IDENTIFYING PROFESSIONAL DEVELOPMENT NEEDS OF GEORGIA PRINCIPALS AS EVALUATORS: PERCEPTIONS OF PREPAREDNESS WHILE IMPLEMENTING THE TEACHER KEYS EFFECTIVENESS SYSTEM

by

TRICIA ANN PATTERSON

BOB JOHNSON, COMMITTEE CHAIR
DAVE DAGLEY
AARON KUNTZ
BRENDA MENDIOLA
PHILIP WESTBROOK

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Education in the
Department of Educational Leadership,
Policy, and Technology Studies
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2015

ABSTRACT

The purpose of this qualitative study was to determine the perceived level of self-preparedness of building level principals to conduct effective teacher evaluations and to identify the future professional development needs while using the Georgia Teacher Keys Effectiveness System. The study conducted a program evaluation needs assessment that addressed further training needs of principals to effectively use the evaluation instrument so future professional development can be planned. As schools work toward meeting the higher standards of new teacher evaluation instruments, finding key areas where principals need further training to use the instrument was the objective of this study. Hence, the questions that were posed in the study addressed how building principals report how they were prepared to implement the instrument and give effective feedback while using the evaluation instrument, what the feelings toward the instrument are in contrast to others used, what the benefits and challenges of using the instrument might be, and what additional professional development they need to become effective evaluators. A qualitative approach was implemented in a three-stage process. The participants were part of two semi-structured individual interviews and a single focus group interview.

DEDICATION

I dedicate this dissertation to my family. They have been instrumental in the completion of this process. My parents, Barb and Randy, wiped away tears, celebrated every small victory, made sure they were at games I had to miss for the kids, and were always with me either in spirit or for real. It is the work ethic and moral responsibility you taught me that got me to this point in my life. My kids have felt each and every bump and bend in this long road to the finish. Ella was five when I started and is now nine – Cooper was one and is now six. They spent countless mornings watching TV or keeping each other occupied as I worked at that lonely corner at the kitchen table. Their hugs and kisses throughout each paper I wrote and chapter I edited meant the world to me and kept me going. Finally, my rock and best friend – the person who showed tough love and never let me think I couldn't finish. Thank you, Bart, for being so much more than a husband and teaching me that it is ok for the kids to have ice cream and doughnuts for dinner when mommy is at class. We all grew over the past four years and I am more than ready to give you all more time as I close this chapter in the book of our lives.

ACKNOWLEDGMENTS

I would like to thank the people who have assisted me in the achievement of a lifelong academic goal. The opportunity to research a topic about which I am passionate and feel is important for the development of our education system has led me to a greater appreciation of the hard work and dedication which school administrators have for conducting teacher evaluations. The professional educators on my dissertation committee who have assisted me in this process made a difference in my research and focus: Dr. Bob Johnson, Dr. Dave Dagley, Dr.Aaron Kuntz, Dr. Brenda Mendiola, and Dr. Philip Westbrook. I would also like to offer a special thank you to Dr. Bob Johnson, my committee chair. Dr. Johnson and I danced this dance together for two years and his questions, process, and guidance led to personal reflection and professional growth.

My professional and personal friends began this journey with me four years ago and supported me every step of the way. My colleagues allowed me to ask questions, participated in conversations to develop an understanding of evaluation and kept my focus on improving our field at the forefront of my work. Thank you, Becci and Karen, for being mentors and always supporting me as I took time to complete assignments or drive to class; Donna, thank you for sharing your professional experience and giving honest feedback. And to all of my friends who were there to help with childcare so I could have un-interrupted time to write and then take me away from the computer when I needed a break.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION TO THE STUDY

> "A profession is not created by certificates and censures but by the existence of a body of professional knowledge, as well as a mechanism for improving it, and by a genuine desire of the profession's members to improve their practice," (Stigler & Hiebert, 1999, p. 146).

The field of education has seen frequent changes in every aspect of the profession, but one change currently under discussion in many states and districts nationwide is that of teacher evaluation. The evaluation process for teachers in America's public schools dates back to the early 1900s when criteria for measuring teacher effectiveness fell into three categories: ratings, systematic observations, and standardized achievement tests (Popham, 1987). Through the twentieth century supervision as inspection was prevalent. Taylor's *Principles of Scientific Measurement* (1911) engaged scientific approaches to teacher supervision and then, in the 1920s, John Dewey's influence introduced democratic forms of teacher supervision, causing a reform from supervision as inspection to supervision as improving instruction (Sullivan & Glanz, 2000). Although teacher evaluation has been evident since the 1900s, research suggests that the standard model used to evaluate teacher effectiveness has remained essentially unchanged. The irony is that almost every other evaluative process in schools has changed dramatically. Danielson and McGreal (2000) contended that, "The traditional approach to teacher evaluation is no longer adequate," (p.7). As evaluation models finally begin to change, this study investigated principals' perceptions regarding their preparedness to effectively evaluate teachers and conduct a needs assessment to determine what future professional development might be beneficial.

A portion of the literature on teacher evaluation related it to supervision as a means of improving teacher practice and ignored the teacher evaluation process, while a larger group of

studies discussed the link between teacher supervision and teacher evaluation as part of supervisory practices (Acheson & Gall, 2003). Charlotte Danielson (2007), who has studied the teacher evaluation process for several years stated that a blueprint for teacher evaluation must include instruments and procedures that provide evidence of teaching (the "how") and clear definition of the teacher (the "what"). Studies of the impact of teacher evaluation are often closely linked to the National Board Standards of Professional Teachers Standards (NBPTS). The NBPTS developed an assessment system that was connected to many teacher evaluation systems and consisted of varied methods for assessing teacher quality. Some of these methods included: Portfolios, on-site performance evaluations, and assessment centered exercise, including monitoring classroom teachers during the various stages of development. The NBPTS was historically situated as part of the standards movement in education during the late twentieth century. Standards for students must be matched by standards for teachers. The purpose for the creation of tougher teaching standards suggested that as the demands on students became more rigorous, guarantees that the educational system is staffed with professionals capable of teaching to achieve these standards became more essential. Moreover, standards for these teachers must be measured by evaluators who are trained to monitor and give effective feedback to effectively understand the standards.

Recently, change in teacher evaluation has become an educational phenomenon that impacts teachers and evaluators and has been of great interest to political agendas and social reform policies. Fullan (1993) found that though there have been many years of new policies being approved or developed it is only recently that the work of education reformers can be analyzed to consider what changes are worthwhile. "The change initiatives in many educational areas involve work grounded in a strong knowledge base because it is deeply grounded in

action," (Fullan, 1993). The area of effective teacher evaluation has been undergoing rapid change, along with the role a principal plays in completing this important piece of accountability in education. As the focus of teacher evaluation has changed from teaching to student learning through accountability measures such as standardized testing, the focus of principal training for teacher evaluation might also need to change to that in which reflects how they give effectively use the instrument.

As a school principal, doing the right thing means doing what is best for students. "Most people will agree that principals are the most important leaders in our school system. Most will also agree that effective leadership is schools is still dismayingly, exceedingly rare" (Schmoker, 2006, p.3). Principals have a good understanding of who the effective teachers in the school are from daily interactions, current teaching rubrics, and working history with them, but as evaluation tools change, the definition of "effective" might also need change. Therefore, principals should consider stretching and deepening their knowledge of what effective teaching is and how to help teachers reach such a platform through the way in which they use the evaluation instrument.

Statement of the Problem

Individual districts in the state of Georgia were given the responsibility of providing the necessary training for principals to implement the Georgia Teacher Keys Effectiveness System (TKES). The districts were given the flexibility to organize and require trainings as they felt necessary and sufficient as an adequate measure of evaluator preparedness. There is some uncertainty as to whether the trainings meet the needs of the principals, and school districts in Georgia must determine how to measure their needs and offer substantial support in using the instrument. An improvement to present training offerings and creation of future professional

development is dependent upon a two-way exchange of information between the client or client system (i.e., evaluator) and the providing agency (i.e., the school district). The task of conducting a program evaluation needs assessment is one that the school district should carry out on a continuing basis with precision and care so that the needs of the evaluators can be identified and prioritized. At that point, the district may allocate resources to the areas of training needs with the highest priority. The problem identified as the objective for resolution by this study was to determine the level of self-preparedness of building level principals to conduct effective teacher evaluations and to identify the future professional development needs while using the Georgia Teacher Keys Effectiveness System.

Purpose of the Study

The purpose of this qualitative study was to obtain data and provide resolutions for the problem stated above i.e. to determine the level of self-preparedness of building level principals to conduct effective teacher evaluations and to identify the future professional development needs while using the Georgia Teacher Keys Effectiveness System. This study contributed to the educational knowledge base by exploring the perceptions of evaluators who participated in using TKES to conduct teacher evaluations. It addressed these evaluators' views on their preparedness to evaluate teachers and assess their own professional growth needs. Using the lens of the building principal, the focus of this case study was to develop a Training Needs Assessment for a Georgia school district to determine the level of preparedness to: (a) determine the benefits and challenges of using the instrument, (b) navigate the instrument in comparison to those used in the past, (c) give effective feedback while using the evaluation instrument, and (d) help guide future training needs for principals in the area of giving meaningful feedback. The qualitative review of the preparedness of principals to give feedback and results of the Training Needs Assessment

were determined through utilization of interviewing current principals who have received initial training, called credentialing, and examining district and state evaluation documents and policies.

Rationale and Justification for Study

Educational research has repeatedly identified teacher "effectiveness" as the most important factor in student learning. The work that teachers accomplish in classrooms matters and public education's top priority should be developing and improving teachers. Teacher supervision and evaluation have different outcomes, but they also complement one another. Supervision is providing coaching to build teachers' capacity while evaluation is assigning merit to teachers' performance. Evaluation is a reflection of human performance in organizational settings (Stronge, 1991). It is an element of every-day life in American culture and people are evaluated by what they say, how they act, where they go and what they do. In the field of education, the inconsistency is the method by which people are evaluated and whether there is room for supervision within the evaluation. One way supervision is included in an evaluation instrument is through the way in which the instrument is used and the feedback that a principal gives a teacher during their evaluation.

The concept of evaluation is closely linked to accountability in education. As Thomas, Holdaway & Ward (2000) stated, "Evaluation of an individual or group is completed to make sure that those being evaluated are accountable to established standards detailing what the individual ought to be doing". Evaluation in the education field has been a central component of instructional improvement that provides rigorous performance measurement and useful feedback (Shakman et al., 2012). The goal of using evaluation as a reflection of one's performance remains the same, but the objectives or outcomes have changed over time. The use of evaluation has changed from a form of documentation to a supervisory tool to help teachers improve their practice and support personnel decision-making. Several factors have contributed to the lack of

effective teacher evaluation, and, until the late 1970s, there was virtually no research on effective teaching. Sweeney and Manatt (1986) noted that there was little serious attention given to teacher evaluation until the late 1970s and that the previous two decades were marked by self-evaluation, ceremonial congratulations, and neglect. Valid criteria for performance evaluation are major parts of the process of evaluating teachers and a lack of research-based criteria has hindered the process of assessing teachers. There has also been a lack of consensus about the purpose of evaluation and which instruments should be used to effectively evaluate staff.

Prior use of teacher evaluation was primarily to document whether a teacher exhibited the minimum competencies to receive tenure or, on the opposite end of the continuum, it was used as a document to assist in the removal of a teacher. Teacher evaluation was not seen as a high stakes activity, but as an annual, routine procedure for which school districts, historically, were allowed to negotiate their own evaluation process. Prior research showed that many teachers did not feel they received enough feedback from their evaluations to tell them how they were doing or to help them grow professionally (Gimbel, Lopes, & Nolan Greer, 2011; Howard & McColskey, 2001; Kimball & Milanowski, 2009).

Evaluations should be designed to promote teacher effectiveness and a great deal of attention has turned to determining effective methods of evaluating teacher performance. "Most authors identify the fundamental purposes of teacher evaluation as improving performance and documenting accountability" (Tucker & Stronge, 2005, p.6). Due to a heightened sense of accountability and the change in objectives, the structure and process for conducting evaluations have begun to change (Danielson & McGreal, 2000). "Better defined goals have shifted the educational dialogue from vague opinions about student progress to factual evidence of student performance," (Peterson & Peterson, 2006, p. 91).

For about thirty years, studies have documented the decline in education and researchers have identified specific areas in need of improvement with teacher evaluation being one of them. One method to increase teacher performance is through the effective use of the evaluation process. Evaluators can only improve the performance of classroom teachers if they are trained to give valid, reliable, meaningful feedback while using the instrument effectively. Over the past decades the federal and state governments have made various attempts at educational reform by taking an advisory role, but it has had little impact. As our students continue to fail to meet the expectations set forth by top ranked countries and guidelines to become college and career ready, it is imperative that something new be done. Education reform has followed changes in the purpose for teacher evaluation by adjusting the systems put in place and the policies that guide them. Reforming the evaluation system through enhancing the skills building principals need to be effective evaluators of teachers is crucial if we are to meet the current performance goals.

In 1983 the National Commission on Excellence in Education (NCEE) reported in *A Nation At Risk* that educational institutions were grossly underachieving. The report stated:

> "Part of what is at risk is the promise first made on this continent: All, regardless of race or class or economic status, are entitled to a fair chance and to the tools for developing their individual powers of mind and spirit to the utmost. This promise means that all children by virtue of their own efforts, competently guided, can hope to attain the mature and informed judgment needed to secure gainful employment, and to manage their own lives, thereby serving not only their own interests but also the progress of society itself" (U.S. Dept. of Ed., 1983).

The report goes on to emphasize the low skills as evidenced through test scores in American youth. Findings in the report were that school content was diluted and lacked purpose, use of classroom time was not effective, the field of teaching was not attracting enough academically able students, and that teacher preparation programs needed substantial improvement. In other words, the findings showed that the teaching profession needed to be

improved and, thus, a new system of evaluation to show improvement was needed. Over the past twenty-seven years, Jon Saphier has looked at data on how administrators have evaluated teachers. He found that, "The incoming skills of administrators tend to be quite low. Administrators have little or no preparation for analyzing instruction when they are hired, and almost never had to show they could do so to get their job" (2008, p. 2). In addition, with the budget cuts many districts are facing in current times, funds to train current administrators on how to use the new evaluation tools and clearly understand the components of them is minimal. The report concluded that the Nation was falling behind globally, and that our country was not as competitive as it once was (NCEE, 1983). The decline in education has had a negative effect on society both socially and economically.

A key influence on education policy was a 1996 report from the National Commission on Teaching and America's Future (NCTAF), titled *What Matters Most*: Teaching for America's Future. The commission called out teacher quality as the key to improving American education and contended that the capability of the teacher has the strongest effect on student learning. "Recruiting, preparing, and retaining quality teachers are the most important ways to improve education" ("What Matters Most," 1996).

In late 1999, Georgia introduced the evaluation of student performance on statewide assessments into the teacher effectiveness debate (Eady & Zepeda, 2007). State lawmakers introduced this concept with the passage of the A Plus Education Reform Act of 2000. This legislation required administrators to consider student achievement to a minimal degree in the annual appraisal of teachers. With the passage of the Act, teacher expectations changed and incorporated "The role of the teaching in meeting the school's student achievement goals, including the academic gains of the students assigned to the teacher" (OCGA 20-2-210, 2000).

The debate about how to identify quality teachers continues today, but it was impacted greatly when the George W. Bush administration passed legislation known as the No Child Left Behind Act (NCLB) (2001). The Act placed the federal government in a more prominent role in American Education and increased accountability among schools and teachers. The legislation required schools to pass yearly standardized tests to measure progress and provide highly qualified teachers to all students, both of which are determined at a state level. NCLB faced heavy resistance from within the educational community for not including enough support to match the high expectations it set forth.

In more recent years, change in teacher evaluation has accelerated. In 2009 the Obama administration provided financial incentives to encourage educational growth. Through the American Recovery and Reinvestment Act (ARRA), over four billion dollars in competitive grants were invested to improve education at the state and local level. The "reward" focused on reform in four areas: Rigorous standards and assessments that would prepare students for college and the workforce; data systems that effectively measure student growth and success while informing principals and teachers how they can improve instruction; recruit, develop, reward, and retain the best teachers and principals; and turn around low-achieving schools. Race to the Top (RTTT) grants were awarded to states that tied teacher evaluation ratings to students' performance on state assessments. Forty-one states submitted applications and the following were awarded grants to be used toward the above-mentioned: Delaware, the District of Columbia (considered a state in this case), Florida, Georgia, Hawaii, Maryland, Massachusetts, New York, North Carolina, Ohio, Rhode Island, and Tennessee.

As states and local districts developed instruments to align with RTTT guidelines, they used criteria and categories, rating scales, response modes, comment sections, and formats based

on concepts from the literature on effective teaching practice. Evaluation was not defined under RTTT, but instead describes the components of evaluation and reference what it can be used for (U.S. Department of Education [USDOE], 2009). Therefore, state legislatures passed bills requiring school districts to use a multi-level tier rating system instead of the old satisfactory/unsatisfactory systems. Within the past four to five years, new teacher evaluation models (i.e. McREL, Stronge, and Marzano) have included well over twenty indicators that represent successful teaching and the new models encourage collaboration and communication between the evaluator and the evaluated. They also allow for self-reflection and growth, establish teaching goals, and incorporate student measurements to coincide with increased observations of the teacher by the evaluator. As a result of RTTT changes, teacher evaluation has transitioned from focusing on teaching to focusing on student learning. Johnson (1997) wrote "By providing the public, administrators, and teachers with data on teachers' skills and performance, a teacher-evaluation system will increase the instructional productivity of teachers, enhance student learning, and ultimately improve the quality of the educational system," (p. 70). A change such as this required that building principals and other principals have a new set of skills that will lead to teacher self-reflection, growth and improved student achievement. The research in this study examined the principals' perceptions of their own readiness to conduct effective evaluations on their teachers and the appropriateness of the professional development provided by the local districts to enhance their skills. Investigating principals' perceptions of their readiness to conduct effective teacher evaluation and what they feel they need to develop their skills, this study used themes derived from the responses given in interviews to determine if there was a difference between what is perceived to be needed and what is actually provided at a local level based on the state requirements.

As part of Georgia's Race to the Top grant, new systems for providing teachers and leaders with meaningful feedback and support opportunities have been developed. The systems were meant to have clear and precise indicators and resources to guide teachers and leaders to improved performance which will positively impact student achievement. The question remains, is this assumption valid and, if so, is the intent actually realized in the results produced by the policy, once it is implemented? In spring 2012, RTTT districts piloted the Teacher Keys Effectiveness System (TKES). Now, as we look toward the 2014-2015 school year, all districts in Georgia will be required to use TKES. Although resources exist, questions continue to linger as to whether proper training has been provided for administrators to implement TKES in an effective way that will lead to teacher improvement and student achievement.

Teacher evaluation reform has clearly happened. Remaining is the debate about whether this current course of action is what is best for teachers, principals, and students. Reports are needed from teachers and principals to determine whether or not educators are philosophically in support of the new models of teacher evaluation. This research does not examine the process or philosophy of different effectiveness models, but rather approaches the topic through the lens of principals to determine which resources and ongoing training they report they need to become effective evaluators of teachers. Krein (1990) wrote, "I propose that the problem is not an inability to determine what the components of a good program should be, but the failure to implement them properly," (p.1).

Evaluations, as defined by Stronge (1991), are a reflection of human performance in organizational settings. The goal of using evaluation as a reflection of one's performance remains the same, but the objectives have changed over time. Evaluation was once a form of documentation and now it is seen as a tool to help teachers improve their practice and support

personnel decision-making. Now a central component of instructional improvement, evaluation should be a rigorous performance measurement that provides useful feedback (Shakman et al., 2012).

Over the past thirty years, research has consistently concluded that effective evaluations should include student data, multiple measures, and evaluator training. The question remained: What specific data provides the most valid and reliable measures of teacher effectiveness and as an evaluator uses multiple data, what specific combination of data are needed and in what proportion? Furthermore, how can this data be collected in a valid, reliable, inexpensive and least obtrusive way? As TKES has been implemented, principals need to decipher whether they are equipped in their knowledge to discern effective teaching and give feedback regarding what needs to change or remain the same. The TKES Handbook terms training as "credentialing". According to the Handbook (2013), "Credentialing is a process of establishing the qualifications and proficiency of evaluators to utilize TKES. The credentialing assessment was a minimal competency assessment that measured participants' understanding of the information and practice provided during training. It included both recall of specific TKES information and practice of all the TKES processes and was a bridge to ongoing learning as evaluators worked with TKES. All evaluators had to pass the credentialing assessment prior to using the Teacher Keys Effectiveness System (TKES)," (p. 16). There is an expectation that evaluators will continue to familiarize themselves with the TKES process as they work within their districts and the credentialing process will provide calibration that further increases the alignment of evaluation ratings. Administrators at local schools were trained and credentialed by representatives from the Georgia Department of Education with each district having a lead instructor in which to direct questions and concerns.

As instructional leaders in the school, building principals serve as the deciding factor for appropriate evaluation practices and coordinate all evaluation activities within the school. A critical feature of TKES is that after initial training and credentialing, principals were encouraged, but not mandated, to review classroom observation videos, observe lessons in classrooms and discuss ratings along with judgment of practice based on the TKES standards rubric. The lack of mandate represented a political compromise between the state and districts that took the strength out of the policy for districts that were not aggressive in pursuing such follow-up training.

"Most authors identify the fundamental purposes of teacher evaluation as improving performance and documenting accountability," (Tucker & Stronge, 2005, p. 6). New evaluation tools, such as TKES, have been used to hold teachers, administrators, schools, and districts accountable. Accountability is the focus of many political agendas and social reform policies today. Change happens daily in the way education is packaged and delivered, and administrators need to have knowledge of what their expectations are going to be, not only in current times, but in the future. It is a professional imperative that they educate, motivate, and advocate for administrators who evaluate to gain the knowledge needed to effectively use the instrument and give feedback that will help their teachers improve their ability to education our students.

One social reform of education was the use of private funding to research and improve current trends. The Measures of Effective Teaching (MET) Project, funded by the Bill and Melinda Gates Foundation, has analyzed thousands of lesson videotapes and studied the shortcomings of current practices (Marshall, 2012). Researchers have found that one classroom observation a year does not give an accurate picture of a teacher's work. However, filling out elaborate rubrics after every visit, as is the current expectation with TKES, creates an impossible

workload for principals, leaving less time for informal classroom visits and interactions with teacher teams. As cited in Marshall, 2012, “School leaders in Tennessee are suffering this fate under the state’s cumbersome new evaluation system”. Time is precious for all educators today and school leaders can’t afford to spend it on the bureaucratic tasks included in new evaluation tools.

Time spent on the evaluation process is only as effective as the principal that takes the time to fulfill the requirements and follow-up with educators to sustain positive growth. The Interstate Leaders Licensure Consortium (ISLLC) defended six standards for school principals including: “A school administrator is an educational leader who promotes the success of all students by advocating, nurturing, and sustaining a school culture and instructional program conducive to student learning and staff professional growth,” (pg. 294). The time spent giving feedback with TKES can help to advocate, nurture, and sustain a school culture conducive to student learning. However, it is crucial that evaluators understand how to navigate the instrument and give appropriate, differentiated feedback. Historically, the two main factors that have motivated the nation’s move to re-design teacher evaluation are a lack of meaningful feedback and lack of differentiation between high and low-performing teachers (Sartain, et.al, 2011). A reason for this is what is effective for one teacher might not be for another. The “technology of teaching” in many cases lacks clarity on how to evaluate teaching because the goals and standards involved are often not entirely understood. Current evaluation systems and training for evaluators do not provide teachers with the information they need to make timely and effective improvements in their instructional practice (McLaughlin, 1990).

Teacher effectiveness through classroom instruction and the learning gains of students should involve evaluation of the act of teaching, as well as the results of teaching. Current

changes in evaluation systems have moved toward providing both feedback and guidance toward improving a teacher's professional practice, but they are still often used as accountability documentation rather than a guide for teacher growth and improvement. As stated by Schmoker (1999), "When we begin to more systematically close the gap between what we know and what we do, we will be on the cusp of one of the most exciting epochs in the history of education" (p. 70). Many of these studies have looked at the teacher evaluation process and have focused on the process and have identified the most common mistakes that evaluators make (Ostovar Namaghi, 2010). Few studies have examined the professional development evaluators need to become effective evaluators of teachers. The following qualitative, program evaluation needs assessment explored the implementation and impact of Georgia's TKES and linked them to the perceptions held by a group of evaluators. This study provided a philosophical sounding board for teachers and principals on how they perceive the new evaluation system being set before them and whether or not this model is a sufficient measure of teacher effectiveness.

The study expanded on previous studies on the topic of teacher evaluation, but is significant because it focused on the preparation of the evaluator in terms of implementing the instrument. Mitgang (2012) highlighted the importance of training and support for principals, after they are hired, that is tailored to both individual and district needs. This study specifically looked at the training and support in regards to the implementation of a new evaluation system. Prior studies showed that many principals were not adequately prepared to be effective evaluators. Many of the studies conducted previously showed that evaluations done by administrators were often based on perception and were often influenced by factors such as the halo effect or leniency (Haefele, 1993; Wise et al., 1985). There were also other barriers that affect the evaluation process and its outcome, including deficiencies in the instrument tool and in

the organization itself (Stronge, 1991). "Indeed, much of the published literature on the principal's role in evaluation consists of explanations of procedural and legal aspects of evaluation and calls for better training of principals," (Painter, 2000, p. 2). Many studies have looked at the process and identified the most common mistakes that evaluators make (Freeberg, 1969; Johnson, 1997). Studies have also concluded that the process is ineffective if it is a tool to help guide personnel decisions and a teacher's professional growth (Wise et al., 1985). Few studies have inquired about the professional development in terms of the amount of support an evaluator needs to conduct an effective teacher evaluation.

The study expanded upon previous studies conducted on the topic of teacher evaluation. It is significant in that it provided valuable information for state and local educational leaders as they work to implement new teacher evaluation systems. Ongoing support will be needed to implement effective evaluation systems and this study was undertaken because there are few studies focused on evaluators and the professional development they need. The study also added to the existing knowledge base of teacher evaluation by providing an assessment of what is needed in the area of supports for principals to conduct valid and reliable teacher evaluations. It brought a stronger understanding of what strategies evaluators felt were effective and beneficial to them as they advanced their skills as evaluators. The results helped the district, other districts in the state, and states using a similar system to integrate practical applications for evaluator staff development in order to produce a greater source of validity and reliability in the evaluations completed.

The research presented has the potential to guide the regulations regarding the training principals receive on the process and way they give feedback while using the instrument to evaluate teachers. It is through the findings of this study that practices can be implemented to

ensure that principals have the skills and knowledge necessary to be effective. Ongoing supports and training are crucial for a transformational change in the way teachers are evaluated because, even those who believe principals support the teacher evaluation program, feel that they lack sufficient resolve and competence to evaluate accurately (Wise et al., 1985). The skills acquired by principals under the new system will help them facilitate teacher self-reflection, growth, and improved student achievement. This study captured the perceptions of a select group of building principals as they reflected on their level of preparedness to conduct effective teacher evaluations and identify future professional development needs while using the Georgia Teacher Keys Effectiveness System.

Research Questions

The research questions for this training needs assessment were as follows:

1. What are the benefits and challenges of using the TKES instrument?
2. What is the level of ease in navigating the instrument in comparison to those used in the past?
3. How is effective feedback given while using the evaluation instrument?
4. What future training needs do principals report they need to use the instrument and give effective evaluations?

Significance of the Study

The findings of this study are relevant to the field of education and specifically teacher evaluation because it is crucial for districts and states to understand the level of preparedness their principals report in order to conduct effective teacher evaluations and determine their future professional development needs. Districts can use this data to determine if they have trained their principals sufficiently. Kimball and Milanowski (2009) found that principals who

conducted more valid evaluations reported that district training helped build their evaluation skills. Therefore, the validity of evaluations increase as principals become more prepared. The comfort level of evaluating is primarily based on preparation and if a principal are not comfortable using a tool, it will not be as effective. In addition, if Georgia principals were going to take time away from other duties to fulfill the vast observation and feedback requirements of TKES, proper evaluation training should be provided so it is not viewed as a waste of time, thus taken less seriously.

Teacher evaluation has intrinsic importance to multiple groups of people. The most obvious group affected by teacher evaluation change is teachers themselves. As teachers modify their teaching to adjust to evaluation changes, principals also adjust to the complexity and time requirements of the new instruments. In addition, students are impacted by the evaluation changes and professionals working at the university level are impacted as they attack the task of preparing the principals and teachers. Finally, employees at the district level are impacted by changes in evaluation as they are given the task of who will evaluate and how they will be qualified to do so. The research from this program evaluation needs assessment provides a concrete, concise outline of what current principals reported to need to become effective evaluators.

