, I. R., & Miller, B. (2006). *Deepening teacher content knowledge for teaching: A review of the evidence*. Retrieved August 12, 2014 from http://hub.mspnet.org/media/data/WeissMiller.pdf?media_000000002247.pdf
- Weiss, R. S. (1994). *Learning from strangers: The art and method of qualitative interview studies*. New York: Toronto: Free Press.
- Wenglisky, H. (2000). *How teaching matters: Bringing the classroom back into discussion of teacher quality*. Princeton, NJ: Millikan Family Foundation and Educational Testing Service.
- Wilcox, D. D. (2000). The National Board for Professional Teaching Standards: Can it live up to its promise? In M. Kanstoroom & C. E. Finn, Jr. (eds.), *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57–67.
- Yeh, C., & Inman, A. (2007). Qualitative data analysis and interpretation in counseling psychology: Strategies for best practices. *The Counseling Psychologist, 35*(3), 369–403.

Yoon, K. S., Duncan, T., Lee, S. W., Scarloss, B., & Shapley, K. L. (2007).

Reviewing the evidence on how teacher professional development affects student achievement. Washington, DC: Regional Educational Laboratory Southwest.

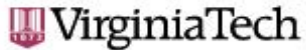
Zacharias, N. T. (2007). Teacher and student attitudes toward teacher feedback.

RELC Journal: A Journal of Language Teaching and Research, 38(1), 38–52. doi:10.1177/0033688206076157.

Appendices

Appendix A

Internal Review Board Approval



Office of Research Compliance
 Institutional Review Board
 North End Center, Suite 4120, Virginia Tech
 300 Turner Street NW
 Blacksburg, Virginia 24061
 540/231-4806 Fax 540/231-0959
 email irb@ut.edu
 website <http://www.irb.ut.edu>

MEMORANDUM

DATE: January 27, 2016
TO: Mary Alice Barksdale, Elizabeth Jean Gilpin
FROM: Virginia Tech Institutional Review Board (FWA00000572, expires July 29, 2020)
PROTOCOL TITLE: Teacher Understandings and Perceptions of the Teacher Evaluation Process in Southwest Virginia
IRB NUMBER: 16-048

Effective January 26, 2016, the Virginia Tech Institutional Review Board (IRB) Chair, David M Moore, approved the New Application request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

<http://www.irb.vt.edu/pages/responsibilities.htm>

(Please review responsibilities before the commencement of your research.)

PROTOCOL INFORMATION:

Approved As: **Expedited, under 45 CFR 46.110 category(ies) 6,7**
 Protocol Approval Date: **January 26, 2016**
 Protocol Expiration Date: **January 25, 2017**
 Continuing Review Due Date*: **January 11, 2017**

*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
 An equal opportunity, affirmative action institution

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

Appendix B

Interview Protocol

Introductory Remarks:

I want to get a sense of what you know about the evaluation process under which you currently operate as a classroom teacher. I will be asking you questions which fall into four categories. First, I will want to get some background information about where you have worked and with which grades you have had experience. Then, I will be asking questions about the evaluation system. Finally, I will have a few questions about your perceptions of the evaluation process and your beliefs about ways in which it possibly impacts your practice.

Demographic Questions:

I would like to gather some information about your overall number of years you have taught elementary school, including where you have taught and in which grade levels. This will give me some sense of your overall experience as a teacher.

1. In which school(s) and in which system(s) have you taught?
2. Tell me about your years of teaching experience.
3. Describe the various age groups with which you have worked.
4. Please describe your teaching assignments. Include ways you have worked with colleagues such as co-teaching, working with special education teachers, any experiences with inclusion, or with paraprofessionals.

Research Question #1: What are teacher understandings of the teacher evaluation process?

Focus: Description of the Evaluation System in Their School Division

5. Tell me what you know about the current evaluation system in your school division and how it was constructed. Describe for me any ways you personally had input into the design of the evaluation system.
6. Describe how the current evaluation system works to evaluate teachers. I want to know how often you meet with your evaluator and what types of feedback you receive about the strengths and weaknesses in your practice.

Research Question #2: What are teacher perceptions about their experiences with the teacher evaluation process?

Focus: Perception of Their School's Process

7. Please describe for me how your evaluator uses the evaluation system in your school.
8. Describe for me what you know about the data that is collected during the evaluation process. What happens to these data and how does your evaluator use them?
9. Describe for me ways you believe the current evaluation system is used to help make improvements to your own practice.
10. How do you think it is working in your school? Include the system of evaluation, as well as your experiences within the process.

Research Question #3: In what ways do teachers believe that evaluation informs their practices?

Focus: Beliefs Teachers Hold About Their Evaluation Experiences

11. Please describe for me how the evaluation system currently informs your practice. Do you believe it is doing the job it was designed to do or do you believe it could work better?