Another significant reason this study was conducted was the lack of research on principal preparedness in terms of teacher evaluation. This study covered a current reality that evaluation systems are changing and the level of preparedness of the principals needs to be addressed. The study took place in a real-life setting where the implementation of teacher evaluation systems was based on relatively new theories that have been widely accepted but minimally evaluated. Timing was the reason a qualitative approach was used in this study because the research took

place during the implementation of the new system. Professionally, the timing was also significant as it occurred early enough in the implementation stage to allow the results to serve as an opportunity to reflect on the process of principal training and future preparation needs to conduct teacher evaluations.

Research Design and Method

Qualitative research was used in this study to explore the needs of evaluator training to use a new evaluation instrument. Qualitative approaches are useful when attempting to understand how actors in a particular context imbue meaning to their behavior. Qualitative methods were used in this study in order to understand the context of teacher evaluation training needs amongst a group of building principals in one school district in Georgia. The principals were selected based on equal representation of the different levels of education in K-12 education (elementary, middle, and high). According to Patton (2002) this type of research is based on "The assumption that there is an essence or essences to shared experience. These essences are the core meanings mutually understood through a phenomenon commonly experienced," (p.15). The shared experience for these participants was that they were all interviewed using guided, open-ended questions related to their use of a new teacher evaluator instrument and the training they received (or need to receive) to implement it effectively.

Data was studied in accordance to the accepted practices of qualitative research education. Data gathered from the interviews were audio taped using an I-pad, transcribed using Microsoft Word, and coded. In addition, field notes were taken during the interviews and at the conclusion they were compiled with observations pertaining to body language, tone, and a description of the interview setting. A coding system was used to link participant responses to the above mentioned research questions by highlighting key terms that related to the literature

studied. After highlighting, the researcher assigned letters to repeating terms to determine consensus and sub-themes.

After the data were compiled, a program evaluation needs analysis was completed. A needs analysis is a type of program evaluation that is a systematic process for determining and addressing needs or gaps between current conditions and desired conditions. The discrepancy between the current conditions and the wanted condition must be measured to appropriately identify the need. A needs assessment is part of a planning process on how to effectively support the positive attributes of the evaluation system and identify the areas of weakness. By clearly identifying the problem and evaluating the process, the resources being used to train and prepare evaluators can be directed toward developing and implementing a feasible solution (Fulgham & Shaughnessy, 2008).

After determining the current results, the desired results will be articulated and the distance between the results will be the actual need that is focused upon. In relation to TKES and what principals need, a training needs assessment will be a systematic inquiry of training needs within an organization for the purposes of identifying priorities and making decisions, and allocating finite resources in a manner consistent with identified program goals and objectives (Triner, Greenberry & Watkins, 1996). There are three levels of needs assessment: Organizational analysis, task analysis, and individual analysis. This study will focus on task analysis, specifically Human Performance Technology (HPT). It will provide data about a job or group of jobs and the knowledge, skills, attitudes, and abilities they perceive they need to achieve optimum performance (Miller, 2000).

Definition of Key Study Terms

A Nation At Risk: A report to the Nation and the Secretary of Education by the National Commission on Excellence in Education describing the erosion of American educational foundations as a "rising tide of mediocrity.

Accountability: The responsibility for implementing a process or procedure for justifying decisions made, and for results or outcomes produced.

Assessment: The process of or instrument for measuring, quantifying, and/or describing those aspects of teaching related to the attributes covered by the evaluation.

Content Standards: As the foundation of a standards-based system, content standards describe what content knowledge and skills students must master (American Federation of Teachers, 2001).

Data: The information and evidence gathered during the assessment process for use in determining the level of teaching performance.

Educational Testing System (ETS): ETS is a large project to provide a framework for state and local agencies to use when making teacher licensing decisions.

Elementary School Principal: An active public elementary school principal in Georgia.

Evaluation: A reflection of human performance is organizational settings.

Evaluation Stages: The major steps in the teacher evaluation process.

Formative Evaluation: An evaluation conducted primarily for the purpose of improving the teacher through identifying the teacher's strengths and weaknesses.

Georgia Teacher Keys Effectiveness System: TKES is the teacher evaluation model, designed by James Stronge, which the state of Georgia implemented as part of the Race To The Top grant.

K-12: Grades K-12

MET: Measures of Effective Teaching Project funded by the Bill and Melinda Gates Foundation.

National Board of Professional Teaching Standards: A recommendation from the Carnegie Foundation in 1986 which stressed performance based tasks and evidence of knowledge, skills, and disposition following the development of authentic assessment of pupils.

No Child Left Behind: The reauthorization of the Elementary and Secondary Education Act of 1965. It creates mandatory testing of students and improvement for schools and sets a minimum level of qualifications for teachings to meet to be "highly qualified".

Principal Experiences: The experiences that principals had with using and analyzing student achievement results as it relates to teacher performance evaluation.

Principal Preparation: The leadership preparation coursework taken to become an administrator and obtained through college level classes with the outcome of gaining a leadership level certificate in Georgia.

Principal Training: The administrative training received post-college level work through professional development activities and local system workshops.

Professional Development: Training and professional learning that has as its purpose the improvement and enhancement of a professional's performance and knowledge.

Public Elementary School: A school funded through public funds that served students where the majority of the population of students fell within the grades of kindergarten through fifth.

Qualitative Research: The basic qualitative study in education typically draws from concepts, models, and theories in education psychology, developmental psychology, cognitive psychology, and sociology.

Race to the Top: A 4.35 billion dollar federal competitive grant program designated to encourage states to and reward states for creating conditions for education innovation and reform.

Reform: A change effort that is undertaken to improve the educational system.

Reliability: The consistency of measurements across evaluators and observations.

Standard: The level of performance on the criterion being assessed that is considered satisfactory in terms of the purpose of the evaluation.

Standards-Based Reform: The change to an educational system that utilizes subject-matter benchmarks to measure student achievement, assessments aligned with standards to measure student performance, and accountability systems that provide rewards or sanctions to the district, school, and students based on student performance.

Student Achievement Results: Student test data gained from different assessments such as the Georgia Criterion-Referenced Competency Test (GCRCT), classroom benchmark tests, teacher-made tests, and norm-referenced tests.

Summative Evaluation: An evaluation conducted primarily for the purpose of making personnel decisions about the teacher.

Teacher Evaluation: A means of assessing a teacher's performance and progress in the classroom.

Triangulation: The attempt to obtain more valid results by using multiple sources of data about one aspect of performance, multiple methods of collecting data, and /or multiple interpretations of the same data.

Validity: Accuracy and comprehensiveness in an assessment as defined by agreed-on criteria.

Limitations of Study

There were limitations to this study because of the nature of the study's design. A case study of a single school district limits the number of participants surveyed and interviewed. Therefore, the capacity to generalize the findings is decreased. The information obtained was limited only to those building level principals who participated in the study and the findings were constrained by the degree of accuracy with which the participants truthfully communicated their perceptions. There were also limitations on the needs analysis due to the fact that it took only the needs of those interviewed into account. The study cannot assess the long-term effects on the participants. Since it is a time-limited study the results will not be tracked over time and prevents links to student achievement.

Organization of Study

Chapter one presented: The introduction to the study; the rationale for the study, along with the statement of the problem; the purpose of the study with the research questions, theoretical framework, and significance of the study; a brief summary of the research design and methods; definitions of key terms; limitations; and the organization of the study. Chapter two provides a review of the literature related to the issues and domain of the study. Chapter three describes the methodology used in the study. Chapter four presents the findings of this study, and Chapter five summarizes and discusses these finding and makes recommendations. References and appendices of pertinent sample documents conclude the study.

CHAPTER 2: REVIEW OF THE LITERATURE

Chapter II reviews the literature that examines the purpose of teacher evaluation, the history of teacher evaluation, federal education reform on teacher evaluation, education reform specific to the state of Georgia, components of teacher evaluation, components of Georgia's Teacher Keys Effectiveness System (TKES), evaluator training, and TKES evaluator training. Literature in the area of Needs Assessments is also included due to the nature of this study. The literatures reviewed framed the study by understanding the background of teacher evaluation and education reform and noting the areas where the gaps in research on effective evaluation systems and evaluator training are present.

Teacher evaluation is not a new concept in education. Beginning from about 500 B.C. and leading into the twenty-first century, thousands of summaries indicate that progress is only recently being achieved in the development of an effective teacher evaluation system that promotes professional growth and student achievement. A primary duty of any school administrator is to conduct periodic evaluations on all teachers working in his or her school. It is considered a central duty because, "research shows a critical link between effective teaching and students' academic achievement" (Mathers, 2008, p. 1). However, "Without high quality evaluation system, we cannot know if we have high quality teachers" (Stronge and Tucker, 2003, p. 3). The evaluation of teachers is a controversial issue that is continuing to change the way teachers and evaluators view their roles and responsibilities. However, reform efforts must

improve and not continue historical inequities (Laguardia & Pearl, 2009). The first step is to understand the purpose of teacher evaluation.

The purpose of the following review is to examine the research on teacher evaluation in a manner that first discovers the reason for teacher evaluation and then not only the history of the development of evaluation systems, but also the way education reform in America and Georgia has impacted the history. The models of evaluation and the components of each individually and those in which are shared will be revealed as a means to compare and contrast which have been proven to be effective and are also included (or excluded) specifically in Georgia's new system. Research from various literatures that expands outside the spectrum of scholars and into the arena held by professionals in the field, being those whom have implemented systems of evaluation will be exposed to help in pinpointing the importance of evaluator training and the mandates that arise from implementation, specifically relating to this study in terms of Georgia's new evaluation system.

This literature review will inform the study by giving relevant background on and purpose of teacher evaluation, specifically the new Georgia system. It will also contribute to the essence of the questions that will be asked to participants who are the actual evaluators implementing the system. In addition, when conducting the needs assessment the outcome will enable sound evidence to be presented for future changes that could help the instrument become more effective.

Purpose of Teacher Evaluation

The purpose of teacher evaluation has changed throughout American history and it is currently transforming from a rudimentary process of documentation to that of an effective tool to build capacity in teachers and ensure that there is an effective teacher in every classroom. Teacher evaluation is influenced by several factors: Past practices, industry, government

influences, sociological factors, values, sentiments, and preferences of American society (Sweeney & Manatt, 1986). Wise, et.al (1984) noted that a consistent finding in almost all successful evaluation systems is the importance of establishing a clear understanding of the purpose of the system, which must then be reflected in procedures and processes and the use of the proper instrument. According to Denham (1987) most researchers seem to agree that the major purposes of teacher evaluation are:

1. Provide a process that allows and encourages supervisors and teachers to work together to improve and enhance classroom instructional practices.
2. Provide a process for bringing structured assistance to marginal teachers.
3. Provide a basis for making more rational decisions about the retention, transfer, or dismissal of teachers.
4. Provide a basis for making more informed judgments about differing performance levels for use in compensation programs such as merit pay plans or career ladder programs.
5. Provide information for determining the extent of implementation of knowledge and skills gained during staff development activities and use in judging the degree of maintenance of the acquired knowledge and skills.

In the research presented, the evaluator is considered the principal and administrative team of a K-12 school building. Kersten & Israel (2005) contended that the school building administrator is defined as the catalyst for the movement towards school improvement by means of teacher evaluation. They noted that the evaluation tools, along with the amount of time, format, and feedback, have changed dramatically from the past procedures used by principals. Millman and Darling-Hammond (1990) examined the evaluation processes, purposes and methods in their edited book and noted the strengths and limitations to the many purposes of

evaluation. In it, they reported Natriello's three major purposes of evaluation to be: Improving performance, personnel decisions, and legitimizing the organizational control system.

Marzano explained that teacher evaluation systems should: Focus on development that is comprehensive and specific, include a developmental scale to develop and track improvement, and acknowledge and reward growth. He concluded that both measurement and development are important aspects of teacher evaluation and the emphasis should be on the latter. Millman and Darling-Hammond (1990) reported that the eight purposes of teacher evaluation are: Pre-service education, teacher selection, certification and licensure, assistance to and assessment of beginning teachers, evaluation for tenure and retention, evaluation for teacher improvement and professional development, evaluation for career advancement and merit pay, and teacher evaluation for school improvement. The purpose of evaluation has changed throughout American history with the changing of evaluation systems.

The History and Evolution of Teacher Evaluation

Teacher evaluation serves many purposes, but the evolution of evaluation and the role of the evaluator has transformed over the course of American history from the late 1800's to the mid twenty-first century. Literature on the topics of teacher supervision and evaluation and the connection to improving teacher practice and student achievement is extensively diverse. Teacher evaluation has changed throughout history from a process of inspection to instructional supervision and, recently, an attempt to improve teacher effectiveness. Sullivan and Glanz (2000) noted that as early as 1709 the city of Boston defined supervision as a group of inspectors who would visit each public school in order to verify the methods of teachers in their performance in the classroom. This form of inspection and supervision continued into the middle of the nineteenth century and focused on maintaining the methods of evaluation that were already in place. By the end of the nineteenth century, school reform mandated a change in the

evaluation process that resulted in more efficient and centralized administrative bureaucracies, with superintendents of school districts in charge of the supervision process (Sullivan & Glanz, 2000).

In the early twentieth century, the building principal (originally called the principal teacher) chose a supervisor, usually a woman, to assist new teachers in mastering subject matter. The special supervisor did not evaluate teachers. Instead, supervision was regulated by the principal and a general supervisor, both male. Eventually, the role of special supervisor was eliminated due to gender bias and two supervisory positions were left: Principal and Vice Principal (Sullivan & Glanz, 2000).

During this time Frederick Winslow Taylor published *The Principles of Scientific Management* (1911) and made an impressive impact on business and education, specifically how supervision and administration were carried out in schools. Taylor suggested that scientific study could determine the proper method of doing every job (Spring, 2008). Management directed the actions of the workers and assumed more duties by reducing workers responsibilities and need for decision making. The influence of Taylor's scientific method led to the standardization of education, which is a cost-effective business concept. Standardization in schools resulted in a standard curriculum and standard attendance procedures, hiring practices, teacher training procedures, student evaluations, and teacher evaluations.

In 1915, Boyce was the first to study an instrument to measure teacher efficiency. The instrument gave the principal a list of criteria thought to be related to teacher effectiveness. The teacher was then rated on each criterion based on the principal's opinion and knowledge in the field. Franklin Bobbit, a professor at the University of Chicago, adopted Taylor's efficiency model and advocated the impersonal scientific approach to supervision, which developed into the

practice of rating teachers on predetermined scales (Sullivan & Glanz, 2000). Bobbitt (1912) and Cubberly (1916) urged administrators to work toward turning out a standard product using scientific measurement. Supporting a social efficiency model, Bobbitt believed that the educational process was analogous to the manufacturing of steel rails. The standardization philosophy might have worked in a car or steel factory, but did not have sufficient experimentation in public schools. The factory metaphor was encouraged by Ellwood Cubberly in his book *Public School Administration* (1929):

> "Our schools are, in a sense, factories in which raw products (children) are to be shaped and fashioned into products to meet the various demands of twentieth century civilization and is the business of the school to build its pupils according to the specifications laid down" (p. 338).

Based on the factory metaphor, Cubberly laid out a set of principles for school administrators that emphasized measurement and analysis of data to ensure that teachers and schools were productive (Marzano, 2011). This began the process of teachers receiving a grade from an A-F scale to indicate their performance for a variety of criteria. The factory-style method of evaluation has been slightly changed, but still remains the essence of twenty-first century teacher evaluation. Cubberly supported the business factory model because of the enormous success that industry was having around the turn of the nineteenth to twentieth century. Cubberly wrote: "From 1897 to 1907, our country experienced an unprecedented period of industrial development and national prosperity" (Cubberly, 1909).

Democratic methods of supervision emerged in the 1920's with John Dewey, the founder of the progressive education movement, centered on creating and sharing the meaning in society that ultimately leads to growth, influenced supervisors to use cooperative problem-solving approaches that would lead to improving instruction. Prominent superintendent, Jesse Newlon, developed and reinforced democracy in supervision and encouraged "participatory school

management and supervision," (p.7), which created a turning point and change in the purpose of supervision from inspection to improvement of instruction.

During the 1930's and 1940's human relations continued to be more influential on teacher evaluation than the scientific notion that social relations, personal characteristics, and non-instructional school services were the three items most frequently used for rating teachers during that period. Self-evaluation, ceremonial congratulations, and neglect marked the style of teacher evaluation in the 1950's.

Sweeney and Mannatt (1986) proclaimed that the 1960's and 1970's were a time when there was a search for relevance in the classroom and a thirst for individuality and human dignity. During this search for relevance, the notion of supervision as a function of leadership emerged in the 1960's and grew in the 1970's with the concept of clinical supervision (Glanz, 2000). Clinical supervision includes a pre-observation conference, observations, data collection with data analysis, and a post-observation conference where the teacher and supervisor review the collected information and develop a plan for instructional improvement. The final step in clinical supervision is completing a final report and giving feedback. Many of the evaluation systems used today were developed in the 1970's based on the clinical supervision model and have not changed despite educational research that has shaped new theories on how students learn and the teaching strategies often required (Danielson & McGreal, 2000). The shift in teacher evaluation during this time changed the focus to what teachers did to improve basic skills in their students. Sergiovanni and Starratt (2007) emphasized in their work that one of the principal's roles was to focus on the foundational understandings that feed the integrity of teaching and supervision.

The field of education has been impacted by the public's demand for accountability. Popham (1987) stated that educational accountability means that the instructional system designer takes responsibility for achieving the kinds of instructional objectives which were previously explicated. Accountability is addressed through the use of teacher evaluation and over the past decade a number of changes in traditional performance evaluation practices have been proposed as policymakers seek ways to screen out less competent teachers and reward more competent ones.

The 1980's moved teachers into a new role of leadership due to the bureaucratic practices entrenched in public school practices. The work of Madeline Hunter and her colleagues at UCLA contributed to the shift in the evaluation process during this era. Hunter began to investigate using a theory-based way of looking at teaching that had its roots in a behaviorist view of basic learning theory. Hunter developed a set of prescriptive teaching practices designed to enhance student learning. According to Danielson & McGreal, (2000), the Hunter model dominated views of effective teaching throughout the 1980's and became a focus for many state and local evaluation practices. Hunter's theories started a trend toward instructionally focused staff development.

The 1980's also saw changes in teacher evaluation through the mandates of *A Nation At Risk Report*, which emphasized teacher effectiveness and spelled out the minimum requirements to be a teacher. In addition, the teacher licensing decisions imposed through the Educational Testing System (ETS) and the PRAXIS series (professional assessments for beginning teachers) started a new type of "first line of evaluation" for teachers (Danielson, 2007). In 1987 the National Board for Professional Teaching Standards (NBPTS) was created to advance the concept of teaching as a profession and attempted to identify and reward the highest caliber

teachers. The NBPTA developed along lines recommended by the Carnegie Foundation (1986), stressed performance tasks, evidence of knowledge, skills, and disposition, following the development of authentic assessment of pupils. Darling-Hammond (2001) summarized the purposes of those now familiar approaches: "…Such strategies may not only improve the validity of a teacher's assessment, but also support the development of teacher education programs organized explicitly around the attainment of important teaching abilities" (p. 763).

The standards movement of the late 1990's increased expectations for student performance and renewed concerns about teacher practice. The traditional model of teacher evaluation, based on scheduled observations of a handful of direct instructional lessons, came under fire. Peterson (2000) stated that, "Seventy years of empirical research on teacher evaluation shows that current practices do not improve teachers or accurately tell what happens in the classroom" (p. 14). Not surprisingly, in this climate, numerous alternative evaluation practices were developed. One of them was portfolios. According to Darling-Hammond (2001), there is no question among educators that teacher portfolios are valuable tools for professional development across teachers' careers. Darling-Hammond (2004) also stated that portfolios give a better picture of the whole teacher rather than using just test scores, which are, "information for an accountability system, but not the system itself."

As we entered the twenty-first century, research sought to understand if teacher effectiveness was directly related to student performance. Clark (1993) contended that researchers appeared to have taken student achievement for granted; they believed that effective teaching techniques would automatically yield positive student achievement. Effective teachers expect their students to achieve and they believe that all students could learn essential knowledge and skills. During the early part of the twenty-first century, state and local school districts

needed to begin to gain appreciation of just how inefficient the current evaluation systems were and start to investigate how to create data-driven systems to raise performance and make the local school districts aware of the issues surrounding the criticism of standard achievement tests. The most significant impact during this time, and the most recent changes in teacher evaluation, came with the implementation of Race to the Top (RTTT). Dinan (2009) reported that RTTT created, "A new culture for accountability" in teacher evaluation.

Prior to the teacher evaluation mandates of RTTT, teacher evaluations were often subjective and focused on teacher traits and the instructional process (Danielson & McGreal, 2000). Instead of a written evaluation that served as a tenure document that verified a person met the minimum competencies through elements such as voice, appearance, emotional stability, trustworthiness, warmth, and enthusiasm (Danielson & McGreal, 2000), Race To The Top transitioned teacher evaluation from one based primarily on judgment to evidence based evaluation. The focus of education shifted from instructional process to student learning by using teacher evaluation as a tool for professional development, growth, and accountability. The reform in evaluation led to change by putting teachers in, "more active and professional roles" in the evaluation process, (Danielson, 2001, p. 14). Peterson & Peterson (2006) wrote, "Principals can foster good teacher evaluation by knowing the developments of the past twenty-five years, taking initiative to support good data gathering for their own teachers and school, and supporting teachers as they become more involved in their own evaluations" (p. 13). RTTT required states to: Design and implement rigorous, transparent, and fair evaluation systems for teachers and principals the (a) differentiate effectiveness using multiple rating categories that take into account data on student growth as a significant factor, and (b) are designed and developed with teacher and principal involvement; conduct annual evaluations of teachers and principals that

include timely and constructive feedback and provide teachers and principals with data on student growth for their students, classes, and schools; and use these evaluations, at a minimum, to inform decisions (regarding personnel decisions), (USDOE, 2009, p. 9).

RTTT transformed teacher evaluation from one of documentation and compliance to that of self-reflection and a professional learning tool. RTTT also deciphered between observation and evaluation. Classroom observations have been the traditional primary source of evidence used by evaluators in their decisions on teaching performance (Haefele, 1993). Observation is a method or technique that can be used to obtain information for diagnosis and development of instructional practice (MET Project, 2012). Evaluation, on the other hand, is a process that uses various methods in order to obtain data to judge a teachers total teaching performance. When administrators observe teachers, they gain valuable data and feedback, but they don't have enough information to judge a teacher's total performance.

The evolution of supervision in the nation's schools has created a present-day dilemma not easily resolved. The necessity to evaluate teachers is a bureaucratic inspection-type function, while the efforts to assist teachers in improving their practice are a democratic, collaborative function (Sullivan & Glanz, 2006). Supervision and evaluation have been described as two separate functions, each incompatible with the other, yet both remaining the responsibility of the same administrator (Holifield & Cline, 1997). Although research shows that an administrator can be both supervisor and evaluator under specific conditions, when there is a lack of distinction between them on the part of the teacher or the administrator, or when the roles of supervision and evaluation are blurred, the relationship quickly dissolves (Zepeda & Ponticell, 1998).

Federal Education Reform

The history of evaluation transformation also includes numerous federal reform movements. The current models of evaluation have been derived from reform movements that occurred from the 1980's to 2014. Johnson (1997) stated that political viability is the potential political divisions that emerge as a result of competing conceptions of effective teaching versus the policies that lead to their evaluations. Kennedy (2005) noted that reforming schools and improving teaching are the battle cries American education has been echoing for twenty years, but teaching has not changed and too many would-be reformers know nothing about the conflicting demands of teaching and if reformers want teaching to change, they must focus on all aspects of reform, not just those that interest them. Many education reform movements have presented themselves over the past decades to give example to such political viability.

For about thirty years, studies have documented the decline in education and researchers have identified teacher evaluation as one of the specific areas in need of improvement. The primary means for increasing student performance is to have distinguished or highly effective teachers in the classroom. Effectively using teacher evaluation is one method to inform and increase teacher performance, but they must be valid, reliable, and provide meaningful feedback. Waters and Marzano (2006) found in a meta-analysis research project they conducted that when building and district leadership "effectively address specific responsibilities, they have a profound, positive impact on student achievement in their districts" (p.6). A review of literature revealed that researchers have been suggesting the same reforms and practices to address this for over thirty years. The federal government has made attempts at educational reform and, because schools continue to produce students whom are not college and career ready, reform has continued to be the focus of many state and federal reform movements.

A Nation at Risk

In 1981, the National Commission on Excellence in Education (NCEE) was created to investigate the quality of education in the US. "The Commission was created as a result of the Secretary's concern about 'the widespread public perception that something is seriously remiss in our educational system,'" (NCEE, 1983, p. 9). A 1983 report published by the United States Department of Education (USDOE) showed that the Nation was falling behind globally and that our country was not as competitive as it once was (NCEE, 1983). In 1983 the Commission released a report titled *A Nation At Risk: The Imperative for Educational Reform*. The report was advisory in that it defined the problems in our education system. Although the report did not contain legislation, many believe that is was the driving force that has launched the wave of changes we have experienced in education over the past twenty years. In the cover letter of the report, Gardner, Chair of the Committee, wrote "The Commission deeply believes that the problems we have discerned in American education can be both understood and corrected if the people of our country, together with those who have public responsibility in the matter, care enough and are courageous enough to do what is required" (NCEE, 1983, p.6). The report began the trend of the invasion of federal regulation on education reform.

The report's findings were focused on four areas within the educational process: Content, expectations, time, and teaching. The report found that teachers were not prepared appropriately through teacher preparation programs and were unqualified to deliver instruction. In addition, the commission's report emphasized the importance for schools to maintain high standards and challenge students so that they reach their full potential (NCEE, 1983). In addition, the commission noted that the educational system goals were two-fold:

> "The twin goals of equity and high-quality schooling have profound and practical meaning for our economy and society, and we cannot permit one to yield to the other either in principle or practice. To do so would deny young people their chance to learn

> and live according to their aspirations and abilities. It also would lead to a generalized accommodation to mediocrity in our society on the one hand or the creation of an undemocratic elitism on the other," (NCEE, 1983, p.13).

The need to reform teaching, teacher preparation programs, and educational standards were some of the suggestions made in the commission's report. Since the release of this report, the focus of education in relation to student achievement has been on the importance of teacher growth and development and instructional productivity.

No Child Left Behind

The federal government abandoned its advisory role when it exercised mandates on an unprecedented scale with the passing of No Child Left Behind, (NCLB), (Brimley & Garfield, 2008). The 2001 legislation was created as a way to increase student performance through accountability, choice, and flexibility in federally funded educational programs (USDOE, 2002). It placed a new emphasis on the achievement scores of previously overlooked student groups such as students with disabilities, minority students, and students learning English as a second language. NCLB was an ambitious federal education law – which proposed to close achievement gaps and aimed for 100% student proficiency by 2014. In the area of funding, NCLB provided provisions for schools receiving federal dollars by tying the federal funds to student performance. Given the growing body of knowledge that the primary means for increasing student performance is to have distinguished or highly effective teachers in the classroom, legislation such as NCLB began to acknowledge the importance of teachers (Tucker & Stronge, 2005).

NCLB required schools to use a standards-based approach, created by each state's department of education, to assess students annually and measure student achievement, (Tucker & Stronge, 2005). The aggregated results of the annual testing are reported on each district's report card, and parents receive reports consisting of their child's individual scores. NCLB

legislation required that "Assessment results and State progress objectives must be broken out by poverty, race, ethnicity, disability, and limited English proficiency to ensure that no group is left behind" (USDOE, 2002, p. 1). The progress of students is measured against grade level proficiency according to the state learning standards and has a goal that every school rid itself of the achievement gap left by race and social class (Rothstein, 2004).

In addition to student results, NCLB also assessed schools against the proficiency goals set by the state in terms of Adequate Yearly Progress (AYP). If a school fails to reach AYP goals it will fall into one of the following categories (based on the number of years it has not reached AYP): Improvement, corrective action, or restructuring (USDOE, 2002). "The No Child Left Behind Act includes options for the state to take over schools that fail to meet adequate yearly progress" (Brimley & Garfield, 2008, p. 153).

NCLB also focused on putting qualified teachers in every classroom (Tucker & Stronge, 2005). The highly qualified teacher provision of NCLB was fueled by 'report after report' of the dire state of education in the nation's public schools, including the fact that twenty-five percent of teachers were 'not qualified to teach in their subject area'. This was especially at the elementary and middle school levels where a majority of coursework in preparatory programs was focused on methodology and pedagogy rather than content. NCLB required all teachers who taught core content subjects to be highly qualified (Tucker & Stronge, 2005).

The federal government's role in education became more active with NCLB and the mandates the legislation gave began the wave of accountability that enforced using research-based practices to improve education and student achievement. "The requirement of NCLB, some maintain, have brought considerable federal control to education in the twenty-first century" (Brimley & Garfield, 2008, p. 209).

Race to the Top

Race to the Top (RTTT) is the most recent legislation aimed at increasing student achievement (USDOE, 2009). RTTT was part of President Obama's 2009 American Recovery and Investment Act and was designed to improve education by encouraging sustained change in the educational system in order to increase student achievement, close the achievement gap, increase graduation rates, and make students' college and career ready as well as globally competitive. President Obama stated:

> "America will not succeed in the 21st century unless we do a far better job of educating our sons and daughters… And the race starts today. I am issuing a challenge to our nation's governors and school boards, principals and teachers, businesses and non-profits, parents and students: if you set and enforce rigorous and challenging standards and assessments; if you put outstanding teachers at the front of the classroom; if you turn around failing schools – your state can win a Race to the Top grant that will not only help students outcompete workers around the world, but let them fulfill their God-given potential" (State of the Union, 2009).

RTTT focused on the following areas: Designing and implementing rigorous standards and high-quality assessments, by encouraging states to work jointly toward a system of common academic standards that builds toward college and career readiness, and that includes improved assessments designed to measure critical knowledge and higher-order thinking skills; attracting and keeping great teachers and leaders in America's classrooms, by expanding effective support to teachers and principals; reforming and improving teacher preparation; revising teacher evaluation, compensation, and retention policies to encourage and reward effectiveness; and working to ensure that our most talented teachers are placed in the schools and subjects where they are needed the most; supporting data systems that inform decisions and improve instruction, by fully implementing a statewide longitudinal data system, assessing and using data to drive instruction, and making data more accessible to key stakeholders; using innovation and effective approaches to turn-around struggling schools, by asking states to prioritize and transform

persistently low-performing schools; and demonstrating and sustaining education reform, by promoting collaborations between business leaders, educators, and other stakeholders to raise student achievement and close achievement gaps, and by expanding support for high-performing public charter schools, reinvigorating math and science education, and promoting other conditions favorable to innovation and reform (USDOE, 2009).

As educational research has advanced and increased our knowledge of student learning so have the goals for student achievement (Danielson & McGreal, 2000). "Research on teaching and learning has yielded much in recent years. Furthermore, our knowledge of how to develop good-teacher evaluation instruments and procedures has increased (Johnson, 1997, p. 70). With educational reform being on the forefront, many legislatures have become aware of research findings and, as a result, have passed legislation increasing teacher evaluation requirements and mandating teacher evaluation training for certified administrators. Toch (2008) recommends dismantling the system that rewards credentials with a single salary schedule and replacing it with a system that rewards performance based on instruction – which requires a credible evaluation system that would be one that: Teachers trust, places less than 50% weight on test scores, and includes a performance pay incentive.

Georgia Education Reform

The federal reform movements mentioned above has been cause for political debate over the topic in the state of Georgia for decades. In most public opinion surveys, Georgians rank improving public education as a primary concern for governmental action. Issues of reform are complicated by competing agendas in the Georgia General Assembly and by differences in the goals advanced by teachers, politicians, and parents. For most of the state's history, education has been a local function, with local school boards making decisions regarding the funding and

operations of schools at the city or county level. Over time the state has assumed more responsibility.

The Politics and Policy Project (2014) provides a summary of education reform in Georgia. Georgia's first major statewide education initiative came in 1916, when the General Assembly enacted legislation that made school attendance compulsory for all children between the ages of eight and fourteen. In 1937 a state Board of Education was established to oversee and coordinate the delivery of education. In the 1950's and 1960's, the issue of education and education policy revolved around attempts to resist federally mandated integration and by the 1970's K-12 education took the largest share of the state's budgetary outlay.

Major educational reform legislation came to Georgia in 1985 with the enactment of Governor Joe Frank Harris's Quality Basic Education Act (QBE) to address the inequality in funding among school systems in the state. Before QBE, school systems received state allocations on the basis of the number of students enrolled, without any adjustment for the fiscal condition of the school system or its ability to raise revenues on its own. Rural systems could not generate as much local funding as suburban districts could and by the early 1980's urban school systems with shrinking tax bases were having difficulty keeping funding at already established levels. Legislators from both urban and rural counties called for funding equalization. QBE was introduced in the state senate by Roy Barnes and it increased the total amount of money appropriated for K-12 education. Under the "local fair share" provision, additional state funds were given to school districts that increased local funding. The act established minimum salary levels for educators and merit pay incentives for outstanding teachers. QBE raised the professional standards for teacher certification and funded continuing-education opportunities for teachers already in the field. Finally, QBE established the Quality

Core Curriculum (QCC), which set guidelines for the specific material to be taught at each grade level.

In his 1990 gubernatorial campaign, Zell Miller promised to institute a state lottery, with the proceeds going directly to fund new educational programs. The legislation for the lottery is very specific about what programs can be funded by proceeds from the lottery. These include offering college scholarships to students who maintain a B average (HOPE Scholarship s), extending Head Start and prekindergarten programs, and developing physical and technological infrastructures in both K-12 and postsecondary institutions throughout the state. These initiatives have made Georgia a leader among states in educational innovations, and the HOPE Scholarship was proposed as the inspiration for a similar federal program.

The A-Plus Education Reform Act of 2000 was crafted by Governor Roy Barnes on the basis of a report issued by a special task force in his first year in office. It provided for an independent office of accountability charged with developing testing standards for all K-12 students and coordinating the distribution of information on test results by issuing report cards on schools and school systems through the mandates of NCLB. The age for compulsory school attendance was lowered from eight to six and children in grades one through three were targeted for early intervention programs. All schools were also required to have advisory councils of parents, administrators, and community members. In addition, teachers were directly affected with salary enhancements guaranteed to teachers who completed national certification and certification was revoked for teachers who receive unsatisfactory ratings for any two years in a five-year span. In perhaps the most controversial aspect of the act, tenure for public school teachers was eliminated. As noted by Marshall (2005), this effort was Georgia's first real attempt at holding educators accountable for student achievement and faced political resistance

in the early stages of development. Moreover, in the early stages of reform, Georgia struggled with assessments not aligned to the curricular standards. Consequently, looking at the process of developing Georgia's response to NCLB, Marshall suggested that "the instability of the political environment and a lack of policy coordination…created obstacles in developing a coherent accountability system" (p. 10).

In 2002 an audit of the QCC conducted by Phi Delta Kappa International found that the curriculum did not meet national standards and could not be completed in twelve years. As a result, state teachers and other education experts developed a new curriculum, the Georgia Performance Standards (GPS), which was implemented in stages beginning in 2005. GPS increased the depth of coverage across content areas and provided instructors with suggested tasks, samples of student work, and commentary by other teachers. GPS also focused on the skills that students must develop in order to complete the curriculum and defined the expectations for acceptable assessment, instruction, and student work. Teachers received training in the year prior to implementation, as well as during the first year of implementation. The first subject areas to transition from QCC to GPS were English, math, science, and social studies.

Although noted difficulties occurred at the onset of NCLB, efforts a clear accountability plan for state educators to follow was submitted in the final plan to the USDOE, Georgia outlined details described in Georgia law, State Board of Education policy, and the *Georgia Consolidated State Application Accountability Workbook* (2010). Included as a component of Georgia's accountability plan was a piece of legislation dealing with the annual performance evaluation of state educators. The law acknowledges the relationship of school personnel and student achievement scores by stating: Annual teacher evaluation shall at a minimum take into

consideration the role of the teacher in meeting the school's student achievement goals, including the academic gains of students assigned to the teacher. Although the evaluation tool in Georgia had not changed with the creation of NCLB, the connection between teachers and student scores clearly came to the forefront with the law's enactment.

The National Race to the Top (RTTT) education law of 2009, changed teacher evaluation in Georgia. Georgia vowed:

> "To equip all Georgia students, through effective teachers and leaders and through creating the right conditions in Georgia's schools and classrooms, with the knowledge and skills to empower them to 1) graduate from high school, 2) be successful in college and/or professional careers, and 3) be competitive with their peers throughout the United States and the world" (GADOE, 2014).

Georgia's application for RTTT was prepared through a partnership between the Governor's Office, the Georgia Department of Education, and the Governor's Office of Student Achievement and education stakeholders. Four working groups and a fifth critical feedback team consisting of teachers, principals, superintendents, higher education faculty, non–profit and informal education organizations, state policy makers, and members of the business and philanthropic communities developed the ideas for inclusion in the state's winning application. Georgia was awarded $400 million to implement its Race to the Top plan and the State Board of Education has direct accountability for the grant.

The U.S. Department of Education released Georgia's Race to the Top grant progress report for education reform and in Years 1 and 2, Georgia made several education reforms including forming partnerships to improve teacher and leader effectiveness. There were delays and problems with the teacher evaluation system, which prompted federal officials to place the educator evaluation project on high-risk status. During year three federal officials praised Georgia for the launch of its Innovation in Teaching Competition and for piloting a professional

learning plan on using assessments to improve instruction. Georgia remains on high-risk status because of the teacher and leader evaluation systems (TKES and LKES), primarily due to problems with Student Learning Objectives in non-tested grades and subject areas. Georgia has tabled plans to implement the performance-based compensation system described in its approved application within the grant period.

Susan Andrews, Georgia Department of Education Deputy Superintendent of Race to Top Implementation stated "While educational reform work is never easy, we are pleased with the significant progress that our state has made as reflected in this report. We have taken on some major initiatives and our success is a direct reflection of the collaborative efforts between our department, other education agencies, and the hundreds of participating teachers, principals, and district leaders around our state" (2014). Georgia and other states in America have chosen models of evaluation, which define what good teaching is, to reform their evaluation systems.

Models of Evaluation

The response to education reform from scholars and experts in the field has been the development of different evaluation models that help identify the characteristics, qualities, and roles of great teachers. Jacob (2012) noted that effective schools emphasized teacher quality as a critical input for improving student outcomes. Teacher certification status, years of experience, and possession of a graduate degree are three factors that were found to be commonly taken into consideration when assessing teacher quality. Research shows that the best way to increase student achievement is to have highly effective teachers in the classroom (Clifford, Behrstock-Sherratt & Fetters, 2012). There have been many studies conducted by leading experts, such as Danielson and McGreal (2000), Marzano (2007), and Reeves (2004), identifying what skills are essential to be a highly effective educator. A highly effective teacher, as defined in the RTTT executive summary, is:

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

Many educators have discussed the complex issues surrounding the problem of identifying the categories and criteria that exemplify effective teaching. The categories and criteria must conform to state statutory mandates, state and local board policies, and negotiated agreements. They do not always lead to a successful evaluation system, but they do represent fairly common perceptions of what makes good teaching as well as the essence of appropriate nonteaching responsibilities. In theory criteria can identify weak spots in the techniques and style used by the classroom teacher. Criteria are most commonly defined as: Accountability, assessment, monitoring, evaluation, or guides to improvement and send a message to teachers, evaluators, and the public as to what is important.

The response mode is the way the evaluator is asked to respond of the evaluation instrument and there are two basic types: Free-response and structured-response. The evaluator produces the response in the form of an essay or narrative in a free-response mode and the answer is selected from a group within a rating scale or checklist in the structured-response mode. When evaluating teachers, the rating scale is not a measurement tool, but rather, a means of placing value on the performance of an educator and the ratings are observations that have been categorized or organized to provide summary information about the behavior of an individual. Several different evaluation models use ratings in different ways to categorize the criteria believed to identify an effective, "good", teacher.

The Hunter Model

The evaluation of teaching relates to the opinion of what constitutes good teaching. A familiar model that defines good teaching is the Hunter Model. Toward the end of the twentieth century, American education was greatly influenced by Madeline Hunter. She is most famous for her model lesson design that included the elements of: Anticipatory set, objective/purpose, input, modeling, checking for understanding, guided practice, and independent practice. Her "Instructional Theory into Practice" (ITIP) model is widely used to plan instruction and define good teaching. In her 1994 book, Enhancing Teaching, Hunter defined three categories to link good teaching and lesson design: Content, learner behavior, and teacher behavior (Hunter, 1994). Professional freedom and creativity is accessible to teachers when using this model and evaluators can observe individual teaching and learning styles in specific classrooms. In regards to evaluating good teaching, Hunter contributed to the field by expanding on observations with checklists and introducing script taping. During script taping, an evaluator recorded teaching behavior and then categorized them into those that: Promoted learning, used precious time and energy but did not contribute to learning, or unintentionally interfered with learning, (Hunter, 1994).

The Danielson Model

Another prominent model that helps define good teaching is the Danielson Model. Charlotte Danielson is a career educator who has taught all levels from kindergarten through college. She holds degrees from Cornell, Oxford, and Rutgers and is highly respected for her work in curriculum planning, performance assessment, professional development, and teacher growth. She wrote about supervision and evaluation in *Enhancing Professional Practice: A Framework for Teaching* (1996) and this book was widely used in the late 1990's and early 2000's before the legislative overhaul of teacher evaluation took place. Danielson's emphasis on

teacher evaluation is to provide a clear definition of the "what" of teaching and then have instruments and procedures that provide evidence of the "how" of teaching (Danielson, 2007). She stated that "Teaching is a fulfillment of public trust and we need to assess teacher effectiveness" (2007). The Danielson Model is a multi-tier evaluation system that includes four domains: Planning and preparation, the classroom environment, instruction, and professional responsibilities. In each domain teachers receive a score of Unsatisfactory, Basic, Proficient, or Distinguished. The Danielson model establishes a common language for professional conversation and the structure for self-assessment and reflection on professional practice (Marzano, 2011). There are seventy-six elements that are broken into four levels of performance evaluation and teachers are given a grade ranging from unsatisfactory to proficient and, ultimately, distinguished performance.

The McREL

The McREL teacher evaluation instrument is a web-based, formative, rubric-driven evaluation system created collaboratively by teachers, principals, and researchers. The model reflects the latest research on effective instruction and aligns with the Model Core Teaching Standards adopted by the Council of Chief State School Officers and has been used for over three years. In the McREL publication, *Simply Better: Doing What Matters Most to Change the Odds for Student Success* (Goodwin, 2011) there are three key characteristics of highly effective teachers defined and measured on the instrument:

1. Highly effective teachers challenge their students with high expectations that develop higher-order thinking skills and take personal responsibility for student progress while encouraging students to be responsible for their own learning.
2. Highly effective teachers create positive classroom environments and develop strong relationships with their students by understanding the learning needs and various

backgrounds of their students and creating respectful and inclusive classroom environments that encourage all students, regardless of background or ability, to contribute to classroom discourse.

3. Highly effective teachers are intentional about their teaching by being clear about what they're trying to teach while mastering a broad repertoire of instructional strategies to help students accomplish their learning goals. They have a deep knowledge of the content and how to teach it while continually monitoring student progress and using classroom assessment data to adapt instruction to student needs.

The Stronge Teacher Evaluation System

James H. Stronge is the Heritage Professor in the Educational Policy, Planning, and Leadership Area at the College of William and Mary, Williamsburg, Virginia. He has worked with numerous school districts and other educational organizations to design and implement evaluation systems for teachers, administrators, and support personnel. His work on effective teachers focuses on how to identify effective teachers and how to enhance teacher effectiveness. The Stronge Teacher Evaluation System is designed with Performance Standards that serve as the job responsibilities or duties performed by a teacher. The framework for the system begins with an understanding of what effective teaching is and is designed and implemented from the empirical research on teacher effectiveness that is synthesized in *Stronge's Qualities of Effective Teachers* (2007). There are a set of standards that retain the diagnostic profile of teacher performance, with seven teacher performance standards:

1. Performance Standard 1: Professional Knowledge - The teacher has an understanding of the curriculum, subject content, pedagogical knowledge, and the developmental needs of students.

2. Performance Standard 2: Instructional Planning - The teacher plans using the state standards, district curriculum, effective strategies, resources, and data.
3. Performance Standard 3: Instructional Delivery - The teacher uses a variety of effective instructional strategies in order to meet individual learning needs.
4. Performance Standard 4: Assessment of/for Learning - The teacher uses a variety of formative and summative assessment strategies and data.
5. Performance Standard 5: Learning Environment - The teacher provides a well-managed, safe, student-centered, academic environment that is conducive to learning.
6. Performance Standard 6: Professionalism and Communication - The teacher maintains a commitment to professional ethics and professional growth and effective communication with all stakeholders.
7. Performance Standard 7: Student Progress - The instructional efforts of the teacher result in acceptable, measurable student progress based on established standards and goals.

James Stronge was the project consultant during the development of TKES. His model of teacher effectiveness was used to guide the addition of components to TKES and plan for implementation.

The Marzano Model

The Marzano Causal Teacher Evaluation Model identifies a direct cause and effect relationship between elements in the model and student learning. Robert Marzano is a well-known and influential writer, speaker, trainer, and researcher in education. He has authored over thirty books and hundreds of articles related to topics in education. The model provides teachers with clear expectations that enable teachers to direct their own professional learning. Inherent within the model is aligned professional development and feedback to support teacher growth and development. The model is based on an aggregation of the research on elements that have

been shown to correlate with student academic achievement and is based on thousands of studies that span multiple decades. The model defines an effective teacher as one who makes instructional decisions that produce student learning gains and includes a well-articulated knowledge base in teaching with four domains directly tied to student achievement through the following domains:

1. Domain 1: Classroom Strategies and Behaviors
2. Domain 2: Preparing and Planning
3. Domain 3: Reflecting on Teaching
4. Domain 4: Collegiality and Professionalism

The four interdependent domains include 60 elements and the model is enhanced by clear descriptions of teaching and levels of performance using rating scales ranging from 0- Not using 1- Beginning, 2-Developing, 3-Applying, and 4-Innovating, which enable teachers and observers to engage in focused practice and feedback. Marzano's model is top heavy with the Domain 1 (Classroom Strategies and Behaviors) having more evaluation rubrics than other domains. Classroom Strategies and Behaviors has thirty-nine rubrics for principals to evaluate teachers while Preparing and Planning has eight categories, Reflecting on Teaching has five, and Collegiality and Professionalism has six.

Marzano's model also includes the Stages of Development guidelines to address the issue as to whether veteran teachers should be evaluated the same as beginning teachers. Teachers are divided into stages of: Initial Status, Professional, Mentor, and Master and there are benchmarks for each stage.

Components of Teacher Evaluation

Although the details vary, there are several components of teacher evaluation that are included in all of the above models and, whichever one a state chooses to implement, the effectiveness of the model can be rated after considering each one separately and collectively. "A good teacher evaluation system must define the teaching task and provide a mechanism for judging the teacher (Wise et al., 1985, p. 65). There must be a clear understanding of what good teaching is, what the evidence of good teaching is, and how it will be measured. The new evaluation system, mandated through RTTT, incorporates multiple measures of effectiveness to determine a composite score that is then used to differentiate teacher effectiveness using four rating categories (USDOE, 2009). Although there are other potential uses of the teacher evaluation process, such as personnel decisions, the teacher evaluation process is most often cited as a way to enable schools to assist teachers in the improvement of their skills by assessing teachers' strengths and weaknesses to determine areas needed for staff development and professional growth. The expectation of teacher evaluation is that it will recognize good teaching and enhance the quality of teaching and lead to more effective student learning and achievement.

Goe et.al (2008) provided a synthesis of research on considerations for implementation on the following seven measures of teacher effectiveness: Classroom observation, principal evaluation, instructional artifacts, portfolio, self-evaluation, student surveys, and value-added test score data. It concluded that a broad definition of teacher effectiveness is important and measures must be varied according to school, district, and state need. Cashin (1996) gave the following recommendations for a successful evaluation system: Clarify institutional goals, decide on the purposes of the data to be collected, use pilot programs when appropriate, significantly involve participants in the department, foster open communication, obtain support

from high level officials, ensure that the system is flexible and legal, define major and minor faculty responsibilities, use multiple data sources, ensure that the data are technically acceptable, define the criteria, train the evaluators, maintain confidentiality, reward effective performance, and review the system periodically.

Evaluations, as defined by Stronge (1991), are a reflection of human performance in organizational settings. Guidelines included under Race To The Top (RTTT) were designed to reform education. In RTTT evaluation is not defined, but instead only the components of evaluation and what it can be used for are addressed (U.S. Department of Education, 2009). The goal of using evaluation as a reflection of one's performance remains the same, but the objectives have changed over time. Evaluation was once a form of documentation and now it is seen as a tool to help teachers improve their practice and support personnel decision making. Now a central component of instructional improvement, evaluation should be a rigorous performance measurement that provides useful feedback (Shakman et al., 2012).

There are many opinions as to what components a teacher's evaluation should include. Kim Marshall, a longtime principal, offered several tips in his 2005 research on teacher evaluation, including: Having a common language of what constitutes good teaching, conducting multiple observations of five to fifteen minutes over the course of the year, giving prompt face-to-face feedback, using a rubric to measure effectiveness, and incorporating some type of teacher portfolio. Marshall (2012) added to the research that using student achievement and feedback from students are important for teacher evaluation, but only when used wisely. According to Kim, student achievement was recommended to be used only if developed locally and after being approved by the principal and student input should only be done if teachers have the ability to review and defend input before reporting.

Peterson (2000) noted that teacher evaluation should take a new approach and incorporate multiple data sources including: Student and parent reports, peer review of materials, student achievement, teacher tests, documentation of professional activity, systematic observation, and administrator reports.

Types of Evaluation

There are two types of evaluation that the components of each model may be categorized: Formative and summative evaluation. Teacher evaluation has two purposes: The improvement of teachers' skills so that they can become more effective (formative evaluation) and evaluation focused on changes in status such as granting tenure, placement on probation, dismissal actions (summative evaluation). It is a mistake to think that one purely formative or summative system can serve the purpose of growth, accountability, school improvement, and personnel decisions and most school districts today are moving toward changing their evaluation instruments to be both formative and summative.

Formative evaluation is typically described as a practice designed to provide teachers with data that can be used to improve the quality of instructional practices and summative evaluation is meant to serve as a data gathering process to evaluate overall performance for an entire school year. Formative evaluation includes walkthroughs and observations, which are accompanied with feedback. Scriven (1988) said that the purposes of personnel evaluation are accountability and professional growth. For personnel evaluation systems to be productive the imbalance between the two should be remedied. The performance improvement dimension reflects the need for professional growth and development of the individual and typically has been considered to be formative. In order for evaluation systems to be feasible there needs to be a rational link between the purposes. Three descriptive studies have examined teacher evaluations (Brandt et al., 2007) and these studies indicated that teacher evaluations were

focused more often on summative goals; they were used more for dismissal and remediation purposes than as part of individualized personal growth to lead to improved teacher, and thus student, performance in the classroom. Most policies on teacher evaluation before RTTT did not include requirements for establishing performance standards and evaluator training (Mathers, 2008). Throughout the years there have been consistent components of teacher evaluation tools that have led to the development of our current instruments.

Observations

Observation is a key component of science, but research shows that teacher observation is flawed. In a 1952 study of teacher observation, Matthew Whitehead surveyed teachers regarding their perceptions as to the importance of each area of supervision (Marzano, 2011). The conclusion to his study pointed to the fact that observational practices failed to capitalize on its potential. Observations did not last an entire period and rarely were there follow up conferences to discuss the evaluation. Whitehead sought to provide teachers a voice concerning how they are evaluated. The report exposed teacher attitudes toward instructional improvement and to describe supervisory practices (Whitehead, 1952). The study used questionnaires given to 115 teachers at multiple schools. The study found that 80% of teachers looked at the classroom visitation by the principal as a method of improving instruction while 20% stated that the purpose of the principal visits were to give ratings to the teachers and to inspect the physical features of the classroom (Whitehead, 1952).

In an article titled *It's Time to Rethink Teacher Supervision and Evaluation*, Kim Marshall (2005), argued that most processes that evaluate teachers are ineffective, inefficient, and a poor use of principal's time. Marshall provides examples of principals observing less than 0.1% and when they do observe, the results are skewed because of an over glamorized lesson or the fact that discipline problems are reduced with an extra authoritative presence. Marshall also

noticed that teachers act atypical because they are nervous while being observed. Marshall found that a lack of data during the observation caused most evaluators to guess as to whether students were really learning or not.

Walkthroughs

Walkthroughs began in the 1960's with administrator Carolyn Downey because she was asked to be more visible in the classroom. During the walkthroughs she realized that they helped her, "get a big picture of the learning environment," and she, "saw how much one could learn from walkthroughs." In addition, they allowed her to show her staff she cared about them and she "came to realize that walkthroughs, along with meaningful dialogue, was an effective approach to help the teacher grow professionally" (Downey, Steffy, English, Frase & Poston, 2004, p. 10). The Downey walkthrough model was created from the information and professional development she did.

Walkthroughs are short and focused observations that range from three to ten minutes long and give the evaluator a snapshot of what is going on in the classroom. "The idea behind walkthroughs is that firsthand observations can paint a picture to inform improvement efforts," (David, 2008, p. 81). Research on walkthroughs is limited, but available studies reveal "Wide variation in their usefulness and effects" (David, 2008, p. 82). Frase and Hetzel (1990) published *School Management by Wandering Around* and followed business models to compare the benefits of having the principal wander throughout the school. The effectiveness of a walkthrough depends on what the focus or target is. Pitler and Goodwin (2008) outline six questions that principals should ask when conducting a walkthrough: Are the teachers using research based teaching strategies? Do student grouping patterns support learning? Are teachers and students using technology to support student learning? Do students understand their goals for learning? Are students learning both basic and higher order level of knowledge? Do student

achievement data correlate with walkthrough data? (p. 9). Walkthroughs can be crucial to improvement, however "only when districts make their purpose clear and carry them out in a climate of trust" (David, 2008, p. 89).

The new evaluation systems that are in place due to the mandates of RTTT require numerous walkthroughs and communication between the evaluator and teacher. The age of accountability is being tested with these new systems and there is not proper support in place to be effective. The opinion of this author is if walkthrough observation quotas are met, the workload of evaluators and stress level for teachers will become unbelievable.

Feedback

"Feedback can be said to describe any communication or procedure given to inform a learner of the accuracy of a response, usually to an instructional question," (Carter, 1984, p. 53). Feedback should be purposeful and deliberate towards the components of the evaluation tool that are being evaluated. Feedback should also be ongoing in order to improve teacher and student performance. "Evaluations should provide all teachers with regular feedback that helps them grow as professionals, no matter how long they have been in the classroom" (The New Teacher Project, 2010, p. 3). For example, if a principal has sixty teachers and one Assistant Principal each will evaluate approximately thirty teachers. The requirement for TKES are four ten-minute walkthroughs, two thirty-minute observations, two formative assessment and one summative assessment per year, per teacher, with feedback required on each; approximately 180 minutes per teacher.

Feedback can be written or verbal, one-sided or two-sided, and immediate or delayed. The two most effective attributes of feedback to promote teacher change are the nature and the timing of the feedback. Feedback should be immediate, specific, positive, and corrective (Scheeler, et. al, 2004). Hattie and Timperley (2007) identified a framework that defined four

types of feedback that motivate learning in students, but can be transferred to evaluators and teachers: Task level, process level, self-regulation, and self-level. Task level feedback is rooted in the work or product; process level is directed at the process of learning; self-regulation is the ability of the learning (teacher) to self-check after learning prompts are given; and self-level is when positive comments are given, but there is no learning connected to them. Feedback can also be descriptive (used to move the learner forward in the learning process), evaluative (used during summative evaluation), or motivational (used to give praise), (Hattie & Timperley, 2007). It has been noted that daily and weekly feedback as opposed to monthly feedback is more effective in an organization and studies show that learning takes place when feedback is given more frequently (Scheeler, 2004). Feeney (2007) stressed that the goal of feedback is to improve the effectiveness of teaching and professional growth. The feedback received should be seen as a formative process that, "Identify opportunities for improvement," (Pitler & Goodwin, 2008). A structure to promote reflective inquiry and conversations for facilitating the learning of teachers is also a critical form of feedback.

Feedback is a critical part of the evaluation process, however administrators may not feel equipped to give such purposeful feedback on evaluations and evaluators often lack specific knowledge about the content areas in which they evaluate teachers. Hattie & Timperley (2007, p. 81) stated that very few studies have "Systematically investigated the meaning of feedback in the classroom". Feedback for teachers should be ongoing and individualized based on a teacher's needs and ability to understand the information given. One reason why principals may feel that they are not providing effective feedback is because the purpose of the walkthrough or observation has not been defined.

Feedback needs to be focused and deliberate on a specific goal that is individualized for each teacher. Dunkleberger (1982) stated that teachers see feedback as one dimensional rating scales used for hiring or firing. According to Frase "Feedback has too often been inaccurate, shallow, and at times mean spirited, rather than helpful and uplifting" (1992, p. 179). Effective feedback can lead to improved teaching and student achievement. Hattie and Timperley (2007) found that when feedback is directed at the right level it can assist students to comprehend, engage, or develop effective strategies to process the information intended to be learned.

Developing the skills to provide effective feedback takes training and practice. Evaluators must be able to ask questions, tell the truth, and encourage feedback from members of the community, (Bolman & Deal, 2003).

Student Performance

Evaluation is an integral and natural component of the learning process and it is imperative for teachers to remember the importance of assessment while being evaluated (Benedict et.al, 2013). However, it is a philosophical issue to consider is whether student achievement should be included in a teacher's evaluation. A teacher only has a limited amount of time to change the study habits and skills of a student and there are many variables within the group of students that can interfere with the teacher doing his or her job to the best of their ability. However, if a teacher spends a full year with a group of students and at the end of the year a bulk of the students have not mastered the basic material necessary to advance them to the next grade level, citizens should stop paying the taxes that support that worthless teacher's endeavors for the next year. The literature backed by teacher unions opposes linking teacher evaluation to student achievement while the literature from the business community supports it. While it is understood that student learning is the purpose of school, educators must use caution in this phase of teacher evaluation because the use of student test scores brings up the question of

equity and reliability in terms of student, parent, or teacher responsibility because, as Ballard and Bates (2008) noted “It is unclear exactly who is responsible for test performance”.

Schools are different in regards to demographics and funding and the high-stakes nature of standardized testing has become very stressful for students and teachers. Jacob (2012) assessed the potential merits of value-added measurement tools, such as standardized testing results, to assist principals in developing teacher talent and making staffing assignment, development, and retention decisions through evaluations and found inconclusive evidence.

The claim for value-added measures is that they capture how much students learn during the school year, thereby putting teachers on a more level playing field as they aim for tenure or additional pay. David (2008) stated that end-of-year test scores do not show how much students learned that year in that class, so measures that take into account where students started are surely an improvement. However, such measures of growth are only a starting point. Making judgments about individual teachers requires sophisticated analyses to sort out how much growth is probably caused by the teacher and how much is caused by other factors. For example, students who are frequently absent tend to have lower scores regardless of the quality of their teacher, so it is vital to take into account how many days students are present. Thus, to be fair and to provide trustworthy estimates of teacher effectiveness, value-added measures require complicated formulas that take into account as many influences on student achievement as possible. Using test scores for teacher effectiveness might be done properly, but it is still an early component to teacher evaluation systems. TKES uses value-added measurements as part of the over-all summative assessment for teachers, but pay is not currently connected to the result.

Surveys

Universities routinely survey students on how professors are performing, but until recently, K-12 students have not been given the chance to evaluate their teachers. This is

because, although students spend hundreds more hours in classrooms than any administrator, people doubt that students can provide valid, reliable, and stable responses about the quality of teaching (Ferguson, 2012). The Measures of Effective Teaching (MET) project has put those doubts to rest. Comparing value-added analysis of test scores, classroom observations, and student perception surveys (using Ferguson's Tripod questions), researchers have found that students provide accurate, helpful information on their teachers' performance. "Students know good instruction when they experience it as well as when they do not" (Ferguson, 2012).

Over the past decade, Ferguson and his colleagues have collected almost a million K-12 anonymous Tripod surveys on their teachers and research refined the questions to group the questions in seven categories that are rated based on agreement or disagreement with each statement on a 5-4-3-2-1 scale. The categories are:

1. Care. This goes beyond a teacher's "niceness" to encompass demonstrated concern for students' happiness and success. A sample question: My teacher really tries to understand how students feel about things.
2. Control. These questions measure management of off-task and disruptive behaviors in the classroom. A sample question: Our class stays busy and doesn't waste time.
3. Clarify. This addresses the teacher's skill at promoting understanding, clearing up confusion and misconceptions, differentiating, and helping students persevere. A sample question: My teacher has several good ways to explain each topic that we cover in this class.
4. Challenge. This covers effort and rigor and measures whether the teacher pushes students to work hard and think deeply. Sample questions: In this class, my teacher accepts

nothing less than our full effort and my teacher wants us to use our thinking skills, not just memorize things.

5. Captivate. Do teachers make instruction stimulating, relevant, and memorable? Sample questions: My teacher makes lessons interesting and I often feel like this class has nothing to do with real life outside school.
6. Confer. This covers teachers seeking students' points of view and allowing them to express themselves and exchange ideas with classmates. A sample question: My teacher gives us time to explain our ideas.
7. Consolidate. This measures whether teachers check for understanding and help students see patterns and move learning into long-term memory. A sample question: My teacher takes the time to summarize what we learn each day.

The MET study showed that student surveys should be a central component of teacher evaluation and that student judgments provide valuable insight as to how well a teacher is supporting and communicating with his or her students. "Only recently have many policymakers and practitioners come to recognize that --when asked the right questions, in the right ways -- students can be an important source of information on the quality of teaching and the learning environment in individual classrooms".

Checklists

Checklists are used to aid an evaluator in consistently comparing and rating criteria for personnel decisions (Burke & Lind, 1987). Checklists can be general, activities related, or lists of standard items. A checklist is an enumeration of a number of behaviors or features that constitute procedure or product (Thorson, Miller & Bellon, 1987). When using a behavior checklist, the evaluator is provided with a series of descriptive statements of job-related

behavior. Checklists usually contain lists of behaviors, traits, or characteristics that are either present or absent. A simple checkmark shows compliance in specific, job related areas. Mielke and Frontier (2012) noted that using frameworks to evaluate teachers with checklists is not enough to improve teacher performance as a summative assessment, but can be used as a formative assessment to help guide teachers in improvement towards their final evaluation. Moreover, they stated that a checklist should be used for personal development and growth of teachers through self-reflection and feedback given from observations by evaluators.

Teacher Keys Effectiveness System (TKES)

As part of the Race to the Top Initiative (RT3) in 2012-13, Georgia conducted a full year implementation of the Teacher Keys Effectiveness System (TKES), a common evaluation system allowing the state to ensure consistency and comparability across districts, and which was based on a common definition of teacher effectiveness, provided by project consultant, James Stronge. The primary purposes of TKES are to: Optimize student learning and growth, improve the quality of instruction by ensuring accountability for classroom performance and teacher effectiveness, contribute to successful achievement of the goals and objectives defined in the vision, mission, and goals of Georgia Public Schools, provide a basis for instructional improvement through productive teacher performance appraisal and professional growth, and implement a performance evaluation system that promotes collaboration between the teacher and evaluator and promotes self-growth, instructional effectiveness, and improvement of overall job performance. The new system for providing teachers and leaders with meaningful feedback and support opportunities was developed with clear and precise indicators and resources to guide teachers and leaders toward improved performance (TKES Handbook, 2013).

The TKES theoretical framework, which was used to guide the research questions for this study, is a theory of action. The theory of action for TKES states: If educators have specific

performance standards for effective teaching and, if educators are provided professional learning support to develop classroom behaviors that meet the performance standards then the professional capacity of teachers to positively impact student learning will increase. Also, teachers will hold higher expectations for student learning and students will learn more and achieve higher levels of success.

According to the Georgia Department of Education (2012) teacher keys evaluation system manual, the Teacher Keys Effectiveness System (TKES) is comprised of three components which contribute to an overall Teacher Effectiveness Measure (TEM): Teacher Assessment on Performance Standards (TAPS), Surveys of Instructional Practice (student perception surveys), and Student Growth (SGP and SLO). The Annual Evaluation Notification states:

> "Official Code of Georgia 20-2-210 requires that each Local Education Association (LEA) shall provide written notice in advance of each school year to each teacher of record of the evaluation measures and any specific indicators that will be used for evaluation purposes. This notification must be completed in the advance of each school year. In order to comply with this it is suggested that LEAs address this in the contract or an attached document disseminated with the contract".

Teacher Assessment on Performance Standards (TAPS)

A fair and comprehensive evaluation system provides sufficient detail and accuracy so that both teachers and evaluators will fully understand their job expectations. Clearly defined professional responsibilities for teachers constitute the foundation for TAPS. Evaluators shall be appropriately trained and credentialed. The Teacher Assessment on Performance Standards (TAPS) component of the Teacher Keys Effectiveness System (TKES) provides evaluators with a qualitative, rubrics-based evaluation method by which they can measure teacher performance related to quality performance standards. TAPS is a three-tiered approach, which defines the expectations for teacher performance consisting of 5 domains, 10 Performance Standards:

1. Professional Knowledge: The teacher demonstrates an understanding of the curriculum, subject content, pedagogical knowledge, and the needs of students by providing relevant learning experiences.

2. Instructional Planning: The teacher plans using state and local school district curricula and standards, effective strategies, resources, and data to address the differentiated needs of all students.

3. Instructional Strategies: The teacher promotes student learning by using research-based instructional strategies relevant to the content area to engage students in active learning and to facilitate the students' acquisition of key knowledge and skills.

4. Differentiated Instruction: The teacher challenges and supports each student's learning by providing appropriate content and developing skills which address individual learning differences.

5. Assessment Strategies: The teacher systematically chooses a variety of diagnostic, formative, and summative assessment strategies and instruments that are valid and appropriate for the content and student population.

6. Assessment Uses: The teacher systematically gathers, analyzes, and uses relevant data to measure student progress, to inform instructional content and delivery methods, and to provide timely and constructive feedback to both students and parents.

7. Positive Learning Environment: The teacher provides a well-managed, safe, and orderly environment that is conducive to learning and encourages respect for all.

8. Academically Challenging Environment: The teacher creates a student-centered, academic environment in which teaching and learning occur at high levels and students are self-directed learners.

9. Professionalism: The teacher exhibits a commitment to professional ethics and the school's mission, participates in professional growth opportunities to support student learning, and contributes to the profession.

10. Communication: The teacher communicates effectively with students, parents or guardians, district and school personnel, and other stakeholders in ways that enhance student learning.

Performance Indicators

Performance indicators provide examples of observable, tangible behaviors for each standard. The performance indicators are examples of the types of performance that will occur if a standard is being successfully met. While it is likely the evaluator will observe many of the indicators during an observation, the list of performance indicators is not exhaustive. The evaluator may also observe many appropriate performance indicators that are not listed. The complete list of TAPS Standards and Indicators can be found in the Appendix.

Performance Appraisal Rubrics

In addition, teachers shall be rated using the performance appraisal rubrics. The performance rubric is a behavioral summary scale that guides evaluators in assessing how well a standard is performed. It states the measure of performance expected of teachers and provides a qualitative description of performance at each level. The resulting performance appraisal rubric provides a clearly delineated step-wise progression, moving from highest to lowest levels of performance. Each level is intended to be qualitatively superior to all lower levels. The description provided for Level III of the performance appraisal rubric is the actual performance standard, thus Level III is the expected level of performance. Teachers who earn a Level IV rating must meet the requirements for Level III and go beyond.

Observations and Documentation

"Most authors identify the fundamental purposes of teacher evaluation as improving performance and documenting accountability," (Tucker & Stronge, 2005, p. 6). New evaluation tools, such as TKES, will be used to hold teachers, administrators, schools, and districts accountable. The Measures of Effective Teaching (MET) Project, funded by the Bill and Melinda Gates Foundation, has analyzed thousands of lesson videotapes and studied the shortcomings of current practices (Marshall, 2012). Researchers have found that one classroom observation a year doesn't give an accurate picture of a teacher's work. Observations and feedback are major components of TKES. According to the TKES Handbook (2013), observations are intended to provide information on a wide variety of contributions made by teachers in the classroom or to the school community as a whole. Observations can be conducted in a variety of settings and take on a variety of forms, including quick, drop-by classroom visits, to more formative, pre-planned observational reviews, using validated instruments for documenting observations. Furthermore, observations may be announced or unannounced. TKES encourages evaluators to conduct observations by observing instruction and non-instructional routines at various times throughout the evaluation cycle.

Appropriately trained and credentialed evaluators conduct multiple classroom observations per teacher. They include: Four classroom walkthrough observations and two formative classroom observations. Observations can be announced or unannounced (based on LEA decision). Walkthroughs observations are at least ten minutes in duration based on a limited number of Performance Standards and the formative observations are at least thirty minutes in duration and are based on all ten Performance Standards. Additional observations may be conducted at the building administrator's discretion.

Documentation of teacher practice and process is the second data source for TAPS. Evaluators may request documentation from a teacher when a standard is not observed during a formative observation(s), walkthrough(s) or when the consistency of a teacher's practice cannot be established. The teacher is responsible for submitting requested documentation in a timely manner. In order for it to be considered, requested documentation must be submitted prior to the completion of the Formative Assessment and/or Summative Performance Evaluation. Additionally, teachers may submit documentation at any time during the evaluation process for consideration by the evaluator.

Formative and Summative Assessment Process

Each formative observation will be directly tied to a formative assessment. Once the formative observation has occurred, evaluators should consider all observations (including walkthroughs) and documentation to rate the ten Performance Standards using the performance appraisal rubrics. The Summative Performance Evaluation is based on formative assessments, additional observations and documentation as well as student perception data generated by the Surveys of Instructional Practice.

Student Surveys of Instructional Practice

The second component of TKES is Student Surveys of Instructional Practice (student perception surveys). Student surveys are administered annually to gather perception data regarding teacher practice. Student survey results inform the rating of Performance Standards 3, 4, 7, and 8 on the Formative Assessment. Student survey results also inform the rating of Performance Standards 3, 4, 7, and 8 on the Summative Performance Evaluation.

The survey is administered in secure conditions outside the presence of the classroom teacher. Students complete the surveys under the supervision of a professional with a valid teaching certificate, service certificate, or leadership certificate. Consistent with state testing

requirements, paraprofessionals may not administer the student surveys unless they also have a valid teaching certificate. Survey responses will be anonymous with no option to enter comments. Evaluators and teachers will be provided with a summary for each standard with a mean score through the GaDOE. Grade 3-5 responses range from 0-2 and grade 6-12 responses range from 0-3, the higher the rating the more positive the student perception.

Student Growth Percentiles

The final component of TKES is the Student Growth (SGP and SLO). The SGP is gathered for teachers of tested subjects (4th - 8th grade Criterion Referenced Competency Test and high school End Of Course Test); this component is comprised of a student growth percentile/value-added measure which shall be calculated annually for student growth based on state assessment data. Student Learning Objective Measures (SLO) is for teachers of non-tested subjects whose students are not assessed using state assessments, this component is comprised of LEA developed and DOE approved Student Learning Objectives. Student growth is calculated annually based on SLO assessment data. For teachers of both tested and non-tested subjects, the Student Growth component is a combined measure that informs the Summative Assessment.

Evaluator Training

An evaluation system and the components that make it up are successful only if the evaluators implementing it at the local level are trained effectively. A 1983 report published by the United States Department of Education (USDOE) stated that the Nation was falling behind globally, and that our country was not as competitive as it once was (NCEE, 1983). The decline in education has had a negative effect on society both socially and economically. For about thirty years, studies have documented the decline in education and researchers have identified specific areas in need of improvement with teacher evaluation being one of them. One method to increase teacher performance is through the effective use of the evaluation process.

Evaluators can only improve the performance of classroom teachers if they are trained to give valid, reliable, meaningful feedback. Over the past decades the federal and state governments have made various attempts at educational reform by taking an advisory role, but it has had little impact. As our schools continue to fail to meet the expectations set forth by top ranked countries and guidelines to become college and career ready, it is imperative that something new be done. Reforming the evaluation system through enhancing the skills administrators need to be effective evaluators of teachers will be crucial.

In the era of changing evaluation systems, Marshall (2005) noted that there is a lack of evaluator (principal) training on how to use them properly. Principals need to develop good skills at knowing the system thoroughly, how to observe using the system, how to effectively write up an evaluation within the system, and then how to have what can sometimes be difficult conversations with the teacher. Marshall stated that principals are often too busy to devote the time needed to evaluate teachers and learn the systems. Being an expert judge of teaching is a learned skill. Hazi (2012) noted that judgment is used in evaluation and can be used to understand the contents of good teaching, but evaluators must be trained to be "expert judges".

Over the past twenty-seven years, Jon Saphier has looked at how administrators have evaluated teachers. He found that "The incoming skills of administrators tend to be quite low. Administrators have little or no preparation for analyzing instruction when they are hired, and almost never had to show they could do so to get their job," (2008, p. 2). In addition, with the budget cuts many districts are facing in current times, funds to train current administrators on how to use the new evaluation tools and clearly understand the components of them is minimal. There are many issues regarding the way principals are trained to become effective evaluators.

Witters-Churchill (1991) discussed serious concerns about current preparation programs of school principals and identified key skills that were documented in a study based on the NAASP University Consortium's "Red Book". Recommendations were made for: Improvement of graduate instruction in job-related skills, extension of opportunities for field experience, provision of practice-oriented university staff, and improvement of generic skills instruction. In addition to higher education, principals can participate in staff development to increase their knowledge in the field of evaluation. McLaughlin and Oberman (1996) found that there are challenges in developing staff development that are specific to the principal's needs as an evaluator. They suggested involving post-secondary partnerships for higher education and the advancement of degrees as one way to help principals.

There have been a number of studies that have looked at the teacher evaluation process and many of these studies have focused on the process and have identified the most common mistakes that evaluators make (Ostovar Namaghi, 2010). Few studies have inquired about the professional development evaluators need in order to become effective evaluators of teachers. In order for teacher evaluation to be meaningful and facilitate professional growth the principal's role needs to change from judge to mentor or coach (Danielson & McGreal, 2000). Although RTTT requires that every administrator, who will be responsible for conducting teacher evaluations, become certified by attending required trainings, there is little research to determine if the requirements are enough. RTTT allows each board of education to decide what the certification process will consist of for the administrators in their district. Although administrators attend these required trainings, in some cases there is no summative evaluation to determine if they have mastered the goals and objective of the training, which were designed to make principals effective evaluators.

There are several factors that link communication to the evaluation of teachers and evaluators must be trained to communicate effectively. Principals need to be able to communicate the process and how it will work in his or her school. Also, there needs to be significant and sincere communication before, during, and after an observation or walkthrough (Marzano, Waters & McNulty, 2005). In order to make evaluations a useful tool for professional growth there must be a sense of trust, which can only be established through the use of honest, unbiased feedback that has been strategically planned. "It is part of principal competence to know who needs reflective talk, who needs encouragement, who needs confrontation, and who works well (or needs to be) left along" (Peterson & Peterson, 2006, p. 70).

Elmore (2000) believed that leaders are not adequately trained to meet the needs of twenty-first Century teacher evaluation. Previous studies show that principals rate their skills in executing the teacher evaluation process very high (Painter, 2000). This finding is in opposition to research, which has found that teachers do not feel that evaluations are meaningful or lead to their professional growth. Danielson and McGreal (2000) report that principals are at a disadvantage as evaluators if they do not have command of the content being taught and not having a background in the subject area could make it difficult for principals to identify inaccuracies in the information being presented, especially in advanced courses.

It is perceived that many administrative preparation programs do not adequately prepare aspiring principals to be effective evaluators. "Historically in the United State, leaders at the school level (principals) and school district level (superintendents) began their careers primarily as classroom teachers. This common professional background has generated a population of educational leaders who share similar motivation, knowledge, and challenges," (Dyer & Renn, 2010, p. 3). Lynch (2012) noted that historically principals served as disciplinarians and the

teachers' boss but under the current federal legislation, principals now must accept the responsibility to manage personnel, funds, and strategic planning while being the school's instructional leader. The skills mentioned by Lynch must be covered in principal preparation programs to help them become better evaluators.

When administrators start their job, they rarely have any experience or training in conducting teacher evaluation. In fact, most beginning administrators do not have experience evaluating teachers until their first assignment. Lashway (2003) stated that "evaluator training is usually at the beginning of preparation programs and that is usually followed by informal, self-guided, and sporadic professional development" (p. 4). Mitang (2012), in a report published by the Wallace Foundation, stated that University-based program are flawed because they consist of criteria that fail to take into account the needs of districts and diverse student bodies; they have weak connections between theory and practice; the faculty have little or no experience as school leaders; they lack opportunities to experience real leadership; and internships are poorly designed and insufficiently connected to the rest of the curriculum.

Schmoker (2006) stated that most people will agree that principals are the most important leaders in our school system and effective leadership in schools is still dismayingly, exceedingly rare. The author noted that the key to improving school leadership begins with "demystifying" it. To do this, leadership must be redefined around professional learning communities: team-based, cooperative arrangements between instructors and administrators. Professional learning communities is a commitment to having all teachers meet regularly with their colleagues for two primary purposes: 1) to determine, in common, the essential standards they will teach in each course on a common schedule, and 2) to prepare lessons and units together, assess their impact on student learning, and refine their instruction of the basis of these assessment results.

Finally, evaluator training must recognize that adults learn differently than students. Adults learn from everything the mind perceives and brains build and strengthen neural pathways no matter where we are, no matter what the subject or the context. Finding better ways to learn will propel organizations (schools) forward and strong minds will fuel those organizations (schools). Capitalizing on an adult's natural style of learning and building systems to satisfy needs will enhance the learning environment for students. Drago-Severson (2009) highlighted that education reform policies need to encompass how educators and evaluators (adults) learn and use their knowledge as a catalyst to promote effective development. Malcolm Knowles' (1998) theory of andragogy is an attempt to develop a theory specifically for adult learning and could be useful for evaluator training. Knowles emphasized that adults are self-directed and expect to take responsibility for decisions. Therefore, adult learning programs must accommodate this fundamental aspect with the following assumptions about the design of learning: (1) Adults need to know why they need to learn something (2) Adults need to learn experientially, (3) Adults approach learning as problem-solving, and (4) Adults learn best when the topic is of immediate value. Instruction for evaluators needs to focus more on the process and less on the content being taught.

TKES Evaluator Training

The research over the past thirty years is consistent in reaching the conclusion that effective evaluations should include student data, multiple measures, and evaluator training. As TKES is implemented, administrators need to be trained effectively. The TKES Handbook terms training as "credentialing". According to the Handbook (2013) "Credentialing is a process of establishing the qualifications and proficiency of evaluators to utilize TKES. The credentialing assessment is a minimal competency assessment that measures participants' understanding of the information and practice provided during training. It includes both recall of specific TKES

information and practice of all the TKES processes. It is a bridge to ongoing learning as evaluators work with TKES. All evaluators MUST pass the credentialing assessment prior to using the Teacher Keys Effectiveness System (TKES Handbook, p. 16). There is an expectation that evaluators will continue to familiarize themselves with the TKES process as they work within their districts and the credentialing process will provide calibration that further increases the alignment of evaluation ratings. As districts are left to determine how to train their evaluators, there is a problem with calibration because, currently, the only trainers providing full TKES training and leading the credentialing for administrators are members of the Teacher and Leader Effectiveness (TLE) Division of the GaDOE. Opportunities for becoming a state certified trainer is available but the consistency for training opportunities among districts in Georgia is unreliable.

Time spent on the evaluation process is only beneficial when used properly. The amount of time dedicated to feedback with TKES can only be used to increase teacher performance and, in turn, student achievement if the feedback is given appropriately and differentiated based how adults learn best. The two main factors that have motivated the nation's move to re-design teacher evaluation is that, historically, teachers have not received meaningful feedback and traditional evaluation systems have not differentiated among high and low-performing teachers (Sartain, et.al., 2011). Current evaluation systems and training for evaluators does not provide teachers with the information they need to make timely and effective improvements in their instructional practice (McLaughlin, 1990). It is of particular concern that most evaluation systems do not differentiate between strong and weak instruction or provide sufficient feedback.

Administrators at local schools will be trained by their districts to use TKES. As the instructional leaders in the school, the administrators serve as the model for appropriate

evaluation practices, coordinates all evaluation activities within the school, and has ultimate responsibility for all evaluation activities within the school. Following the TKES training and credentialing, evaluators are encouraged to, but not mandated to, review classroom observation videos, observe lessons in classrooms and discuss ratings along with judgment of practice based on the TKES standards rubric.

"Most evaluations are not used to target the needs of individual teachers and help them select professional development to address those areas in which they need additional knowledge of skills" (National Board Resource Center, 2010, p. v). Although the policies included in the legislation for RTTT mandate specific components for teacher evaluation, it is difficult for administrators and teacher to make the transition from being an action or routine that they need to complete at the end of the school year (Frase, 1992), to an ongoing process for continued improvement. In a report conducted by the Center for American Progress, thirty principals were interviewed regarding hiring, assigning, and evaluation practices used in the United States and principals agreed that the evaluation process is used to improve teachers' practices, but in reality, they do not believe that this purpose is actually carried out (Donaldson, 2011).

Due to the lack of training on giving specific feedback throughout a school year, evaluations tend to be highly inflated and, "used as an opportunity for cheerleading and motivating, rather than providing critical feedback" (Varlas, 2009, p. 6). The Joyce Foundation (2010) studied over 100, 000 teachers from Illinois and found "evaluations typically contain little detail to show who are the outstanding teachers' or the poor performers and they do not relate school goals or identify strengths and weaknesses that could be used to start turning things around" (p. 3). According to the Center for American Progress "current teacher evaluation systems do not give proper feedback, so effective teachers are not identified and rewarded, and

ineffective teachers are not given the tools to improve. Students and teachers lose out as a result" (2010, p. 1).

It is imperative that individual school administrators understand and articulate to the staff what components of the evaluation instrument he or she is looking for in an observation or walkthrough. It is also important that the perceptions of principals and teachers in regard to the evaluation system and the impact it has on teaching and learning are assessed. The level of support teachers and principals in Georgia have for the implementation of TKES as their evaluation system is unclear at this time and a needs assessment will help uncover the greatest areas of concern for both teachers and principals. Principals need to be familiar with the policies and an assessment of the evaluation tool itself is needed so that Georgia can ensure its new policy is "in pace with new ideas about what, when, and how teachers learn…and focus on developing schools' and teachers' capacities to be responsible for student learning" (Darling-Hammond & McLaughlin, 2011).

Needs Assessment

A needs assessment is a systematic process for determining and addressing needs or gaps between current conditions and desired conditions. The discrepancy between the current conditions and the wanted condition must be measured to appropriately identify the need. Clark (1993) presented a literature review that recommended states and school districts evaluate the process at hand before evaluating teachers directly. A needs assessment is part of a planning process and, in the case of TKES implementation, a planning process on how to effectively support the positive attributes of the evaluation system and identifying the areas of weakness. By clearly identifying the problem and evaluating the process, the resources being used to train and prepare evaluators can be directed toward developing and implementing a feasible solution (Fulgham & Shaughnessy, 2008).

Roger Kaufman first developed a model for determining needs defined as a gap in results (Fulgham & Shaughnessy, 2008). Kaufman explained that to conduct a quality needs assessment first determine the current results, articulate the desired results and the distance between results is the actual need. In relation to TKES and what evaluators need, a training needs assessment is a systematic inquiry of training needs within an organization for the purposes of identifying priorities and making decisions, and allocating finite resources in a manner consistent with identified program goals and objectives (Triner, Greenberry & Watkins, 1996). Miller (2000) stated that the training needs assessment is a critical activity for the training and development function. There are three levels of needs assessment: Organizational analysis, task analysis, and individual analysis. The focus of this study will be on task analysis: Provides data about a job or group of jobs and the knowledge, skills, attitudes, and abilities needed to achieve optimum performance (Miller, 2000).

Strauss (2012) reported that professors from colleges and universities in Georgia wrote a letter to the governor, Nathan Deal, to protest the new evaluation system. Badertscher (2014) claimed that the state's new teacher evaluation system needs some work and it is skewed because the data from the pilot districts shows only a small numbers of Georgia teachers are ineffective while test scores do not reflect that. McManus (2011) questioned the intent of the Georgia students' role of evaluator with the student surveys and Shearer (2013) strongly encouraged Georgia to seek a waiver to wait on implementation of the controversial new evaluation system. As Georgia implements TKES, a Training Needs Assessment will address what some of the professionals in the field of education have doubts and reservations about by pinpointing where training needs to occur. Success in Georgia's new system will come only if careful planning and reflection occurs. As Millman and Darling-Hammond reported, there are specific factors which

are critical to implementing and sustaining a teacher evaluation program at the district level including technical and managerial concerns, accountability, bureaucratic control, professional autonomy, individual needs, and institutional goals.

Summary

In this chapter the literature was reviewed in regards to teacher evaluation systems and how evaluator training impacts the effectiveness of implementation and use. The review of the literature indicates that having highly effective teachers in the classroom is the primary way to increase student achievement. Recognizing which teachers are effective and giving a means for professional development to those who are not has been identified as the primary use of an evaluation system. Transforming the teacher evaluation process from a simple documentation of competency to a tool for teachers to become proficient and hopefully distinguished in their field requires evaluators to possess new skills and a better sense of understanding on how to use the data derived from evaluation.

The study presented evolved from the literature reviewed by building a foundation of where teacher evaluation comes from, what changes have contributed to the components found within the different models of evaluation, and the future of how Georgia's TKES will be used effectively by trained evaluators. The literature and theoretical frameworks derived from it were used to provide insights into what questions to ask during the interview process, what codes to use, and how to think about organizing and analyzing the data collected.

As education continues to change, the public is more aware of what public education promotes and the way the teachers are rated. The State of Georgia releases Report Cards that communicate school performance to the public and gives them the opportunity to watch and either support or retract the teachers in their local school based on the assumption that the performance is directly correlated to teacher effectiveness.

Brimley and Garfield (2008) reported school systems are slow to respond to public pressure and adapt, but the federal government has taken legislative action to bring about the change for necessary sustainability of changes that will lead to greater student achievement. NCLB was the first mandate that focused on the recruiting and retaining of highly qualified teachers in order to increase student achievement and it was through this legislation that there was a focus on standards for both students and educators. RTTT is the latest reform legislation aimed at improving the quality of our schools and increasing student achievement. Teacher evaluation is one of the major components found in the RTTT legislation making it the first mandate to design a rigorous evaluation system. It is due to RTTT the Georgia has implemented TKES in hopes that their evaluation system is transformed into a valuable tool to strengthen instructional delivery and the effectiveness of every classroom teacher.

Included in the review of literature was a historical overview of the function and purpose of teacher evaluation. It was not until recently that teacher evaluation became a collaborative effort between evaluator and the evaluated. The systems in place now, specifically TKES, were developed to reflect a teacher's competence, establish goals, and give incentives to grow professionally. The teacher evaluation process, as mandated through RTTT, encourages communication and self-reflection about a teacher's performance when aligned with a clear set of criteria. It is through the change in theory and concept behind the purpose of evaluation and the structure of it that will lead to ultimate growth, but only if both the evaluator and the evaluated are trained to implement such a transformation and make the idealistic theory a realistic concept.

The study presented focuses on the importance of evaluator training, which was found in limited resources throughout the literature reviewed. Although evaluators can learn to avoid the

most common mistakes made by evaluators and place more trust in the new effectiveness systems, the research on how often, how much, and how effective reflection, training, and communication has not yet been determined. This study will add to the research on evaluator preparedness and the importance of communication and collaboration between the evaluator and the evaluated. The research showed that the professional development and support is not consistent among different evaluation systems and the mandates on training vary from state to state and, specifically in regards to TKES, from district to district. It can no longer be sporadic, but rather on-going and intense enough to ensure that evaluators develop the skills necessary.

The study examined teacher evaluation systems and the research adds to the knowledge base of teacher evaluation by providing important insights in the area of development, implementation, and supports needed by current and future evaluators to conduct valid, reliable, and effective teacher evaluations. This study was designed with the intent of increasing the knowledge about what is currently in place in Georgia with TKES, but the hope is that the outcome of the research leads to a deeper understanding of what strategies and supports would be beneficial for evaluators to feel confident and productive in their skills to evaluate teachers while using any evaluation system.

CHAPTER 3: METHODOLOGY

The purpose of this qualitative study was to determine the perceived level of self-preparedness of building level principals to conduct effective teacher evaluations and to identify the future professional development needs while using the Georgia Teacher Keys Effectiveness System. The main concept to be discussed using qualitative methods was the context of teacher evaluation training needs amongst a group of building principals in one school district in Georgia. By clearly identifying the problem and evaluating the process, resources used to train and prepare evaluators were suggested through the development of a needs assessment. The program evaluation needs assessment was designed in an attempt to help develop and implement a feasible solution to any reported issues with the training provided to implement the new evaluation system (Fulgham & Shaughnessy, 2008). Multiple methods' of interviews and document analysis were used in this study and the data collection techniques, the instrumentation tools, and data analysis are described in this chapter.

Chapter three is divided into six sections that describe the components of the research design and methodology to be applied when exploring the readiness of evaluators during the implementation of a new teacher evaluation instrument in the state of Georgia, specifically, Cobb County School District. The first section is the specific research design and will provide an overview of the qualitative study and the reason this type of method was chosen. The second section provides a perspective of the unit of analysis and gives explanation of the framework. The third section is the method of data collection, including a summary of the research questions. The fourth section is a discussion of the sample used and sampling strategy. The fifth section is

a discussion of the data analysis and limitations and the final section is a summary of the information provided in chapter three.

Research Design

Qualitative research was used in this study. Qualitative research, according to Merriam (1998), lies with the idea that meaning is socially constructed by individuals in interaction with their world. There are not fixed explanations, but rather interpretations of reality that are in flux and that change over time. Qualitative researchers are interested in understanding what those interpretations are at a particular point in time and in a particular context. The nature of this study allowed for simultaneous processes of analyzing while data was collected (Creswell, 2008) and was an effort to understand the level of preparedness principals have toward conducting effective evaluations. There are several key characteristics in qualitative research: Researchers strive to understand the meaning people have constructed about their world and their experiences, the researcher is the primary instrument for data collection and data analysis, the process is inductive in that researchers gather data to build concepts, hypotheses or theories rather than deductively deriving postulates or hypotheses to be tested, and the product of a qualitative inquiry is richly descriptive. Each of these characteristics will be discussed in terms of how they apply to the current study (Merriam, 1998).

Qualitative researchers strive to understand the meaning people have constructed about their world and their experiences. Qualitative research is an effort to understand the situations in their uniqueness as part of a particular context and the interactions there (Patton, 1985). The understanding revealed is an end in itself and attempts to understand the setting and what it means to the participants. In this research study, school principals who conducted evaluations using a new instrument were examined based on their experiences within this unique setting and the understanding they take away from using the new evaluation instrument.

The researcher as being the primary instrument for data collection and data analysis can be described in another way by the participant having a perspective that entails identifying the perspective of the participants from the world in which they are living (Hatch, 2002). In the context of this study, this particular characteristic was apparent in the fact that the principals created a lens in which the researcher viewed the interpretation and understanding of their preparedness for using the new evaluation instrument. The view of the researcher is not as important to this process as that of the participant and, since the researcher has significant involvement in the public schools as a principal and evaluator, it was imperative that the focus remained solely on the participants.

The process was inductive in that the researcher gathered data to build concepts, hypotheses and theories, rather than deductively deriving postulates or hypotheses to be tested. The researcher gathered data through a variety of means (interviews, transcriptions of notes, observations, and review of historical documents) and it was then organized and "processed using the human intelligence of the researcher," (Hatch, 2002). In attempting to understand the meaning a phenomenon has for those involved, qualitative researchers build toward a theory from observations and intuitive understandings gleaned from being in the field. Typically, according to Merriam (1998), findings inductively derived from the data in a qualitative study are in the form of themes, categories, typologies, concepts, tentative hypothesis, and even substantive theory. In the context of this study, the researcher conducted semi-structured interviews with a sample consisting of primary and secondary school principals from a large suburban school district in the state of Georgia. These data was then coded to determine if there is a need for further preparedness in terms of using a new evaluative instrument with their

teachers. Themes will then be analyzed by the researcher which will provide an insight into the understandings of these particular principals.

Finally, the characteristic that the product of a qualitative inquiry is richly descriptive comes from an understanding that it is an extended firsthand experience. Words and pictures rather than numbers are used to convey what the researcher has learned about a phenomenon. The researcher is expected to spend time within the setting and, thus, derive descriptions of the context, the participants involved, and the activities of interest (Merriam, 1998). The data, in the form of quotes, tapes, electronic communication, or a combination thereof are always included in support of the findings of the study, which contribute to the descriptive nature of the study. In this study, as an evaluator herself, the researcher spent a great deal of time within the context of the environment in which the evaluative measures are implemented by the school district. Hatch (2002) states that research indicates the need for extensive time in the field to be able to "capture what they claim". Corbin and Strauss (2008) explained that it is important to understand what the people in a study are doing in regards to the scenarios in which are being recorded. In regard to this study, the researcher devoted specific time so that there was engagement between the principals' environments and her understanding.

As cited in Fitzpatrick, Sanders & Worthen (2011), the Joint Committee of Standards for Educational Evaluation (1994), defined program as simply, "activities that are provided on a continuing basis" (p. 3). In a newer edition, it was noted that a program is much more than a set of activities. They defined a program as having the following characteristics: A set of planned systematic activities, use managed resources, achieve specified goals, relate to specific needs, specific, identified, participating human individuals or groups, specific contexts, results in

documentable outputs, outcomes, and impacts, follow systems of belief (implementation theories on how the program works), and investigable costs and benefits.

Program evaluation is different than program research because its main purpose is not to add knowledge to the field, but rather provide useful information to those who hold a stake in whatever is being evaluated, which often helps them to make a judgment or decision (Fitzpatrick, et.al. 2011). In this study, the information provided by the researcher will enable numerous groups of stakeholders to make decisions about the current training principals receive before providing feedback to teachers through the use of a new teacher evaluation instrument.

Assessment is generally defined as the preamble of the current state and evaluation is the opposite – it is the summative activity that determines what change has occurred. Time and efforts can be saved by taking advantage of the parallels between needs assessments and evaluation. Assessment without evaluation is a diagnosis without follow-up to see if a treatment works or not. Without evaluation, an intervention cannot be labeled a success or failure. Evaluation without assessment is a final exam given without a baseline comparison. Evaluation prompts continuous improvement and analyzing the data from an evaluation provides the direction for process or activity change. Needs assessment can be represented as a formula of what "*Should Be*" minus what "*Is*" equals the need.

Using qualitative design, the research will specifically focus on a needs assessment as a type of program evaluation. Evaluation has been used to test the effectiveness of programs and consider the effectiveness of new and existing programs, procedures and/or interventions at producing some form of outcome or change. The findings from evaluations focus on the strengths and weaknesses of various aspects of innovations as well of their overall outcome. This information is, in turn, used to consider how such interventions might be modified,

enhanced or even eliminated in the effort to provide a better service, fulfill a particular need or meet a specific challenge. In education, evaluation has served a somewhat similar purpose.

Specifically, a needs assessment within the context of this research will be a systematic process to acquire an accurate, thorough picture of the preparedness of principals in administering a new evaluation instrument and extracting the needs for future training. The needs assessment was derived from information that was collected and examined about the school-wide evaluation of teachers. The data can be used to plan and allocate funds and resources toward future professional development, which will lead to an ultimate goal of having meaningful evaluations that lead to improvement in instruction and, ultimately, student success.

Needs assessment should be seen as an integrated set of procedures rather than groups of disproportionate elements. Needs assessments have the following characteristics: Needs are identified as gaps between current results and desired results (Hobbs, 1990); the *Should Be* state (desired results) is defined in terms of competency-based job requirements (Wright & Geroy, 1992); employees are involved in their own *Is* state assessment (Freire, 1970); the needs assessment process yields proactive rather than reactive results for individuals or organizations (Wright & Geroy, 1992); a statement of need is free from any indication of what solutions and resources will be used to meet the need (Kaufman, 1994); information on the desired *(Should Be)* state and on the assessment of the employee's current *(Is)* state is validated using multiple data points gathers through a variety of data collection methods (Wright & Geroy, 1992); the *Is* state is analyzed in their organizational context (Mager, 1972); guidelines are provided for prioritization of gaps (needs) based on cost, importance, consequences, and/or value-add for both internal and external clients (Hobb, 1990).

Through the interviews given, the Principals of the school district defined the processes which were identified qualitatively through open coding. The researcher immersed herself in the data by reviewing the transcriptions numerous times and continuously linking responses to one another and the research questions. The key terms that related to the literature reviewed were highlighted using three different colors and then letters were assigned to sub-themes that emerged from within the consensus themes. The processes for improving employee performance (Human Performance Technology, HPT or Human Performance Improvement, HPI) lead to the identification of appropriate solutions to resolve the causes of poor current human performances (Sission & Swanson, 1990). In this research, the performance of the Principals is dependent on their preparedness to implement the evaluation instrument. Proponents of HPT believe that people are one part of the "performance engine" which has a number of components that influence performance (Rummler & Brache, 1995), such as staff development and a Principal's education. HPT is a systematic approach to improving productivity and competency by achieving the desired organizational and individual results. In this research, the desired result is improved teacher performance through the feedback given in an evaluation. HPT emphasizes the systematic analysis of the factors that influence behavior and performance to measurably improve human performance and job satisfaction (Wikoff, 1994). Everything in an organization's internal and external systems is connected and must be understood to improve organization and individual performance (Rummler & Brache, 1995). The qualitative approach, through interviewing, allowed the researcher to examine the different factors that impact Principal preparedness. HPT can be applied to individuals, small groups, and large corporations to influence human behavior (Lane, 2000).

Unit of Analysis

The unit of analysis in this research study was a group of principals in public schools (K-12) in a suburban school in a large district within the state of Georgia. The researcher chose nine principals with various experiences, education, backgrounds, and training resources. These building leaders were chosen on the basis of their length in time being evaluators (over two years), level in which they lead (three principals from each level: elementary, middle, high), and training on using the new teacher evaluation instrument (standard training without additional development). Each principal has started using the new evaluation instrument through their own professional duties and responsibilities. These principals provided the unit of analysis for this research.

Data Collection

The qualitative research model, which has emerged from research efforts in the social sciences, is particularly appropriate for this study because it is effective in understanding the total environment in which training takes place (Creswell, 1998). The four research questions acted as a guide for the researcher for data collection and analysis. The researcher collected data from multiple sources to determine evaluators' perceptions on their preparedness to conduct effective evaluations using TKES. The majority of the data came from individual interviews and then a focus group interview. A small portion of the data was taken from school district and state documents describing the teacher evaluation process and instrument used.

The interview process offers an opportunity to gain insight into the participants' perspective (Creswell, 2007). Individual interviews with the nine principals initially focused on basic demographic information from the participant to include: education background, years of experience as an evaluator, and training to implement TKES. The years of experience could be a key component when determining preparedness to conduct effective evaluations. The

background data provided in this section will serve a dual purpose in that the participant will be provided an opportunity to answer basic demographic questions which are very familiar to them, as well as allow them to become more comfortable and increase their familiarity with the researcher (Hatch, 2002). The steps for interviewing that will be followed are those cited by Creswell (p. 163-164):

*Decide on the Focused Research Questions***.** The research questions were open-ended, general, and focused on understanding the central phenomenon of the study: What is the level of preparedness principals report to have while effectively implementing a new evaluation instrument. Open-ended questions are recommended by qualitative researchers because these types of questions lead the participant to discuss their experiences as they understand them (Hatch, 2002). The questions were formatted and asked of each principal who participated in the study. The researcher listened attentively and asked appropriate follow-up questions. The questions were intended to provide the most accurate picture of the lived experiences of the participants. Hatch (2002) recognized six generalizations to be used when writing good interview questions. These will be used in this study: Open-ended, formatted so language and vocabulary used provides familiarity, clarity (no hidden agenda), neutral in nature (no influential questions), respectful, and relevant (directly related to the objectives of the research).

The interviews generated data in order to best describe the perceptions of evaluator readiness to conduct effective evaluations as the participants answered questions from the interview and the researcher described the responses given. To ensure reliability, carefully crafted questions were created. The questions were open-ended with the goal of getting closer to the truth about how evaluators feel they are prepared to give effective feedback. Open-ended questions allow for a greater variety of responses from participants by are difficult to analyze

statistically because the data must be coded or reduced in some manner. Closed-ended questions are easy to analyze statistically, but they seriously limit the responses that participants can give. To accomplish meaningful dialogue on evaluator readiness to conduct effective evaluations, nine principals within the Cobb County School District in the state of Georgia were selected for interviews. These participants were intentionally selected based on experience and the level of education in which they work (elementary, middle, or high).

Identify Interviewees. Interviewees were chosen based on who could best answer the questions based on maximum variation sampling. The interviewees were those that could best answer the research questions. In order to understand the differences between elementary, middle, and high school principal responses, three from each level were interviewed and their differences were analyzed when coding the data. After IRB approval, a letter requesting participation was sent to eligible interviewees (*Figure 3.1)*. Respondents were assured that neither their personal identity nor the identity of their school would be released during the dissertation. A return address, email, and phone number was provided for the selected participants to respond back with either a yes or no, or to have any questions answered. Selected participants who did not respond received a follow-up email from the researcher. If a selected participant declined, the researcher determined and selected a new participant with similar characteristics from the one who declined. Once a group of participants were selected, interviews with the building principals were scheduled.

Design and Use an Interview Protocol. A protocol was created and used that will allowed for space to write responses. The interviews lasted approximately one hour while the researcher recorded responses. Adequate recording procedures were used so that the researcher could record the interviews using instruments that were effective and comfortable for both the

interviewer and interviewee. The protocol to be used is outlined in Figure 3.2. During interviews, the researcher kept analytical notes to address personal understandings of comments made.

Determine the Place for Conducting the Interview. In order to accommodate a setting that was comfortable for the interview, the researcher offered a location that was comfortable and easily accessible. The natural settings were within the Cobb County public schools in Georgia. The researcher attempted to schedule these interviews at convenient times for the principals, while at the same time utilizing a time frame which allowed them to be fully engaged in the interview process. An effort was made to schedule it during natural school breaks or after school hours. By scheduling in this way, the researcher believed the data would be more accurate than if the school principal were faced with focusing while interacting with daily responsibilities. Follow up interviews were scheduled as needed to clarify understanding based on the responses during the initial interviews.

After the individual interviews were completed, the researcher conducted a focus group interview as another qualitative method (Creswell, 2013). Focus group interviews probe deeper than surveys. They explore topics individually and can lead to a thick, rich description of responses. According to Marshall and Rossman (2011) the method of interviewing participants in focus groups comes largely from marketing research but has been widely adapted to include social science and applied research. The researcher created a supportive environment and asked focused questions to encourage discussion and the expression of differing opinions and points of view. The focus group format triggered conversation and thoughts within the group. The strengths of focus-group interviews are that this method can be socially oriented, studying participants in an atmosphere more natural than artificial experimental circumstances.

Participants may generate new ideas between each other that will lead to further questions asked by the researcher. The results of a focus-group interview have high "face validity (Marshall & Rossman, 2011) because the method is readily understood, the findings appear believable, the cost of focus-group interviews is relatively low, they provide quick results, and they can increase the sample size of qualitative studies by permitting more people to be interviewed at one time. Again, analytical notes were kept by the researcher to record personal understanding of comments made.

After the interviews were conducted, the researcher analyzed the field notes obtained from each participant. During the individual interviews and the focus group interview, the researcher took notes about interactions and, after obtaining permission, recorded the interviews to be transcribed and analyzed inductively later. The notes were compiled and served as a way to capture the essence of the interview and provide the researcher with an opportunity to view the data from an introspective viewpoint. Transcripts were created based on the recordings of the interviews so that the data was reviewed at a later date. This is a critical point of the process and was exhausting; as Patton (1990) noted, transcriptions typically take about four hours per one hour of interview time. The researcher compared field notes with the transcriptions in order to analyze the data into themes for this study. Data analysis will be discussed in greater detail later in this chapter.

The primary ethical issues that may arise in conducting focus-group interviews center on the dynamics of power and influence that may play out in any group. The researcher was sensitive to the dynamics of the group and facilitated the process. There are additional challenges regarding the interview process that Creswell (2013) cited. Unexpected behavior by the participant, the researcher's ability to create questions, formulate clear instructions, deal with

sensitive issues, and conduct transcriptions, are all issues which are faced by the researcher. In addition, interviewing is also time-consuming and exhausting. "Undoubtedly, conducting interviews is taxing, especially for inexperienced researchers engaged in studies that require extensive interviewing," (p. 173). As data is being collected, the research must search through a plethora of data for analysis so that a determination may be made regarding the specific questions surrounding the research. The research questions for this study were centered on the lived experiences of the principals in conducting teacher evaluations and giving feedback. These challenges were met and addressed by the researcher in order for clear and concise data collection to take place and create a valid study.

There were ethical issues during data collection in the field and in analysis and dissemination of qualitative reports. The data collection phase of research present ethical issues into informed consent procedures, deception or covert activities, confidentiality toward participants, benefits of research to participants over risks, and participant requests that go beyond the social norms. The researcher protected the anonymity of the participants, explained the purpose of the study, and did not engage in deception about the nature of the study. Finally, an ethical issue that presented itself in this study, due to the fact that the researcher is a principal who conducts teacher evaluations, was that the researcher shared personal experiences with participants in the interview setting. Maintaining professional interactions and staying in compliance to the interview questions asked, helped maintain ethical behavior.

Document Review

In addition to the individual interviews and the focus group interview, documents were collected and reviewed. Some of the documents were public documents, such as the actual TKES evaluation instrument, the TKES handbook, and others related to the study. According to Creswell (2003), there are distinct advantages and disadvantages in document analysis. Marshall

and Rossman (2011) stated that documents can provide background information that helps establish the rationale for selecting a particular site, program, or population. Researchers often supplement interviewing and observation with analyzing documents constructed specifically for the research at hand. The advantage of using documents is that it will not disrupt ongoing events and the researcher determines where the emphasis lies after the data have been gathered. A weakness is the span of inferential reasoning by the researcher (Marshall and Rossman, 2011).

In this study, the researcher analyzed documents presented to the principals during the focus group interview. The documents reviewed included the Teacher Keys Effectiveness System Handbook, the TKES instrument itself, and surveys sent by the Georgia Department of Education. The written explanation of these documents, their implementation, and their uses for training evaluators will be described by the researcher.

Sampling

The sample for this study was consistent with the research questions being posed. Creswell's (2013) explanation of a purposeful sampling approach was used to determine the specific principals to be interviewed during this study. The researcher selected nine principals in a large suburban school within the state of Georgia who currently complete teacher evaluations and give feedback as part of the process. The sample was interviewed individually and in a focus group because individuals may want to say something in private that they may not want to say in a group or reactions to questions might change when questioned in a group instead of individually. The sample was interviewed both ways to provide a robust amount of data in the research topic.

A major strength of purposeful sampling is that it allows the researcher to select participants that can provide insight into the question being studied (Creswell, 2007). In this study, only principals from one district were considered because the researcher was looking

specifically at principals who were in the first year of participation using TKES. In addition, it provides the researcher with a homogenous sample. The homogenous group can be reduced to sub-groups so that the researcher may examine the sample characteristics in greater detail (Hatch, 2002). This research consisted of sub-groups of elementary, middle, and high school principals. The sub groups allowed the researcher to identify commonalities which relate to the research questions. As purposeful qualitative sampling was used in this study, the selected principals were interviewed to best explain the phenomenon of validity of evaluator preparedness. The intent of selecting this purposeful sample was to develop a further understanding of evaluator preparedness to conduct effective evaluations. The interviews of the selected participants provided a voice to evaluators who might otherwise be silent on the subject.

The weaknesses of purposeful sampling include it being considered a "convenience sample" (Hatch, 2002) and the relationship between the researcher and the participants. The researcher remained aware of the issues which may have been caused by having a biased view of the participant or the participants having limitations in their feeling to discuss topics with the researcher without fear of information being used for unintended purposes.

The researcher selected building level principals that had more than two years of evaluation experience. The participants all had the same level of training and experience implementing this specific new teacher evaluation instrument. The participants generated from a pool of school principals in the school district in Georgia. They were analyzed based on the level in which the principal works (elementary, middle, or high school). These participants varied in age, race, gender, and educational background. There was also a variance among participants in the length of service in a particular school. This characteristic addressed the question of level of familiarity with teachers when evaluating them.

Data Analysis

This study will consist of individual interviews and a focus group interview with participants who range in experience as school principals. During the interview process, the researcher listened reflexively as the participants responded to the questions developed. These questions were developed to provide the participant the opportunity to explore the experiences which they have had as evaluators. As data was collected for this study, the researcher examined the responses provided in order to view the phenomenon being studied as the lived experiences of the participants (Hatch, 2002). Through the interview experience, the principals had an opportunity to freely comment on the questions being posed. These data provided the researcher with insight into the realities of the experiences of the participants.

Marshall and Rossman (2011) stated that, in qualitative studies, data collection and analysis typically go hand in hand to build a coherent interpretation. The initial concepts of this study (i.e. conducting effective evaluations) were understood more thoroughly and shifts in the researcher's understanding of the concept occurred as the data was analyzed and coded. The preliminary research questions and related literature were used as guidelines for data analysis and used for theory-generated codes. As the analysis of data occurred, incomplete data led to new collecting and analysis that served to strengthen the interpretation of the focus of the research.

Analytic procedures fall into seven phases that took place for this research: Organizing the data, immersion in the data, generating categories and themes, coding the data, offering interpretations through analytic memos, searching for alternative understandings, and writing the report or other format for presenting the study (Marshall & Rossman, 2011). Each phase entail data reduction (translating the data into manageable chunks) and interpretation (bringing meaning and insight to the words and acts of the participants in the study. The interpretive act

brings meaning to the data and displays that meaning to the reader through the written report. Patton (2002) noted, "Qualitative analysis transforms data into findings. No formula exists for that transformation. The final destination remains unique for each inquirer, known only when – and if – arrived at," (p. 432).

*Organizing the Data***.** After listening to the recordings the researcher transcribed the data into MS word documents. Next, the data was highlighted and the field notes were added. The researcher filed the transcriptions based on the dates and times when, the places where, and the persons with whom they were gathered.

Immersion in the Data. The researcher read and re-read the data and became intimate with the material. Descriptive data from the transcriptions will display the events of the phenomenon under study in pre-developed data-recording charts. Over-arching themes and sub-themes were generated from the coding.

Coding, or the formal representation of analytical thinking, was accomplished after the themes and sub-themes were discovered. Open coding was conducted first to relate to the conceptual categories. The researcher worked through the data and coded it using corresponding colors and letters related back to the set of research questions. After that, axial coding was done to group the first codes according to conceptual categories that reflected commonalities among codes. Responses were charted for commonality among level (elementary, middle, high).

As the data began to present itself in clusters or patterns around the coding, analytic memos helped the researcher work through inductive reasoning. Throughout the process, writing notes on questions that arise from clusters generated unusual insights that moved the analysis from the mundane to the creative (Marshall & Rossman, 2011). Patton (2002) described the processes of inductive analysis as "*discovering* patterns, themes, and categories in one's data".

The process allowed for a reflection on the understanding expressed by the participants. As categories and themes were developed the researcher offered integrative interpretations of what had been learned. The "telling of the story", according to Patton (2002) brings meaning and coherence to the themes, patterns and categories, developing linkages and a story line that makes sense and is engaging to read. The most useful data points were used to begin this process. Due to the fact that the researcher is also an evaluator in Georgia, using TKES, typologies were avoided if possible. However, themes created by the researcher that were grounded in data, but not necessarily in the participants responses, might be beneficial to the overall findings.

After the data was compiled, a needs assessment was completed. A needs assessment is a systematic process for determining and addressing needs or gaps between current conditions and desired conditions. The discrepancy between the current conditions and the wanted condition must be measured to appropriately identify the need. A needs assessment is part of a planning process on how to effectively support the positive attributes of the evaluation system and identify the areas of weakness. By clearly identifying the problem and evaluating the process, the resources being used to train and prepare evaluators can be directed toward developing and implementing a feasible solution (Fulgham & Shaughnessy, 2008).

After determining the current results, the desired results were articulated and the distance between the results are the actual need that is focused upon. In relation to TKES and what evaluators need, a training needs assessment was a systematic inquiry of training needs within an organization for the purposes of identifying priorities and making decisions, and allocating finite resources in a manner consistent with identified program goals and objectives (Triner, Greenberry & Watkins, 1996). There are three levels of needs assessment: Organizational analysis, task analysis, and individual analysis. The focus of this study will be on task analysis:

Provides data about a job or group of jobs and the knowledge, skills, attitudes, and abilities needed to achieve optimum performance (Miller, 2000).

Standards of Rigor for Study

Qualitative research must meet rigorous standards that strive for "understanding" that deep structure of knowledge that comes from visiting personally with participants, spending extensive time in the field, and probing to obtain detailed meanings (Creswell, 2013). The data analysis is proven to be rigorous when the techniques used produce the data necessary to answer the research questions in a reliable and valid method. In conducting a needs assessment for principal training to give meaningful feedback, the interview process is the most efficient and reliable method to use to attain the amount of detail necessary to answer the research questions.

In order to meet the standards of rigor for this study, the researcher developed interview questions that described in detail the lived experiences of the principals interviewed. The interviews conducted in a way that ensured the participants that the data provided were used for the sole purpose of answering the clearly stated research questions. The researcher attempted to not lead the interview in a direction to answer a specific research agenda.

Writers have searched for and found qualitative equivalents that parallel traditional quantitative approaches to validation. Creswell (2013) summarized validation in qualitative research to be an attempt to assess the accuracy of the findings, as best described by the researcher and the participants (p. 249-250). It is the distinct strength of qualitative research in that the account made through extensive time spent in the field, the detailed thick description, and the closeness of the researcher to participants in the study all add to the value or accuracy of the study. Validation, according to Creswell (2013), is done through one or more of the following procedures: Prolonged engagement and persistent observation, triangulation, peer review or debriefing, negative case analysis, clarifying researcher bias, member checking,

rich/thick descriptions, and external audits of the account. The use of at least two of these methods is recommended in proving validity for qualitative research. This study included the following standards of rigor: Triangulation, clarifying researcher bias, member checking, and thick, rich descriptions.

As the data was collected, the researcher looked for two to three points of the interview during different topics that presented the same information. By connecting the multiple points, the researcher showed that the participants' real views and authentic behaviors were included. Connecting the points of view in the individual interviews, focus group interview, and document review served as a way to validate the data.

The researcher clarified bias from the outset of this study so that the reader understands the researcher's position as a principal in the school district being studied and any biases or assumptions that impact the inquiry are addressed (Merriam, 1998). The researcher is bias in the fact that she is an evaluator in the large school district and is experiencing the same phenomenon as the members of the focus group. In addition, the restrictions of the evaluation tool impact her professional duties and responsibilities. Finally, the researcher is a colleague of those that were interviewed, so the bias toward one or another was clarified in terms of their relationships.

Lincoln and Guba (1985) describe this to be "the most critical technique for establishing credibility" (p. 314). Also called *member check* and *respondent validation*, member checking is when the researcher submits materials relevant to a study for checking by the people who were the source of those materials. Member checking relates to the researcher taking the answers to interview questions, once transcribed, back to the interviewees to determine the accuracy of the transcriptions. The exact transcriptions and any themes that evolved from the data analysis were given to the participants to review. Due to the nature of this study, the use of member-checking

was crucial for the researcher to understand what was going on in the social setting of the focus group. When conducting this member checking exercise, the researcher decided what was to be validated. In particular, she determined the comments that were submitted with her interpretations.

The researcher will enable the readers to transfer information to other settings and determine whether the findings can be transferred by using rich, thick descriptions. Stake (2010) stated, "A description is rich if it provides abundant, interconnected details," (p.49). Thick description is described by Lincoln and Guba (1985) as a way of achieving a type of external validity. By describing a phenomenon in sufficient detail, the researcher can begin to evaluate the extent to which the conclusions drawn are transferable to other times, settings, situations, and people. Specifically, the descriptions provided through this study will be transferable to other evaluators in Georgia who are using the TKES instrument. The questions in this study will help lead the members of the focus group into discussions that thoroughly explain the phenomenon being studied.

Situation of Self as Researcher

The researcher in this study has completed fifteen years of service in public education, all of which have been in the state of Georgia. Three years were spent in a school district other than the one studied. The researcher has been an administrator performing teacher evaluations for five years and a principal for one. It is the first year for the researcher to conduct teacher evaluations using TKES. Over the past years, she has worked with numerous school principals whom have different levels of education and experience. The researcher became intrigued about the level of preparedness for principals to give meaningful feedback while beginning to train to implementing the new evaluation instrument herself over a year ago. This, along with

discussions in courses at The University of Alabama, led the researcher to pursue this study on principal preparedness to give meaningful feedback during evaluation.

Due to the fact that the researcher is a principal herself, the perspective of the researcher in this study is that of an active participant. In light of this acknowledgement and the fact that teacher evaluation is part of her own job description, the researcher will be acutely aware of this when conducting data collection and data analysis. The above mentioned standards of rigor, especially member checking, will be used intensely throughout the study.

Summary

The researcher utilized qualitative methods to conduct a needs assessment as a type of program evaluation to examine principal preparedness to conduct effective evaluations. The study will seek to provide a needs assessment that pinpoints the various types of professional development needed for principals in different levels of education. The results of this study will contribute to the literature in that it will provide data which will speak to teacher evaluation and giving feedback to adults for optimal learning.

Chapter III is organized so that two purposes could be achieved. The first purpose was to describe the research design and methods chosen to conduct this research. The research design of this study is a qualitative program evaluation approach. The researcher has provided a description of the design, as well as the methods to be used to answer the research questions. The second purpose of the chapter was to examine how the design will address the research questions raised in the study. The study will specifically target principals who are beginning to use a new instrument for teacher evaluation. Finally, the needs assessment will allow for optimal use of professional learning resources in school districts similar to the one studied, in the future. The following chapter will discuss the results obtained from the interview processes and documents researched.

CHAPTER 4: DATA ANALYSIS

The purpose of this qualitative study was to determine the perceived level of self-preparedness of building level principals to conduct effective teacher evaluations and to identify the future professional development needs while using the Georgia Teacher Keys Effectiveness System. Through evaluations of teacher effectiveness a building principal is one of the most important factors in helping teachers improve their practice and changes in teacher evaluation, brought on primarily through federal legislation, has mandated changes in the current teacher evaluation process. These two foci primarily led to the inquiry of perceived principal preparation to conduct effective teacher evaluations and identification of future professional development needs and resulted in the formation of this study.

In this chapter, findings from this study are analyzed in the context of the research questions initially presented in Chapter one and connections to the literature reviewed are made. Data were collected from multiple responses to determine the specific preparation and professional development or resources building principals believe are to help provide more effective feedback as they evaluate teachers. The following research questions were developed as the framework for this study and served as the foundation for the data analysis:

1. How do building principals describe their preparation to give effective feedback while using the evaluation instrument?
2. What are their feelings toward the instrument in contrast to others used in the past?
3. What are benefits and challenges of giving feedback?

4. What additional professional development resources do evaluators report they need to become effective evaluators?

Within this chapter, data collected for each of the four research questions, as well as the data analysis methods used is examined. The primary procedure used to collect data for this study was an interview protocol. Along with some document review, an interview protocol is sometimes described as a conversation with a purpose. The conversation is a person-to-person encounter in which one person elicits information from another (Merriam, 1998). During the interviews, notes were taken as well as a taped recording made of the conversations. Each recording was transposed at a later time. Each of the nine semi-structured interviews took between forty-five to sixty minutes and the focus group interview took around ninety minutes. At the end of each, the respondents were given the opportunity to ask questions or add to the information they provided. In addition, each transcription was sent to the individual respondents to validate responses recorded.

All of the interviews for this study were semi-structured, a mixture of both structured and flexible questions. The primary reason to choose a semi-structured format was the fact that this format allows the researcher to fully investigate the respondents' realities through the process of adding questions for clarification when necessary. Part of the preparation for data collection was to develop interview questions that helped clarify the respondents' answers to the research questions posed by the study. In addition, the interview questions addressed each research question in several ways. Some of the questions asked the principals to draw upon their experiences or behaviors while other questions elicited what the principal thought about their experiences or behaviors, which could reveal their opinions or values. Additionally, the interview questions addressed the research questions by probing for feelings – how the

respondent reacts emotionally to their experiences, knowledge – factual knowledge the person has or what the person considers factual, and background or demographic information – how the respondent describes him/herself in relationship to other people (McMillan & Schumaker, 2006). Once the interviews were completed, decisions regarding analysis of the collected material were made.

In answering the four research questions the data were analyzed through the use of descriptive statistics taken from the transcriptions. The data displays the events of the phenomenon under study by identifying subthemes within the research and a way to interpret the collected data. Each subtheme was linked to an overarching theme, which corresponded to an interview question. Through open coding, a process of discovering concepts that are the basic building blocks of theory (Strauss & Corbin, 2008), the data generated three themes: Inter-rater reliability, training, and experience.

This chapter is organized in seven sections: An introduction, a demographic profile of the study sample, reflections on the data collection and analysis process, emergent codes, logic, and schemes, research questions presentation, residual study data, and a summary of study findings.

Demographic Profile of Study Sample

Descriptive statistics is used when the purpose of an investigation is to describe the data that have been, or will be, collected. Knowing the demographic background and professional experiences of the respondents is important as it is from these experiences and perspectives that they answered the questions during the individual interviews and focus group for this study. Nine participants from a suburban school district in Georgia participated in this study, four (44.5%) were male and five (55.5%) were female. There were eight Caucasian and one was African-American. Numbers were assigned to the principals to protect their identity. The results

from this study can be applicable when looked at from a gender perspective, as it would be expected that all females and males would respond similarly, regardless of location, since all have similar guidelines and mandates.

All principals' currently work in the same school district. The district serves approximately one-hundred and eleven thousand students and is the second largest in the state, twenty-fourth largest in the United States. There are one-hundred and fourteen campuses: Sixteen high schools, twenty-five middle schools, sixty-seven elementary schools, two charter schools, two special education centers, one adult center and one performance learning center. Twenty-two of those schools have been named National Blue Ribbon Schools and fifty-six have been named Georgia Schools of Excellence according to the results from the state assessment scores. There are five thousand, seven hundred-seventy-seven classroom teachers and one thousand, three hundred-thirty-one special education teachers, all of whom will be evaluated using the TKES instrument. Demographics reveal that the student population is made up of: White 39.8%, Black 31.9%, Hispanic 19.5%, Asian 5.1%, Multi-Racial 3.4%, American Indian <1%, and Pacific Islander <1%. The district has a graduation rate of 78.2% (2014), the percentage of students qualifying for free/reduced lunch is 45.0%, and the transiency rate is 22.64%. Although it was not specifically analyzed in this study, the size of the district could have determined the evaluation systems used prior to RTTT mandates and TKES, therefore impacting the perceptions of teacher evaluation. Student enrollment gives a general sense of the type of district and setting in which respondents work. This could be used for comparison with similar future studies.

The principals that participated in this study were nine building level principals. After approval from the district and University of Alabama to conduct this study, the researcher

contacted the principals of three elementary schools, three middle schools, and three high schools. Each principal has been employed by the district for over ten years and has been a building principal for at least one full year and has implemented at least one evaluation instrument other than the current one under study (TKES). Table 1 below shows the experience of each principal in this study.

Table 1: *Principal Years of Experience*

Principal	Level	Experience in Education	Experience in CCSD	Experience as an Administrator	Experience as a Principal
1	Middle	22 Years	22 Years	11 Years	7 Years
2	Middle	26 Years	24 Years	11.5 Years	3 Years
3	Middle	27 Years	22 Years	16 Years	11 Years
4	High	25 Years	22 Years	10 Years	4 Years
5	High	18 Years	11 Years	5 Years	6 Years
6	Elementary	18 Years	18 Years	9 Years	1 Year, 5mths
7	High	21 Years	21 Years	15 Years	12 Years
8	Elementary	22 Years	02 Years	12 Years	6 Years
9	Elementary	25 Years	25 Years	12.5 Years	1.5 Years

Teachers are evaluated every year by the campus administrative team and this is the first year for the school district to use TKES. TKES requires evaluators to complete four individual walk-throughs, two observations, two formative assessments, and one summative assessment on each student. In addition, evaluators are charged with the responsibility of analyzing student survey data and end of course test results when awarding teachers with a rating in each of the ten standards addressed in TKES. All of the above mentioned principals went through initial credentialing with the State of Georgia to be able to implement TKES.

Reflections on Data Collection and Analysis Process

There were a few significant challenges that the researcher encountered during the data collection and analyzing process. First, a very surface level challenge that was anticipated prior

to the start of the research was time. Time was a challenge because every building principal is extremely busy and usually booked from the beginning of the day until the end of the day. The challenge was met and overcome by the researcher being flexible in her own time and willing to travel to make the interviews as convenient as possible. In addition, the researcher was precise and did not waste any time when conducting the interviews. The researcher was also at the mercy of the participants when member checking to make sure the transcriptions were correct prior to moving on with the coding process.

The second challenge was recognizing and keeping constant awareness of researcher bias. As a principal herself, the researcher is also conducting evaluations using TKES and has strong opinions on the instrument and how feedback is given to teachers. The challenge was met by conducting mock interviews with a couple principals not involved in the study. Upon listening to the interviews, the researcher recognized that she spoke too often and gave many of her own opinions. During the actual research interviews, the researcher kept constant reminders on her recording device to keep herself from giving personal thoughts on how the participants were responding or how she felt about the actual questions. Although this was a challenge, it was also a great way for reflection while listening to the interviews and transcribing them. Separating herself from the data helped the researcher become aware of the creation of the overall themes with limited personal bias.

The challenge of getting the focus group together was the first one the researcher faced. Scheduling at a time convenient for nine principals is almost impossible, so the researcher conducted the focus group at a location as neutral as possible and offered food and beverages. The nine principals came on time, but the urgency of completing the interview was apparent in a few of the participants in regards to their body language. As the questions were asked there was

not a challenge of engaging the participants, there was a natural progression of the conversation that the researcher was quite pleased with.

A challenge that did arise was in asking the participants to look at the TKES training manual as part of the document review section of this study's data collection. The purpose was to gain an understanding of how familiar the participants were with the document. It was apparent that they were not particularly clear on the three-hundred, fifty-eight page document and have not navigated through it very much. The researcher took note of this and used that information as a follow-up question in the second individualized interviews.

The biggest epiphany came to the researcher rather quickly during the data collection process. After the third interview an overwhelming theme consumed the data being collected. It was one that the researcher had not anticipated. The theme of principals needing inter-rater reliability within their building developed from the first question and maintained frequency throughout the last question. Although the term was anticipated and known by the researcher, the intense desire and need for more opportunities to reach success in this area of the evaluation process for their teachers was not. It was overwhelming to the researcher to hear the urgency in the participants' responses during the interviews and after while transcribing the interviews.

Another realization was that the researcher began the data collection process expecting to hear how principals wanted more training on how to give appropriate feedback. What revealed itself during the interviews was that principals *did not* feel they needed future professional learning in that area, but rather on the actual instrument and the way their ratings compared to those in their building and the district. As the direction of *need* changed according to the codes evolving from the responses, the researcher used the flexibility of the semi-structured interview process to ask follow up questions that directly related to *how* the need could be addressed

through professional experiences. The eye-opener was that the professional development was not necessarily a nicely packaged class or session that principals could attend and take ideas away from, but rather real-life experiences that evolved into conversations with other professionals about how they use the instrument and rate/evaluate teachers.

A discovery arose during the focus group interview. It was that the conversation was guided and completing led by the group itself. The researcher knew that this was the intention of the group interview, but was pleasantly surprised when the event took a specific, participant-generated path. The researcher started the interview with the question of what future preparation the principals would like to have in the area of TKES and the themes that were generated during the individual interviews were naturally brought up and discussed as a group. Again, it was apparent when asked about the TKES handbook that the participants have not accessed it often, and the agreement that there just isn't enough time in the day to sit down and look through such an in-depth, long document was reached much quicker than the researcher anticipated.

For the most part, the interviews were conducted at the participants' schools. Table 2 shows the time, location, and date of the interview. The school name has been changed to keep the participants' identities confidential.

Table 2: *Interview Information*

Principal	School/Interview Location	Date	Time
P1	SMS/SMS	1/23/15	10:00-10:45am
P2	DMS/DMS	1/23/15	2:30-3:15pm
P3	HTMS/HTMS	1/23/15	3:45-4:30pm
P4	PHS/PHS	1/23/15	5:00-5:45pm
P5	SCHS/SCHS	1/26/15	12:45-1:30pm
P6	MBES/WHS	1/27/15	12:00-12:30pm
P7	WHS/WHS	1/27/15	12:45-1:45pm
P8	SPES/Off-Campus	1/28/15	12:00-12:45pm
P9	MES/MES	2/2/15	9:00-9:30am
Focus Group	ALL/TES	2/2/15	5:00-6:30pm

The first interview with Principal One (P1) was in his office at his middle school around 10:00 in the morning. The interview lasted about forty-five minutes and, although she did not know the participant well, it was very comfortable for the interviewee. As the interview began, the researcher gave a brief over-view of the study and the nature of the data that would be collected. She also gained a signature for the appropriate documents for IRB. The questions were followed as planned and additional, clarifying questions or statements were made when needed. Upon reflection of this first interview, there are questions the researcher could have asked, but being the first interview, did not recognize at the time.

The second interview with Principal Two (P2) was in her office in her middle school at 2:30 in the afternoon. The interview lasted around forty-five minutes as well, but this interview was a little more focused on the nature of the way she and her administrators give feedback to teachers (rather than her background in conducting evaluations). The interview lent itself to more of a conversation than the first one, but the researcher remained very aware to not add her own bias in the formation of statements. Toward the end of the interview, an assistant principal came into the conversation and changed the tone of the interview. The principal went from answering the questions and having a conversation to addressing the concerns that her administrative team has brought up throughout the process.

The third interview with Principal Three (P3) was in her office in her middle school at 3:45 in the afternoon. The time spent with this principal lasted longer than the first two because of conversations had after the actual interview. Although all questions were answered, the interview focused on the consistency between the members of her administrative team due to the fact that two of the three are new to evaluating teachers. A great deal of time was spent in

describing how the administrative team has worked together to reach agreement in using the instrument, mostly due to the seniority of the principal and her knowledge of the staff.

The fourth interview with Principal Four (P4) was in his office at his high school at 5:00 in the afternoon. The setting and tone of the interview seemed much more "professional" than the first three. The researcher felt as though the principal was more timid in answering the questions and the first fifteen minutes of the interview were spent clarifying why the questions were important. Once the conversation turned to giving teachers feedback, the principal was more open in his responses. The interview lasted around forty-five minutes.

The fifth interview was in a school that serves a more diverse population than the first four. The researcher traveled to the South end of the county to interview Principal Five (P5) at 12:45 in the afternoon. After a brief tour of the school, the interview lasted around forty-five minutes. The focus of this interview seemed to be on the principal making a specific point to mention that he felt he needed to evaluate all of his new teachers and help them grow. He answered all of the questions, but always put an emphasis on the importance of his being the lead evaluator for the new teachers and helping his administrative team grow and learn through his lead.

The sixth and seventh interviews were conducted after a principal's meeting at 12:00 in the afternoon. They both took place in Principal Seven's (P7) high school. Due to the fact that they were after a meeting and the time constraints of the principals, the first with Principal Six (P6) seemed a little more rushed, lasting about thirty minutes. Although all of the interview questions were answered, the researcher made notes on the topics that were not covered as in depth to re-visit, if need be, during the second individual interview. In addition, this principal tended to be very cautious of how she answered the questions. Principal Seven was very willing

to talk and the interview last about an hour. He was open and very opinionated about the way the evaluations are done and his thoughts on how they could be better. He answered all of the questions and did not need much guiding in terms of follow up comments.

Principal Eight (P8) was interviewed at a church following another principal meeting at 12:00 in the afternoon. She was willing and able to answer all of the questions and was the first to reference the use of the TKES handbook and other State resources. Her interview lasted about forty-five minutes and was guided by the interview questions. The school served by this principal is a lower socio-economic school and she made mention of the numerous other duties she has (being a Title One school) and the time required by the evaluation instrument.

Principal Nine (P9) was interviewed at her school for about thirty minutes following a collaboration meeting with Assistant Principals and other Principals. The principal was very serious and engaged in answering the questions provided, but did not give much detail outside of the exact questions asked. The room that the interview was given, as was the case for all of the above mentioned interviews, was quiet, comfortable, and private.

The Focus Group interview was one of the most intriguing to the researcher in terms of how the participants engaged with one another. The setting was the researcher's elementary school at 5:00 in the evening. The researcher provided beverages and food for the participants and made sure that the room was comfortable with enough seating, which was strategically placed so the participants could look at one another. As the conversation started, the participants were reluctant to begin. Quickly, that changed, and the researcher had to do very little to guide the progression of the interview. During the interview there were subtle cues that participants were not comfortable with what or how others responded. Sometimes they said something, sometimes they just made a gesture (rolling eyes, re-crossing legs, moving in the chair), and

sometimes they made a verbal remark that was not a comment but more of a tone of agreement or disagreement. The researcher spent the majority of this interview listening and only had to re-engage the group once.

It is through following the standards of rigor of qualitative research that the researcher learned the most about the topic of future needs for principal preparation in using an evaluation instrument, specifically, TKES. Understanding the deep structure of knowledge that comes from visiting personally with participants, spending extensive time in the field, and probing to obtain detailed meanings (Creswell, 2013) led to a deeper submersion in the data being collected through the interviews. In order to produce the data necessary to answer the research questions in a reliable and valid method, rigorous standards were followed throughout the data collection period. In conducting a needs assessment for future principal preparation to implement an effective teacher evaluation with meaningful feedback, the interview process proved to be the most efficient and reliable method to use to attain the amount of detail necessary to answer the research questions. In order to meet the standards of rigor for this study, the researcher developed interview questions that described the lived experiences of the principals interviewed. The interviews were conducted in a way that ensured the participants that the data provided would be used for the sole purpose of answering the clearly stated research questions. The researcher attempted not to lead the interview in a direction to answer a specific research agenda. After each interview, the researcher validated the data in several ways. First, the researcher clarified her bias from the outset of this study so that the interviewees understood the researcher's position as a colleague in the school district being studied. Through a "practice" round of interviews with principals not in the same school district, the researched realized she talked about her personal experiences too much. In an attempt to avoid this, the researcher kept

a visual at the top of her recording device to remind her to not give her own feedback to the questions. Therefore, through a careful and conscious attempt, any biases or assumptions that might impact the responses were avoided by not discussing personal opinions or practices while the interview questions were being presented and answered.

Lincoln and Guba (1985) describe member checking as "the most critical technique for establishing credibility" (p. 314). Also called *respondent validation*, member checking is when the researcher submits materials relevant to a study for checking by the people who were the source of those materials. After the interviews were completed, the researcher transposed the interviews into a Word document and sent the document to the individual participant through an email attachment. The researcher verified that each transposition was valid and correct in the eyes of the interviewee. The importance of this, as stated by Creswell (2013), is that qualitative research is an attempt to assess the accuracy of the findings, as best described by the researcher and the participants (p. 249-250). If something was interpreted incorrectly or a piece of information left out, the researcher clarified the original document sent to the participant.

As the data was collected, the researcher looked for two to three points of the interview during different topics that presented the same information. In connecting the multiple points, the researcher found that the participants' real views and authentic behaviors were included in the results. Connecting the points of view in the individual interviews, focus group interview, and document review served as a way to validate the data through triangulation. During the focus group, especially, specific points of view from various participants are highlighted and explained later in this chapter.

During the explanation of the results, rich and thick descriptions will enable the readers to transfer information to other settings. Thick description is described by Lincoln and Guba

(1985) as a way of achieving a type of external validity. By describing the phenomenon of future principal preparation to conduct effective teacher evaluations, the researcher has begun to evaluate the extent to which the conclusions drawn are transferable to other times, settings, situations, and people. Specifically, the descriptions provided through this study will be transferable to other evaluators in Georgia who are using the TKES instrument.

Document Review

The Teacher Keys Effectiveness System Handbook was the main document reviewed for this study. The TKES Handbook was developed on behalf of the Georgia Department of Education to assist in the implementation of Georgia's Race to the Top (RT3) plan. School systems involved in the pilot/full year implementation are required to use the handbook. The materials in the handbook are copyrighted by either the Georgia Department of Education of Dr. James Stronge. The handbook is organized into four chapters. Throughout chapter one, the TKES implementation handbook, there are links to chapters two through four. The handbook is organized in the following manner:

> Chapter One: Teacher Keys Effectiveness System Implementation Handbook
> Chapter Two: Teacher Keys Effectiveness System Fact Sheets
> Chapter Three: Teacher Keys Effectiveness Research Synthesis
> Chapter Four: Teacher Keys Effectiveness System Endnotes for Handbook, Fact Sheets, and Research Synthesis

All of the information was not studied in this research. However, the components that were are explained as needed throughout the description of how the coding was completed and themes derived. In addition, the researcher accessed the handbook, when needed, to explain or clarify comments made by the participants. When asked how often the handbook was used in this first year of implementation, responses varied from, "I only referenced the manual in the very early stages of the implementation. As I recall, I used it to double check the rubrics and make sure I had entered everything correctly" (Principal one) to "I use the handbook during A-

team meeting when we have a discrepancy in what we each heard during our trainings. We referenced it several times at the beginning of the year but have not gone back to it since probably September," (Principal seven). The remaining principals noted that they have never referenced the handbook. The contents of the TKES Handbook are illustrated in Figure 3.

Codes and Themes

As the coding for the data began, the researcher found herself changing the process many times until the data made sense. Through the coding process, future professional development needs for principal preparation to implement the evaluation instrument effectively are identified using the three consensus themes that emerged: inter-rater reliability, training, and experience.

Inter-rater reliability, inter-rater agreement, or concordance is the degree of agreement among raters. It gives a score of how much homogeneity, or consensus, there is in the ratings given by those who judge something. In this case, the principals are the judges (i.e. evaluators). The inter-rater reliability in regards to TKES is the concurrence between ratings that are earned by the teacher. For example, if one evaluator rates an area as a "3" will another rater do the same? There are multiple visits to a classroom by different evaluators mandated with the TKES instrument to help reach inter-rater reliability. Training, for the use in this research, can be defined as the staff development, or professional development that principals have already, or wish to, experience. TKES training, specifically, was the same for most of these principals. It was a two-day experience led by a Georgia Department of Education professional. Professional development encompasses all types of facilitated learning opportunities including credentials such as academic degrees to formal coursework, conferences and informal learning opportunities situated in practice. Within the context of TKES, professional development lends itself to the number of times an evaluator was able to attend a workshop or meeting that helped give him or her further explanation as to how to properly use the instrument. Experience is the knowledge or

mastery of an event or subject gained through involvement in or exposure to it. A principal can gain experience in education through on-the-job training and book-learning. Experience for the participants in this study is vast and varied. All of the principals have at least two years of evaluator experience; some have post-graduate degrees; most have attended further professional development for TKES; all have a background as being an instructional lead teacher; and the majority have been principals for over five years.

TKES is obviously an instrument that involves using humans as the collector of that data. When humans are part of the measurement procedure, one must worry about whether the results he or she gets are reliable or consistent. The participants in this study noted that there were several steps they have taken in controlling the inter-rater reliability in their building, but it is an area to focus on for future preparation. The principals stated throughout the data collection period, in many different ways, that this is an area they would benefit from more consistency throughout their levels, areas (Cobb County Schools is separated into six areas), and the entire district and state. The codes led to the following sub-themes: standards, feedback, and interpreting survey data. After highlighting the key terms, the researcher used a red pen to note the theme of inter-rater reliability and then labeled the sub-themes with a letter (A, B, C).

There are ten standards in TKES that the evaluators use to rate the teachers. The ten standards, as described in the TKES Handbook are:

1. Professional Knowledge – The teacher demonstrates an understanding of the curriculum, subject content, pedagogical knowledge, and the needs of student by providing relevant learning experiences
2. Instructional Planning – The teacher plans using state and local school district curricula and standards, effective strategies, resources, and data to address the differentiated needs of all students.
3. Instructional Strategies – The teacher promotes student learning by using research based instructional strategies relevant to the content to engage students in active learning and to facilitate the students' acquisition of key knowledge and skills.

4. Differentiation – The teacher challenges and supports each student's learning by providing appropriate content and developing skills which address individual learning differences.
5. Assessment Strategies – The teacher systematically chooses a variety of diagnostic, formative, and summative assessment strategies and instruments that are valid and appropriate for the content and student population.
6. Assessment Uses – The teacher systematically gathers, analyzes, and uses relevant data to measure student progress, to inform instructional content and delivery methods, and to provide timely and constructive feedback to both students and parents.
7. Positive Learning Environment – The teacher provides a well-managed, safe, and orderly environment that is conducive to learning and encourages respect for all.
8. Academically Challenging Environment – The teacher creates a student-centered, academic environment in which teaching and learning occur at high levels and students are self-directed learners.
9. Professionalism – The teacher exhibits a commitment to professional ethics and the school's mission, participates in professional growth opportunities to support student learning, and contributes to the profession.
10. Communication – The teacher communicates effectively with student, parents or guardians, district and school personnel, and other stake holders in ways that enhance student learning.

A teacher is visited, at minimum, six times throughout the year from his/her evaluators. There are two complete cycles including two ten-minute walkthroughs and one thirty minute observation. Upon completion of the observation, a formative assessment is completed using the totality of evidence from the entire cycle. A summative assessment is completed at the end of the school year. A walkthrough is defined as a more frequent, brief observation that focuses on a limited number (typically 1 to 4). The purpose of these walkthroughs is to help establish the frequency and consistency of appropriate classroom practices as identified within TKES. Walkthroughs provide glimpses into the regular practices of teachers and should be reflective of observations and of documentation that teachers provide. Data from the Walkthrough observations will be used to support and enhance performance ratings on the Formative Assessment and in the Summative Assessment. As evaluators conduct observations in a teacher's classroom, they continually build a portrait of that teacher's approach to and implementation of instructional practices. Over time, these observations should demonstrate the

consistency of a teacher's performance. Although many practices and instructional strategies should be directly observed both in the walkthroughs and in the formative observation, other information can and should be considered in the ratings for the formative assessment (i.e. walkthroughs, lesson plans, student work samples and other forms of documentation). Each Formative Assessment will be directly tied to a formative observation. Once the formative observation has occurred, evaluators will consider the full scope of a teacher's practices that have been observed and documented up to and including the formative observation. Following the formative observation, if sufficient evidence is not present to rate a teacher's performance on a given standard or standards, evaluators can request specific documentation relevant to the aligned standard(s). Documentation should supplement evidence and practices observed in the course of a teacher's professional practice, but documentation alone should not account for a rating on the formative assessment. Practices cited in documentation should be supported by observed practices. Following the first Formative Assessment, the cycle of observing practice and collecting documentation should begin again for the second Formative Assessment cycle.

During the two Formative Assessments and the final Summative Assessment, the teachers receive a rating on each of the standards. During the Walkthroughs, building level evaluators determine how many and which standards to rate. The ratings are based on a rubric and the teachers receive a number from 1(low) to 4 (high). Proficient (3) is the expected level of performance from teachers. The following figures provide a richer description of the levels of performance. Figure 4 illustrates a richer description of the levels of performance as provided by the Georgia Department of Education. Figure 5 (front and back) is a reference guide to help define the ratings for evaluators.

Georgia Department of Education · TAPS Standards Reference Sheet
Performance Standards and SAMPLE Performance Indicators *
*Performance Indicators are not inclusive and should not be used as a checklist.

Planning

1. Professional Knowledge

The teacher demonstrates an understanding of the curriculum, subject content, pedagogical knowledge, and the needs of students by providing relevant learning experiences.

- Addresses appropriate curriculum standards and integrates key content elements.
- Facilitates students' use of higher-level thinking skills in instruction.
- Demonstrates ability to link present content with past and future learning experiences, other subject areas, and real-world experiences and applications.
- Demonstrates accurate, deep, and current knowledge of subject matter.
- Exhibits pedagogical skills relevant to the subject area(s) taught and best practice based on current research.
- Bases instruction on goals that reflect high expectations for all students and a clear understanding of the curriculum.
- Displays an understanding of the intellectual, social, emotional, and physical development of the age group.

2. Instructional Planning

The teacher plans using state and local school district curricula and standards, effective strategies, resources, and data to address the differentiated needs of all students.

- Analyzes and uses student learning data to inform planning.
- Develops plans that are clear, logical, sequential, and integrated across the curriculum (e.g., long-term goals, lesson plans, and syllabi).
- Plans instruction effectively for content mastery, pacing, and transitions.
- Plans for instruction to meet the needs of all students.
- Aligns and connects lesson objectives to state and local school district curricula and standards, and student learning needs.
- Develops appropriate course, unit, and daily plans, and is able to adapt plans when needed.

Instructional Delivery

3. Instructional Strategies

The teacher promotes student learning by using research-based instructional strategies relevant to the content to engage students in active learning and to facilitate the students' acquisition of key knowledge and skills.

- Engages students in active learning and maintains interest.
- Builds upon students' existing knowledge and skills.
- Reinforces learning goals consistently throughout the lesson.
- Uses a variety of research-based instructional strategies and resources.
- Effectively uses appropriate instructional technology to enhance student learning.
- Communicates and presents material clearly, and checks for understanding.
- Develops higher-order thinking through questioning and problem-solving activities.
- Engages students in authentic learning by providing real-life examples and interdisciplinary connections.

4. Differentiated Instruction

The teacher challenges and supports each student's learning by providing appropriate content and developing skills which address individual learning differences.

- Differentiates the instructional content, process, product, and learning environment to meet individual developmental needs.
- Provides remediation, enrichment, and acceleration to further student understanding of material.
- Uses flexible grouping strategies to encourage appropriate peer interaction and to accommodate learning needs/goals.
- Uses diagnostic, formative, and summative assessment data to inform instructional modifications for individual students.
- Develops critical and creative thinking by providing activities at the appropriate level of challenge for students.
- Demonstrates high learning expectations for all students commensurate with their developmental levels.

Assessment Of and For Learning

5. Assessment Strategies

The teacher systematically chooses a variety of diagnostic, formative, and summative assessment strategies and instruments that are valid and appropriate for the content and student population.

- Aligns student assessment with the established curriculum and benchmarks.
- Involves students in setting learning goals and monitoring their own progress.
- Varies and modifies assessments to determine individual student needs and progress.
- Uses formal and informal assessments for diagnostic, formative, and summative purposes.
- Uses grading practices that report final mastery in relationship to content goals and objectives.
- Uses assessment techniques that are appropriate for the developmental level of students.
- Collaborates with others to develop common assessments, when appropriate.

6. Assessment Uses

The teacher systematically gathers, analyzes, and uses relevant data to measure student progress, to inform instructional content and delivery methods, and to provide timely and constructive feedback to both students and parents.

- Uses diagnostic assessment data to develop learning goals for students, to differentiate instruction, and to document learning.
- Plans a variety of formal and informal assessments aligned with instructional results to measure student mastery of learning objectives.
- Uses assessment tools for both formative and summative purposes to inform, guide, and adjust instruction.
- Systematically analyzes and uses data to measure student progress, to design appropriate interventions, and to inform long- and short-term instructional decisions.
- Shares accurate results of student progress with students, parents, and key school personnel.
- Provides constructive and frequent feedback to students on their progress toward their learning goals.
- Teaches students how to self-assess and to use metacognitive strategies in support of lifelong learning.

Georgia Department of Education • July 1, 2014 • Page 1 of 2

Figure 1. TAPS Standards Reference Sheet

This figure was one of the rubrics to help principals navigate the standards.

Georgia Department of Education
TKES Performance Standards and Rubrics

Performance Standard 1: Professional Knowledge *The teacher demonstrates an understanding of the curriculum, subject content, pedagogical knowledge, and the needs of students by providing relevant learning experiences.*			
Exemplary *In addition to meeting the requirements for Proficient…*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually demonstrates extensive content and pedagogical knowledge, enriches the curriculum, and guides others in enriching the curriculum. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher consistently demonstrates an understanding of the curriculum, subject content, pedagogical knowledge, and the needs of students by providing relevant learning experiences.	The teacher inconsistently demonstrates understanding of curriculum, subject content, pedagogical knowledge, and student needs, or lacks fluidity in using the knowledge in practice.	The teacher inadequately demonstrates understanding of curriculum, subject content, pedagogical knowledge and student needs, or does not use the knowledge in practice.
Performance Standard 2: Instructional Planning *The teacher plans using state and local school district curricula and standards, effective strategies, resources, and data to address the differentiated needs of all students.*			
Exemplary *In addition to meeting the requirements for Proficient…*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually seeks and uses multiple data and real world resources to plan differentiated instruction to meet the individual student needs and interests in order to promote student accountability and engagement. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher consistently plans using state and local school district curricula and standards, effective strategies, resources, and data to address the differentiated needs of all students.	The teacher inconsistently uses state and local school district curricula and standards, or inconsistently uses effective strategies, resources, or data in planning to meet the needs of all students.	The teacher does not plan, or plans without adequately using state and local school district curricula and standards, or without using effective strategies, resources, or data to meet the needs of all students.
Performance Standard 3: Instructional Strategies *The teacher promotes student learning by using research-based instructional strategies relevant to the content to engage students in active learning and to facilitate the students' acquisition of key knowledge and skills.*			
Exemplary *In addition to meeting the requirements for Proficient…*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually facilitates students' engagement in metacognitive learning, higher-order thinking skills, and application of learning in current and relevant ways. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher consistently promotes student learning by using research-based instructional strategies relevant to the content to engage students in active learning, and to facilitate the students' acquisition of key skills.	The teacher inconsistently uses research-based instructional strategies. The strategies used are sometimes not appropriate for the content area or for engaging students in active learning or for the acquisition of key skills.	The teacher does not use research-based instructional strategies, nor are the instructional strategies relevant to the content area. The strategies do not engage students in active learning or acquisition of key skills.
Performance Standard 4: Differentiated Instruction *The teacher challenges and supports each student's learning by providing appropriate content and developing skills which address individual learning differences.*			
Exemplary *In addition to meeting the requirements for Proficient…*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually facilitates each student's opportunities to learn by engaging him/her in critical and creative thinking and challenging activities tailored to address individual learning needs and interests. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher consistently challenges and supports each student's learning by providing appropriate content and developing skills which address individual learning differences.	The teacher inconsistently challenges students by providing appropriate content or by developing skills which address individual learning differences.	The teacher does not challenge students by providing appropriate content or by developing skills which address individual learning differences.
Performance Standard 5: Assessment Strategies *The teacher systematically chooses a variety of diagnostic, formative, and summative assessment strategies and instruments that are valid and appropriate for the content and student population.*			
Exemplary *In addition to meeting the requirements for Proficient…*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually demonstrates expertise and leads others to determine and develop a variety of strategies and instruments that are valid and appropriate for the content and student population and guides students to monitor and reflect on their own academic progress. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher systematically and consistently chooses a variety of diagnostic, formative, and summative assessment strategies and instruments that are valid and appropriate for the content and student population.	The teacher inconsistently chooses a variety of diagnostic, formative, and summative assessment strategies or the instruments are sometimes not appropriate for the content or student population.	The teacher chooses an inadequate variety of diagnostic, formative, and summative assessment strategies or the instruments are not appropriate for the content or student population.

Georgia Department of Education
TKES Performance Standards and Rubrics

Performance Standard 6: Assessment Uses *The teacher systematically gathers, analyzes, and uses relevant data to measure student progress, to inform instructional content and delivery methods, and to provide timely and constructive feedback to both students and parents.*			
Exemplary *In addition to meeting the requirements for Proficient...*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually demonstrates expertise in using data to measure student progress and leads others in the effective use of data to inform instructional decisions. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher systematically and consistently gathers, analyzes, and uses relevant data to measure student progress, to inform instructional content and delivery methods, and to provide timely and constructive feedback to both students and parents.	The teacher inconsistently gathers, analyzes, or uses relevant data to measure student progress, inconsistently uses data to inform instructional content and delivery methods, or inconsistently provides timely or constructive feedback.	The teacher does not gather, analyze, or use relevant data to measure student progress, to inform instructional content and delivery methods, or to provide feedback in a constructive or timely manner.
Performance Standard 7: Positive Learning Environment *The teacher provides a well-managed, safe, and orderly environment that is conducive to learning and encourages respect for all.*			
Exemplary *In addition to meeting the requirements for Proficient...*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually engages students in a collaborative and self-directed learning environment where students are encouraged to take risks and ownership of their own learning behavior. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher consistently provides a well-managed, safe, and orderly environment that is conducive to learning and encourages respect for all.	The teacher inconsistently provides a well-managed, safe, and orderly environment that is conducive to learning and encourages respect for all.	The teacher inadequately addresses student behavior, displays a negative attitude toward students, ignores safety standards, or does not otherwise provide an orderly environment that is conducive to learning or encourages respect for all.
Performance Standard 8: Academically Challenging Environment *The teacher creates a student-centered, academic environment in which teaching and learning occur at high levels and students are self-directed learners.*			
Exemplary *In addition to meeting the requirements for Proficient...*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually creates an academic learning environment where students are encouraged to set challenging learning goals and tackle challenging materials. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher consistently creates a student-centered, academic environment in which teaching and learning occur at high levels and students are self-directed learners.	The teacher inconsistently provides a student-centered, academic environment in which teaching and learning occur at high levels or where students are self-directed learners.	The teacher does not provide a student-centered, academic environment in which teaching and learning occur at high levels, or where students are self-directed learners.
Performance Standard 9: Professionalism *The teacher exhibits a commitment to professional ethics and the school's mission, participates in professional growth opportunities to support student learning, and contributes to the profession.*			
Exemplary *In addition to meeting the requirements for Proficient...*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually engages in a high level of professional growth and application of skills and contributes to the development of others and the well-being of the school and community. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher consistently exhibits a commitment to professional ethics and the school's mission, participates in professional growth opportunities to support student learning, and contributes to the profession.	The teacher inconsistently supports the school's mission or seldom participates in professional growth opportunities.	The teacher shows a disregard toward professional ethics or the school's mission or rarely takes advantage of professional growth opportunities.
Performance Standard 10: Communication *The teacher communicates effectively with students, parents or guardians, district and school personnel, and other stakeholders in ways that enhance student learning.*			
Exemplary *In addition to meeting the requirements for Proficient...*	**Proficient** *Proficient is the expected level of performance.*	**Needs Development**	**Ineffective**
The teacher continually uses communication techniques in a variety of situations to proactively inform, network, and collaborate with stakeholders to enhance student learning. *(Teachers rated as Exemplary continually seek ways to serve as role models or teacher leaders.)*	The teacher communicates effectively and consistently with students, parents or guardians, district and school personnel, and other stakeholders in ways that enhance student learning.	The teacher inconsistently communicates with students, parents or guardians, district and school personnel, or other stakeholders or communicates in ways that only partially enhance student learning.	The teacher inadequately communicates with students, parents or guardians, district and school personnel, or other stakeholders by poorly acknowledging concerns, responding to inquiries, or encouraging involvement.

The contents of this form were developed under a grant from the U. S. Department of Education. However, those contents do not necessarily represent the policy of the U. S. Department of Education, and you should not assume endorsement by the Federal Government.

Figure 2. TKES Performance Standards and Rubric.

This figure illustrates the reference guide provided to all evaluators to help define the ratings.

Feedback occurs when outputs of a system are "fed back" as inputs as part of a chain of cause-and-effect that forms a circuit or loop. In this research, the system is inferred to be evaluation instrument and all components that comprise what a teacher is rated on within it. Simple causal reasoning about a feedback system is difficult because the first system influences the second and second system influences the first, leading to a circular argument. This makes reasoning based upon cause and effect tricky, and it is necessary to analyze the system as a whole. Specifically, with teacher evaluation, the feedback given to a teacher from one experience influences the outcome and reaction to the next and so forth. In TKES it is strongly recommended that all observations include commentary on all of the rated standards. The feedback is recorded in the Georgia Department of Education electronic platform within five business days. If an observation is not shared with the teacher within five school days, it will be invalid as a formative observation or walkthrough.

Evaluators conduct a Pre-Evaluation Conference, Mid-Year Conference and Summative Conference for all teachers evaluated by TKES. The evaluator is also responsible for providing timely feedback to the teacher on observations, whether recorded as a walkthrough, or as part of a formative assessment on the Formative Assessment through the electronic platform. Feedback and commentary from both types of observations are shared with the teacher via the electronic platform. A formal conference after each formative observation is optional. There are no guidelines for how feedback is given, just when it must be given (within five days of a walk through or observation).

Interpreting the survey data and the use of the data to contribute to final ratings for teachers is the last sub-theme for inter-rater reliability that was derived from coding. Another component of the Teacher Keys Effectiveness System consists of Surveys of Instructional

Practice. These surveys provide a means for collecting client (in this case student) perception data and will be administered through the electronic platform. Student Surveys of Instructional Practice are administered in grades 3-5, 6-8, and 9-12. Students in K-2 grades do not participate in the survey. Among the advantages of using this survey design are the rapid turnaround in data collection, the limited cost in gathering the data, and the ability to infer perceptions of a larger population from smaller groups of individuals. In the Teacher Keys Effectiveness System, student Surveys of Instructional Practice is used as a measure of teacher effectiveness and documentation in that they provide support for four of the standards. These four standards (Standard 3 - Instructional Strategies, Standard 4 - Differentiated Instruction, Standard 7 - Positive Learning Environment, and Standard 8 - Academically Challenging Environment) reflect the direct experience of students in classrooms. Data from student surveys must be used to inform the rating of a teacher's performance on these standards on the Formative Assessment (when data is available) and on the Summative Assessment. Commentary specifically related to the surveys of instructional practice must be included in the feedback for standards 3, 4, 7 and 8 in the summative assessment. If the rating on any of these four standards differs significantly from the rating indicated by the survey data, the evaluator is required to provide written justification to explain why the performance rating on the standard is not aligned with the survey data.

The second theme that came from the coding of the data was training. The sub-themes found within the coding included training on: the TKES platform, the Professional Learning Plan process, whether training should be conducted at the district or state level, and for survey administration. After highlighting the key terms, the researcher used a blue pen to note the theme of training and then labeled the sub-themes with a letter (as noted in the table below).

Training on the TKES platform for initial implementation during the current school year (2014-2015) was conducted in the Spring and Summer of 2014. All evaluators must be fully trained and credentialed by a state and/or district credentialed trainer in using the components of Teacher Keys Effectiveness System (TKES). All administrators who are responsible for evaluating teachers must be credentialed prior to using TKES. Credentialing is a process of establishing the qualifications and proficiency of evaluators to utilize TKES. The credentialing assessment is a minimal competency assessment that measures participants' understanding of the information and practice provided during training. It includes both recall of specific TKES information and practice of all the TKES processes. It is a bridge to ongoing learning as evaluators work with TKES. All evaluators MUST pass the credentialing assessment prior to using TKES. The expectation is that evaluators continue to familiarize themselves with the TKES process as they work within their districts. The credentialing process provides calibration and further increases the alignment of evaluation ratings. Currently, the only trainers providing full TKES training and leading the credentialing for administrators are members of the Teacher and Leader Effectiveness (TLE) Division of the Georgia Department of Education. As the instructional leader in the school, the principal serves as the model for appropriate evaluation practices, coordinates all evaluation activities within the school, and has ultimate responsibility for all evaluation activities within the school. Following the TKES training and credentialing, evaluators are encouraged to review classroom observation videos, observe lessons in classrooms and discuss ratings along with judgment of practice based on the TKES standards rubric. Specific navigation of the TKES platform, where the data is gathered, was not a large part of credentialing.

A second area where the research participants mentioned needing additional training was on implementing professional development plans. In TKES, these plans are referred to as Professional Learning Plans (PLP's) and one is created by the evaluator and approved by the principal within the electronic platform. It may be developed by the evaluator in collaboration with the teacher, coaches, mentors, or other qualified individuals. The PDP provides guidelines and timelines for specific, mandatory professional learning which supports immediate improvement of teacher practice and increased teacher effectiveness. The TKES handbook provides a summary of professional learning resources and teacher development opportunities located in the platform. The PLP may include any other enhancement opportunity with clear expectations about changes needed in performance to be demonstrated in the classroom and school. The PLP is an intensive effort toward improvement of teacher practice and effectiveness. A PLP may also be used when a teacher does not meet the professional duties, responsibilities and ethical expectations required by the teacher.

The TKES process was the third sub-theme that arose from the coding. There are six main steps in the annual TKES process. The first step, teacher orientation is when all teachers must receive an orientation regarding the requirements of TKES prior to the beginning of the evaluation cycle. Orientation materials and guides are provided by the Georgia Department of Education and should be used by the district and/or building principal to orient teachers. The orientation must take place prior to the first observation. Documentation of the orientation for each teacher must be maintained within the electronic platform. Teacher familiarization is the second step and it consists of ongoing professional learning utilizing provided materials on each of the ten performance standards. These activities may occur and/or be repeated at any time during the school year. Teachers who participate in familiarization activities earlier in the year

will have a clearer understanding of the ten performance standards and the expectations for classroom practice and performance. The electronic platform is used to record the teacher's participation in activities related to familiarization. The next step is Self-Assessment. In a full implementation year, all teachers complete a self-assessment on the ten TKES standards as soon as possible following orientation. Teachers complete the Self-Assessment electronic platform, and it is available to both the teacher and the school evaluator for review and professional learning planning. The Self-Assessment is used to guide discussion during the Pre-Evaluation Conference. The aggregated self-assessment data from a staff or group within the school can be used to make decisions regarding appropriate professional learning for individuals or groups. The next step in the process is teacher Walkthroughs. The purpose of these walkthroughs is to help establish the frequency and consistency of appropriate classroom practices as identified within TKES.

Data from the Walkthrough observations will be used to support and enhance performance ratings on the Formative Assessment and in the Summative Assessment. Formative Observations and Assessments are the next step. As evaluators conduct observations in a teacher's classroom, they continually build a portrait of that teacher's approach to and implementation of instructional practices. Although many practices and instructional strategies should be directly observed both in the walkthroughs and in the formative observation, other information can and should be considered in the ratings for the formative assessment (i.e. walkthroughs, lesson plans, student work samples and other forms of documentation). Each Formative Assessment is directly tied to a formative observation. Once the formative observation has occurred, evaluators consider the full scope of a teacher's practices that have been observed and documented up to and including the formative observation. Following the

formative observation, if sufficient evidence is not present to rate a teacher's performance on a given standard or standards, evaluators can request specific documentation relevant to the aligned standard(s). Following the first Formative Assessment, the cycle of observing practice and collecting documentation begins again for the second Formative Assessment cycle. After a second cycle, a Summative Assessment is done as the final step in the process. The Summative Assessment consists of a totality of evidence from the formative assessments and student instructional surveys.

The next two sub-themes that were derived from the coding was the different training provided from the district and the state to implement TKES. Specifically, the participants in the study noted the way they were supported through the state and local training on the above mentioned platform, process and PLP components. In addition, most of the participants noted needing more training on administering the Surveys of Instructional Practice and deciphering between not only who needed to take them, but also when they were to be administered and commented on in the electronic platform.

Experience was the final theme that came out of the data collected. The sub-themes that emerged from the coding were: years' experience as an evaluator and, specifically, principal, use of other evaluation instruments, TKES training already received, and giving feedback. Experience developed as a theme due to the nature of the interview questions and relation to the above mentioned themes. While answering the questions, each principal noted and spoke to the experiences that helped them implement this instrument. Through coding the responses, the researcher noted that a principal's experiences can be described as the over-arching theme of the entire study because it directly impacts the way training and inter-rated reliability are interpreted and implemented in his/her building.

The logic behind the themes emerged from the data and the coding led to the three main themes. While coding the interview questions, the research questions were kept in mind. The process began with the researcher highlighting all of the significant words pertaining to future principal preparation. After coding with letters, (A, B, C, etc.), the sub-themes and over-arching themes arose from the data. The researcher then used the data to underline in red, blue, and green to help organize the data into the three over-arching themes. Table 3 shows the color that relates to each theme, along with the codes that were used to derive the sub-themes (A, B, C, etc.).

Table 3: *Data Coding*

Theme	Color Used	Codes for Sub-Themes
Inter-Rater Reliability	Red	A – Standards B – Feedback C – Interpreting Surveys
Training	Blue	A – Platform B – Professional Learning Plans (PLPs) C – Process D – State/District E – Surveys
Experiences	Green	A – Years B – Use of Other Instruments C – Previous TKES Training D – Giving Feedback

Research Questions

This section focuses on the four research questions that addressed the focus of this study: future principal preparation needs to conduct effective evaluations. Each response is documents to show the congruencies and differences of the participants' perceptions regarding the research questions. Responses will be divided into two categories: over-arching theme and sub-theme. Direct quotations and excerpts from the interviews are explored. There were no themes

generated strictly from one level – all themes reached across the three different levels within the system and those principals interviewed.

Research Question 1: What are the benefits and challenges of using the TKES instrument?

The following questions were asked of the principals to answer this question:

1. What experiences prepared you to start using TKES this year?
2. What is the biggest challenge of using TKES?
3. How do you feel about the standards and rating them for the walkthroughs and observations?

Data collected for question one shows that the nine principals interviewed had several indicators that led to the overall arching themes of experiences and inter-rater reliability. The principals commented on the specific experiences that they have had leading up to this year's implementation of TKES. The components of the instrument that the participants commented on that they find are challenges included the standards and the number of times they are in a classroom. However, there was a prominent theme of inter-rater reliability that was mentioned, overwhelmingly, to need improvement. The biggest challenges were expressed by all nine principals as the time the instrument mandated and the lack of inter-rater reliability in giving ratings and using the platform.

What experiences prepared you to start using TKES this year?

The principals answered this question in a school-specific manner. Five of the nine principals, with at least one in each level (i.e. elementary, middle, and high), noted that practices, such as previously requiring lesson plans to be turned in or doing multiple walk-throughs, in

which they put in place prior to TKES being mandated helped for a smooth transition. Due to the fact that more than half of the principals responded in this manner, the data coding presented the theme of experiences playing a major role in preparing principals to implement TKES this school year. The quote below, from Principal Four, is representative of the group because the above mentioned five principals all talked about how requiring lesson plans and doing multiple walk-throughs prepared their teachers for some of the components of TKES and helped them manage the time that is required to complete such tasks. Although each principal spoke of his or her own school respectively, the theme of experience emerged from the data in the way each explained the specific routines that helped his or her building adjust to the new evaluation instrument.

> The walkthrough process that we've done at our school over the past four years and specific look-fors have helped us prepare for TKES. The other thing is that I have always required teachers to have written lesson plans. They are to be opened and readily available anytime we do our walks. Last year we collected data on how many times we went into a classroom and there were no lesson plans. So teachers were used to seeing us and meeting our expectations.

All of the nine principals interviewed noted that the training they went through with the Georgia Department of Education prepared them to implement TKES. The credentialing process, as defined earlier in this chapter, was similar for all of the participants in this study and they all mentioned it in their responses. The principals all attended a two-day workshop that was led by a representative from the Georgia department of Education. Cobb County has its own representative and she was the trainer for all evaluators.

Another experience that became evident through the data collected was how the role of instructional lead teacher benefited many of the participants as an experience that helped them implement the instrument. Of the nine participants, six of the principals explained how their involvement as the instructional leader in a building, helped them to interpret and analyze the

standards that are to be rated on each teacher. All three of the high school principals discussed how the area of content they consider themselves to be experts in was easier to monitor and evaluate than others, while one of the middle school and two of the elementary school principals discussed how being an Area Lead Teacher for the district gave them greater insight in all content areas. As the quote below, from Principal Two infers, the principals recognize that the more experience they have with the curriculum being taught and the manner in which it is taught, the more they can effectively evaluate the teachers in their building.

> Being an instructional lead teacher and having state mentor training have helped me to understand what the standards mean and what to look for. I am kind of old fashioned in that I learned to script and I have a hard time letting go of that. But I walk out of the classroom with a wealth of information that I can think about and figure out how each of the pieces I saw fit into the standards I am addressing.

Finally, all nine of the principals expressed the benefit of having used similar instruments in the past and made a connection with how semantics of evaluation is still the same, just with different requirements from the instrument. The school district has had three different evaluation instruments over the past ten years. The GTOI (Georgia Teacher Observation Instrument) was designed to evaluate teachers based on their years of experience (i.e. teachers with less than three years received multiple observations and teachers with more than three years received only one observation). The Cobb Keys was the most recent instrument prior to TKES and had eight standards that closely match those in TKES. Four of the principals focused on the language used in TKES compared to the language in the Georgia Keys. Moreover, there is an indication, as seen below, that all instruments can be implemented in a similar manner if particular language is used, which can be developed by principals and all evaluators after experiencing different instruments and the environments in which they are used.

> The instruments that were developed and used in CCSD in the past have been very similar. The way I give feedback and the terms I use are similar, it is a matter of semantics and how we are delivering it (number of times) is different with the new model. The way we evaluate is the same, though. In my four years as a principal and six as an AP these experiences and trainings on the other models have helped me implement TKES. Every evaluation experience is useful to learn the language.
>
> My most useful experiences came from the previous observation instrument since that is what TKES is – which requires knowing what the components of the instrument are and knowing what good instruction looks like and practice – and what areas teachers need help. So, the previous walkthroughs and observations helped me prepare to use TKES.

What is the biggest challenge of using TKES?

The overwhelming response to this question, meaning that all nine principals began their response with this topic, was time. The walkthroughs and observations take a great deal more time than instruments that were used in the past. Each teacher receives at least four walkthroughs that are between ten to fifteen minutes each. In addition, there are two observations that last between thirty to forty-five minutes each. During the walkthroughs and observations, evaluators must comment on the standards and address specific concerns. Two to three standards are evaluated during walkthroughs and all ten standards are evaluates *and* commented on during the two formative observations and the summative assessment. The amount of time the components of the instrument take has led to a feeling of "checking off" the requirements instead of focusing on the results and improving the quality of what they are actually evaluating. The frustration felt by the principals due to the time the instrument takes to complete is not grounded in the idea that they have more to do, but the realization that the mandates are creating a feeling of sterilization and institutionalization.

> The ability to be an instructional leader has been sterilized – and institutionalized. Because you can't go in all the time and give formative feedback – can't talk because I am busy typing.

Much like the evaluation instruments developed by Taylor and Cubberly during the scientific method era of teacher evaluation, this instrument lays out a set of principles for school administrators that emphasizes measurement and analysis of data to ensure that teachers and schools are productive. However, it is apparent that some of the participants in this study believe it is too scientific and not "human" enough – missing the interaction and relationship building due to the amount of documentation that is required. The principals spoke about how they love being in the classroom – but the instrument makes it constraining by having to fill the requirements and it is just too much right now without any flexibility.

> A frustration with the time is I am finding myself not in as many classrooms because I am focusing on my list that I evaluate (all principals agreed and said the same) – I feel like I am seeing less of what is going on in my building, but I am documenting more.

In addition to time, seven of the principals mentioned that the use of the platform was cumbersome for not only their teachers, but themselves as well. The TKES platform is all electronic and the initial training did not cover the multiple resources available on the platform. Optional trainings were offered, but, as one principal noted, "You don't know what you don't know" so many evaluators passed on attending the additional trainings. It was only after the cycle of walkthroughs and observations began that principals' and evaluators started discussing amongst themselves how many resources were available on the platform. One component of the platform that individual schools were left to figure out and navigate on their own was the student surveys. The surveys presented themselves as a challenge for two reasons, as noted in the quote from one principle below: 1) Validity of comments made by students who range in age from nine to eighteen and 2) Timing and scheduling electronic surveys that were to be given at random, but still with students familiar to the teacher(s) being surveyed.

> The student perception surveys were interesting – I am not exaggerating when I tell you it took me 16 hours to do the schedule – and we don't have enough kids to give a true

> rating – in addition, they are nine to twelve year olds giving their "perception" of a good teacher – it just turns into a popularity contest.

During the focus group, the impact of the surveys came up during the discussion, pointing toward an inconsistency in the messages received and making them challenges of TKES. The comments about the amount of time the instrument takes were similar to those given during the individual interviews, but the survey comments focused on when evaluators needed to comment on the results of the surveys. Some principals were told in training that they needed to comment on how a teacher was rated by the students on standards three, four, seven, and eight in the formative assessments during cycle two and on the summative, yearly assessment. On the other hand, some were trained that you only commented on them in the summative assessment. As one principal noted, "If you wait to do the surveys – you can play the system and wait to do them until you are done with the formative assessments and don't need to worry about the difference in trainings."

Another challenge one principal spoke directly toward was how his experiences in labeling a teacher "just proficient" has been, and continues to be, a major concern of his. Although multiple instruments have been used over the years, the end result of giving a teacher a rating based on a small amount that is actually seen remains troublesome for many evaluators. As noted below, it is through experience that principals begin to understand the complexity and vast realm of responsibilities our teachers are required to perform on a daily basis. It was noted that, as the components of an instrument change, some concerns remain the same.

> The platform, change of rating system, and number of visits is what differs with TKES. I had issues with the rating scales of the former ones as well (calling a teacher just proficient when they did so much more), so that is a continuous issue with not only TKES, but all evaluation instruments I have used.

A challenge that four principals, one at the middle school level and three at the high school level, reported was that the messages received from trainings were different for their administrative team and themselves. The state of Georgia requires state training for all evaluators, but there are often several different dates and training teams that facilitate the information. Even though Cobb County has their own "go-to" person, there are often different individuals to administer the actual training. So, as mentioned below, the training might be described the same way, but individual delivery and interpretation play a part in the message that is sent.

> I'd say yes, technically we were all trained in the same manner. However, because it was so intensive, we came back together to talk about our experience and compared notes and found that things were interpreted differently since they were subjective based upon your experience. We answered each other's questions and collaborated.

One principal made a point to distinguish the type of school she works in and the researcher took note that this spoke to one of the limitations of the study: different populations are at each of the nine schools where the principals evaluate. The reason why this might be a factor in the interpretation of the instrument is that some schools, quite frankly, deal with more day-to-day problems than others. In schools where discipline and other environmental factors are not a major factor, the requirements of the instrument might be easier to execute.

> I appreciate that TKES takes an in depth look at all ten of the standards. The time it takes is a lot and I feel grateful that I am in a school that is autonomous – the discipline is taken care of – so when it is time to walk we can do that. I don't think it has stressed our teachers out because we have worked really hard to give them the tools they need. We prepared our professional learning communities (last year) to two of the standards so the teachers are working on the wording and rubrics of standards.

A principal that is visible in a building knows what is going on in the classrooms and is able to evaluate happenings consistently. Three of the principals explained that a challenge to

them was documenting things that they already knew about their staff. There is a forced component of the instrument that requires evaluators to write comments and check off ratings after each walkthrough or observation. Therefore, in order to get a "green check" and be in compliance, an evaluator must make the comments and do the proper number of walkthroughs and observations. At times, these principals felt that they had to jump through hurdles to make sure the right places are "checked off".

> If you are an administrator that isn't very visible and doesn't know your staff, it is good. But if your leadership philosophy is one where you are visible and you know your staff, it is forcing you to document what you already knew. So, for me it is somewhat cumbersome sometimes – I am more frustrated with the hurdles you have to jump to make sure you are signed off properly. I know my staff and I don't need to massage them. It forces everyone to be more visible, though, which is a good thing.

In contradiction to the above, one principal referenced the amount of time the instrument takes to implement properly, but found benefits in it. She mentioned how the standards allow evaluators to be more specific about the performance of a teacher. Therefore, it seems that the instrument might help principals who were not as visible before become so while completing the steps of an annual evaluation cycle.

> The time involved is cumbersome, however TKES does a better job breaking out the standards and explaining what makes an effective teacher so from that aspect I think it is more helpful to find a teacher's strength. When you want to encourage a teacher leader or an area where they need to grow it is easier to address those – instead of just being "overall".

The way TKES breaks down the standards and the rubric that is used to determine a rating for each standard is cumbersome, but it is also thorough and allows for evaluators to fully describe what they have seen throughout the year as they walk through a classroom or observe a lesson. In the past, a teacher only had one "shining moment" to show an evaluator what he or she could do – now there is an opportunity to do so on many occasions in different settings. As the principal below noted, if used correctly, the challenge of time can be countered by the

opportunities available through the feedback and professional development found within the instrument and on the platform.

> I am one of the few that really love TKES; I appreciate how it is broken down and the rubric vs the yes/no or emergent/proficient; I like the bullet points under the standards; I like the process; I have only been in half of the classrooms this year, though (the teachers on my list to evaluate); I like the feedback and the opportunities it offers – it has a wealth of professional development opportunities.

How do you feel about the standards and rating them for the walkthroughs and observations?

The question above emerged after the first two interviews because the first two principals included the topic in their responses about conducting walkthroughs and observations. In addition, they alerted the researcher to the idea that several principals were not aware of the consequences that came with a lower ranking. The researcher added the question for the remaining principals and found that there was not much consistency in how the principals rated the standards.

Two of the principals responded about the way they rate the standards and their knowledge, or lack thereof, of what giving a teacher a "2" as a rating would mean for him/her. The issue was that through the course of the school year principals were told that if a teacher had two or more "2s" on TKES they needed to be turned into the PSC (Professional Standards Commission). These two principals noted that they would have liked to have known that going into the school year, not because it would have changed things, but to have knowledge of it and inform the teachers.

> The other thing I wish we would have known is about the "2" being reported to the PSC – the rollout of the information was not appropriate – I would have liked that knowledge at the beginning so I could have discussed what a 2, 3, 4 looks like before giving one so I could help them grow and come back the next day (or next week) to hopefully see growth – before marking them down and having to report that to the PSC eventually. We didn't know that until January!

The inter-rater reliability from building to building was a major theme that emerged from the data gathered from this question because it demonstrates how principals rate differently from building to building and how they use the standards (or don't use them) to make comments about performance that is seen. Administrative teams, from building to building, have different processes. Although it might be beneficial for their teachers, the question remains if it is fair when the ratings are compared from building to building. Below are examples of how some principals reported their administrative teams are rating the standards during the walkthroughs and the observations.

> During the first set of walkthroughs, all three admin looked for the same standards. This time, we know we are looking for 3,4,7,8. I go in and I have ten minutes – I look for information to fit under those four standards. I do the walkthroughs different than observations. Walkthroughs are very visual – what do I see, how prepared is the teacher. If I come out and feel like I saw more for two standards than the others then I write on those only. If necessary I comment on all four, but I don't often.
>
> For our walks we are focusing on instructional strategies and differentiation – so we make sure to rate them regardless of is we see it or not. We use the same standards throughout the school year.
>
> I rate the standards that I see. I am not going to rate a teacher down in an area just because I don't see it. I tell the teachers what standards I am looking for and if I see another one I rate it. If I see differentiation, I rate it.
>
> We tried to choose the standards that we thought we would always be able to see. For example, with the four walkthroughs we are definitely going to look at classroom environment and instructional strategies. However, if I walk into a room and it is during a transition or snack time, I leave and come back later.
>
> My administrative team knows what standards we are looking for and we always rate the teachers on those standards. If we see another standard while we are in the room that will help the teacher, we rate it. We only do that to help teachers, though.

During the focus group interview, one of the elementary school principals discussed how she had contacted the Georgia Department of Education for clarification on what to rate during an observation or walkthrough. She did so because she felt an issue that happens at the

elementary level is that if you choose safe standards, depending on the time of day, you may not see something and you have to mark them down. At the middle and high school levels, because the day is divided into content blocks, there is less concern for this. The state advised her not to rate a teacher low if you don't see something, but rather give them credit for what they are doing at the moment. During initial training it was clearly not articulated that way and, still, not all elementary principals and evaluators have been given this message.

Finally, two principals comments on a very important topic. TKES is an effectiveness measure – an instrument that was developed with the theory of helping teachers grow to become better educators. However, if an evaluator gives a teacher a "2" under the assumption that they are growing and learning, which is quite common, it will negatively impact his or her overall standing with the PSC. These principals stated that there needed to be a realization that a two should not be viewed as "poor", but the weight it brings in terms of turning a teacher over to the PSC or eventually impacting their pay is detrimental – causing many evaluators to not give a teacher a two.

> A two isn't needs improvement (missed opportunities – could have been done better), a one is not seen. And if you are supposed to help grow teachers, it is OK to see some two's, especially if you have some new teachers.

Research Question 2: What is the level of ease in navigating the instrument in comparison to those used in the past?

The following questions were asked of the principals to answer this question:

1. What other instruments have you used in the past?
2. How do the others compare to TKES?
3. How do you share the work load with your administrative team?

Data collected for question two shows that the nine principals interviewed had several indicators that led to the overall arching theme of experiences. The principals commented on the specific experiences that they have had leading up to this year's implementation of TKES. The importance of this question was that it helped the principals recall instruments they have used in the past and asked them to become aware of the similarities and differences and why they may find one more beneficial that another. In addition, the question about how the work load is shared amongst the administrative teams rose from the first interview. For the first two questions, there was not a difference between the levels in which a principal worked. However, the last question saw a theme among high school principals that was not there in the elementary and middle school principals.

What other instruments have you used in the past?

There is not a perfect evaluation instrument, but different aspects of previously used ones, compared to the requirements of TKES, was the reason behind this question. In referencing the others used, these participants were able to highlight the benefits and challenges of TKES. All nine principals have used at least one other form of evaluation instrument. Some have used up to three different instruments. The most recent evaluation instruments used were the Georgia Teacher Observation Instrument (GTOI) and the Cobb County KEYS. The GTOI was the classroom evaluation instrument used as part of the Georgia Teacher Evaluation Program (GTEP) in Georgia classrooms before 2009. It included specific teacher behaviors categorized under the rubrics of instruction and management. The Cobb KEYS was developed from the Georgia CLASS Keys in 2009 and used until 2014. The purpose was for improvement and accountability and was a formative and summative tool that identified a teacher's level of performance across five strands of teacher quality. The five strands were: Curriculum and

Planning, Standards Based Instruction, Assessment of Student Learning, Professionalism, and Student Achievement. The five strands were defined into performance standards and elements with rubrics.

How do they compare to TKES?

Comparing TKES to the other instruments mentioned above was embedded in several of the questions by the way the principals mentioned what they used to do and what they are asked to do now. In addition to the question of rating the standards, it was mentioned often in the remarks about giving feedback, which will be addressed later in this section. All of the principals mentioned how the instrument compares to others used in the past and, again, the experiences that the principals have seem to lead to that being the over-arching theme. In addition, the future training that the principals would like to receive came up in the way they referenced former instruments that they used and what they would like to see come from TKES in the future.

One of the principals explained how TKES being a growth model is helpful because you can tell teachers, "This is what I have seen multiple times," and they can grow from it – or stay the same. Whereas, by comparison, the other instruments were more along the lines of "Here is the observation and I'll talk to you at the end of the year with your summative assessment." Another principal said that she spent so much time in the classroom anyway, there was never a question from a teacher on what they were doing or a question about what she was seeing. However, the difference with TKES is now she is spending more time observing (ten minutes instead of six to seven).

The use of the platform was something that all of the principals explained as a difference between TKES and instruments used in the past. As one principal said, "I don't fault the TKES

platform at all – for all of us it is a learning process. The biggest problem is ranking all 10 standards and gaining enough information that might not be obvious within the time frame of an observation. The navigation of the platform is what has been tricky for me." TKES is electronic and all components are uploaded to a state platform, which is different than those used in the past.

Evaluation is part of the job and principals understand this. A high school principal summed it up rather nicely by stating, "All of the instruments we have used are the same, they are just a different instrument to record it on. TKES is more intensive due to the time constraints and it seemed, at first, more lock stepped. As we have found different ways to do it, there has been some flexibility, but it is just a lot more time involved – even though we always had done two observations on everyone in the past. TKES is a little more intensive with adding the mid-year assessment and the technical aspects of having to meet, check off categories and follow up to make sure the teachers have those "green checks", but in terms of *what* we are evaluating, it is the same as those used in the past."

The focus group interview compared TKES to other instruments in terms of the training that might need to come to help TKES improve the evaluation process like some of the others the principals noted to have used. The principals agreed that they need more training on the platform and the process. For example, some administrators did the walkthroughs and formative assessments in a different manner/timeline than others. Again, consistency among administrative teams, schools, and the district was noticed to be absent from the way TKES has been implemented and future training could address this.

How do you share the work load with your administrative team?

There was a distinct difference between how the elementary and middle school principals answered this question compared to the high school principals. High school administrative teams are made up of eight to ten administrators, while elementary and middle school teams are made up of only two or three. Therefore, all of the middle and elementary principals split their teachers evenly amongst their administrative teams and high school principals only saw, typically, one group or department.

One principal talked about how he took on more of the workload at the beginning of the year in an effort to make sure there was inter-rater reliability. He mentioned that his administrative team conducted the first five walkthroughs with his team and then compared ratings. They compared what they gave and why and found they were actually identical. During the observations the team went together for the first five as well. After conducting the first walkthroughs and observations together, the workload was divided evenly.

In high school, the teachers are divided by content area, not grade level. Most content areas have close to the same number of teachers to evaluate. Although an evaluator might not be an expert in the content area, he or she is still the expected expert in knowing good teaching. A high school principal recognized this by stating:

> High school is a little different in the sense that elementary and middle deal with grade level and high school deals with content areas. So, when I take a content area, it is usually something you didn't teach. So you have to explain that, even though you didn't teach it, you can still evaluate good teaching. Other admin (principals) could share with me how to talk to teachers about how you are evaluating them on something you didn't teach. You recognize what feels right/looks right. Conversations with the students and looking for engagement are huge. Content specific – I might not know – but I know good teaching. You feel a good teacher – you can see it in the students and the classroom (and results).

Another high school principal explained why he evaluates all of the new teachers and, even though they might need more time and attention, there are often fewer new teachers, so the

workload evens out. In addition to the explanation of why he evaluates the new teachers, he confirmed what the above principal stated in terms of "good teaching is good teaching":

> I observe and evaluate new teachers – I feel that as a principal I need to be hands on with the people we are bringing into our teaching culture. I can do observations on everything – good teaching is good teaching. Some people like one content, but I tend to focus on the goals we have (math this time). I should be comfortable with all content – that is my job. Strong instructional strategies are good no matter the subject area and engagement is engagement.

Research Question 3: How is effective feedback given while using the evaluation instrument?

The following questions were asked of the principals to answer this question:

1. How do you differentiate the way you give feedback?
2. Did you give feedback verbally or just on the platform? Explain.
3. How has your experience or graduate work helped you in giving feedback?

Question three data indicated that all nine principals gave feedback in a different manner. The over-arching themes included experiences (what has worked and what has not worked), training (in terms of what they were told to do, what they thought they were mandated to do, how their administrative teams differentiated the way they gave feedback), and inter-rater reliability (explanation of how one school handled the feedback compared to others). All three levels answered with different trends and there were no consistent themes found within one level that differed from another level.

How do you differentiate the way you give feedback?

The term differentiate refers to how one changes the presentation, content, environment, or product of an assignment or activity. In the case of evaluation, it is important to differentiate on the basis of teacher understanding and level of competence. When completing the

evaluations, there are several areas where principals can differentiate in the way they give feedback. The feedback is not mandated to be in person. Feedback can be given in oral form if an evaluator wishes, but the only conference that *has* to be in person is the summative assessment at the end of the year. During the other formative assessments, feedback can be written and loaded on the platform or given in a small or large group setting. As the interviews took place, it was clear that all principals gave feedback in a different manner.

During the walkthroughs, four of the principals looked for the same three standards and when they gave feedback they gave sentence-form feedback such as: "We saw this…" "This was a good example of…" Other principals stated that they changed how they gave feedback as the year progressed. For the first set of walkthroughs and observations they wrote everything down and tried to write something for every standard, so each teacher got feedback on the specific standards that they saw. The comments were unique for each teacher and if there were questions about the written comments on the platform the evaluators had follow up conversations with the teacher. There were also comments made in person that were not loaded on the platform.

One middle school principal, who has been at her school for over twenty years as a teacher, assistant principal, and now principal, recognized the individuality of teachers and the need to keep the happenings of their classrooms and profession protected. The personal relationships that she has with many of the teachers in her building run deep and strong, which might be why she spoke to the fact that an electronic platform monitored by the state is a little scary. There is no certainty of who sees the information that evaluators enter about their teachers.

> Teachers' classrooms are very personal. Giving different feedback comes naturally to me. I comment on the platform but I am very careful knowing that is "out there" and if

> they move out of this district, the information will be there. I am honest, but if it is needed to have a 1:1, I ask them to come and talk to me in person. An example is a teacher who has a very intense personality. I wanted to let them know that their intensity might come across wrong and they should be mindful – knowing that intensity is not always "joy" in a student's eyes. I have been here for a lot of years and I have raised some of these teachers from baby teachers. I appreciated the opportunity of being an AP and then principal and it brings a level of understanding where my teachers are and how they function. It allows me to concentrate on other things – not who they are.

Walkthroughs and formative assessments have to be shared with the teacher and finalized within five days. The only face-to-face follow ups that are mandated are the summative assessments. The others just require the teachers to look at them and accept. As one participant said, "No verbal pat-on-the -back even made a three seem not sufficient and the scores were not delivered as well. Therefore, we are changing that next semester and having a short conversation that allows us to give positive feedback and show appreciation for what they do without just coldly pushing the share button." Many teachers need personal interaction to feel accredited, appreciated and accomplished. This system eliminates that essential human need.

The main reason principals and other evaluators are by-passing the face-to-face conferences is time. One high school principal said, "Having one-hundred and ten teachers and needing to have face to face meetings is hard. I am going to pick and choose who I have the face and face conversations with. I know the teachers well enough to make some of those judgment calls." Although this might be a reasonable solution, it is hard to determine if a person really needs individual feedback or not, no matter how well someone believes they know him or her. As a middle school principal stated:

> We have met verbally because you need to meet with teachers to give them that feedback, even if it is a quick one – even though TKES does not require it. To me, if you are trying to build strong instructional leaders and teachers, you need to have conversations face to face – especially if there are issues. If there are issues you need to address it in person to see if the teachers have anything to add to the conversation. The data was shared individually.

One elementary principal agreed and mentioned leaving personal notes for teacher who did a great job. She also met with teams to deliver over-all data and trends as long as all teachers received three's or four's.

Did you give feedback verbally or just on the platform? Explain.

After listening to the answers to the first question of how the feedback was given, the second question was more direct and specifically asked if the feedback was verbal or just on the platform. In addition, this was the format of the question on feedback for the focus group. The question was similar to the first, but was needed to clarify and understand what the environment was when the principals met with their teachers to deliver their ratings. Three of the principals explained that they met with their entire staff as a whole group to go over data trends. Again, noting that with over eighty teachers, it was impossible to meet face-t-face. However, if a teacher had a two or one, they met with them one-on-one to discuss an improvement plan.

The focus group responded to the question of how they gave feedback in a similar manner as the individual interviews. As was evident in the first interviews, there were many differences in the way principals reported they gave feedback. The coding of the data pointed toward a theme for future training because a few of the principals mentioned where they felt the issues lie and how the topic of feedback could be improved in the future. Future training should, as all of the principals agreed, be consistent. One of the principals had the understanding that if a rating is anything other than a three you had to give a comment, but found out that other principals at his level were not leaving comments for fours. Another participant stated, "I thought if you give a two or four you have to comment – but then I asked the state department and they said, 'No, just give a rating.'"

A high school principal explained that he thought the conflict lies within the instrument and future training on giving feedback should be focused on what is acceptable from the state and district and what is appropriate to help teachers grow professionally. Again, needing consistency within the district was a strong feeling shared by all participants.

How has your experience or graduate work helped you in giving feedback?

The above question was primarily answered in terms of developing relationships in the building and being able to balance the requirements of the job while implementing this instrument. Six of the participants noted that on-the-job experience prepared them to give feedback more so than graduate work with one stating, "I think having more experiences goes further than a degree on a wall." However, one elementary principal reflected on the fact that graduate school helped her mature and talk to adults in a more appropriate manner." A high school principal agreed and determined that his graduate work helped him develop his confidence, but not necessarily the way he chooses to give feedback. He stated:

> If you don't know how to do that you are in the wrong profession. We are paid to have those difficult conversations. I tell my admin if you can't have those conversations professionally, then you are sitting in the wrong seat. You have to have the best people in your building. If they are not you need to counsel them into another profession or at least out of your building. You don't have time to put everyone on a PLP – you need to have those strong conversations. I think graduate school helped develop my confidence, but not necessarily the language needed for the specific directives – that comes with experience.

Research Question 4: What future training needs do principals report they need to use the instrument and give effective evaluations?

1. What professional development would you like to see principals get to help with using the instrument?
2. Should future professional development be provided by the district or the state? Why? Should it be by level or area?

Question four was the meaning behind this needs assessment. After collecting information and data on the background knowledge and use of the instrument over the past year, the researcher felt that a simple, straight forward question should conclude the study. Asking directly about the type of professional development principals felt they needed was intended to lead to data that could help plan future principal development sessions and positively impact teacher evaluation instruments and implementation. It was clear from the beginning of the data collection period that the principals felt strongly about areas of weakness or ones in which they would like to see changes or improvements. However, the following questions allowed them to discuss their ideas in better detail. The overwhelming theme for future professional development was inter-rater reliability. All three levels reported the same trends in answering these questions.

What professional development would you like to see principals get to help with using the instrument?

As the nine principals answered this question, the sub-themes for professional development emerged as needs in training on the platform and the process, implementing professional learning plans (PLPs), and scheduling, conducting, and interpreting the student surveys. Three of the teachers requested statements termed "Glows and Grows" from the district that would help them have difficult conversations without tearing the teachers down while continuing to foster better instruction. During the TKES state training, the participants were shown what a teacher deserving a four would look like, but it was based on a teacher in a district very different that Cobb. The principals all agreed that the example would more than likely be a three in their district. That realization showcased that there is subjectivity to TKES – which might be improved upon if the evaluators were given the opportunity to look at the rubric together and flesh out what a three or four teacher looks like in Cobb.

The principals all mentioned needing consistency among their administrative teams as well. They stated that the same terminology and rating guidelines should be used regardless of which administrator is doing the evaluation. Completing a few walkthroughs and observations together is one way that a principal stated he addressed this during the first year of implementation and the others in this research agreed that it should be a district requirement. As one middle school principal stated:

> I think just more collaboration between myself and other administrators to see what they are doing at their schools and how they are doing things is what I need. I think I am seeing the process correctly. However, some of my administrative team was trained differently than I was and there was a difference in how we interpreted several things. Completing some of the requirements together and in other buildings could help eliminate some of the discrepancies.

The process that this group of principals went through for this study is an example of what the principals noted they needed in the future to become more effective evaluators. The ability to sit together and talk about their interpretations of the information gathered, procedures used during the TKES process, and implementation of the components of the instrument was beneficial during the focus group interview and would be a productive and effective way to plan for future professional development.

One way that a middle school principal is addressing future preparation for herself is by planning staff development on the instrument for her entire staff. While she claims to always try and find ways to do an observation and be able to give feedback that teachers are able to take and grow from in order to be the best they can be, she recognized that teachers need to see what that "four" teacher looks like in her eyes. She felt that the district should prepare the evaluators to show their teachers what that looks like in their building, throughout the district, and in the state. The evaluators could then share with one another to note and comment on consistency. The

ideal, as mentioned by one principal would be, “Done with the admin team so we are consistent and done in the district so there are the same expectations.”

A different area of future training and preparation requested by the principals was to help them quantify and qualify what each standard equals so there is a better connection for each standard and how they can be measured. This will help evaluators to know and understand what they are looking for and what is appropriate. In doing so, evaluators can then train their staff. The principals that mentioned this were from middle school and elementary school and they referenced being able to set expectations at the beginning of the school year so teachers could expect a certain rating based on the performance they gave throughout the year.

During the focus group interview, the researcher spent a great deal of time on this question. As the principals started dialogue with one another the overwhelming response was that there should be differentiated professional development for veteran and new principals and it should be hands-on. Being able to walk one another’s buildings and have conversations about what a rating should be seemed to be the most agreed-upon need for this group of principals. All of the principals felt that the training needs to be improved on or they need more training on the platform in order to navigate it better. The principals also asked for more training on the process and less on the theory of “how to be a principal”. For example, during the credentialing training, the state spent six hours talking about what good principals do every day. However, they did not mention the word *dashboard*, which is where all of the performance indicators and professional development is located.

One principal noted that there needed to be consistency in the training because she and her administrative team went to two different trainings and they had two different messages – from the same trainer. Or, if they can’t make the training the same, offer different options for

those who have already completed initial credentialing. Having gone through it for a year, the principal said that she didn't need to be re-trained, but it would be great for principals that have been doing it for a year to have training on the platform and maybe PLPs. Other principals also spoke to the fact that there was not any training on when teachers should be placed on a PLP and they desire future training in that area because they were not aware, until well into the school year, that if a teacher gets a two overall they would be reported to the state and should be on a PLP.

The principals also mentioned needing more professional development in the area of the student surveys. They mentioned the surveys in terms of their experiences and how they could make them better, but also where they could benefit from future trainings. On elementary principal said that the training should be for the Assistant Principals who are in charge of the surveys. She stated, "The AP's need more staff development on how to set the surveys up and to know how much logistic planning it was going to take to get the kids their access keys.

Should future professional development be provided by the district or the state? Why? Should it be by level or area?

Most of the participants in the study felt that the district could serve their needs better than the state. It has been their experience, especially through this process, that the district does a better job developing their people. Although they were not knocking the state, they recognized that Cobb knows their people and they do a great job identifying, differentiating, and addressing areas of need. All of the principals also felt that the training needed to be done by level: elementary, middle, and high. They realize that the needs differ by level and what you are rating looks different as well. Along with separating by level, all but two of the participants felt that the training should be divided by area. There are six areas in Cobb County, all monitored by an

assistant superintendent. They felt the training should be differentiated this way because teachers in different areas are dealing with students and staff that have different struggles so the professional development could be geared toward what those principals need to know and understand to help their own teachers.

During the focus group, one principal mentioned that she thought the training should be done by the state because then it would be coming from one source and since it is a state instrument and they are evaluating from it, it would be helpful to know what they see as those being appropriate.

Residual Study Data

There was some data collected in this study that was not used. Although there was not much data not used, the information collected that did not fit in the themes that emerged were those of personal feelings toward teacher evaluation. In asking questions about TKES, the principals spoke, at times, in terms of their over-all feelings toward teacher evaluation. The purpose of this study was not to determine if teacher evaluation should be completed or not, but rather if the principals were prepared to conduct effective evaluations. The nature of the interviews led to open communication and the researcher did not limit the amount of information that each participant wanted to share. Instead, during the interview, all data was transcribed and considered valuable until the coding took place.

Another area where residual data was collected was during the focus group interview when the principals started to talk about the Leader Keys Effectiveness System (LKES). LKES is similar to TKES, but it used to evaluate administrators and district leaders. This study did not include a description of LKES, nor was it analyzing how principals felt about the instrument, so that data was not included. It was hard for the principals to not relate TKES to the instrument

that will lead to their own evaluation, so the researcher tried to allow some discussion before directing the conversation back to the questions being asked.

At times the principals gave examples of specific teachers and how they were rated. In listening and reviewing the transcriptions, these examples were not relevant to the themes that emerged because they spoke more to the teacher and not the evaluator. Since this study was focused on the preparedness of the principals to give effective evaluations, the researcher felt that descriptions of specific teachers were not relevant. A teacher's ability to do their job is absolutely part of reflecting on a teacher evaluation instrument, but in relation to this study, the answer would be more appropriate if directed toward how the ratings were reached in relation to how the performance was documented.

Finally, there were many comments made in relation to the administrative teams that the principals worked with and how their competence or lack thereof impacted the overall environment of the school in which the principal was answering questions about. Although the inter-rater reliability was important to document, this study was not researching the ability of the entire team to evaluate teachers, but rather solely the principal. For example, one principal mentioned that one of her administrators could not type very fast and it made the evaluations last a lot longer and take more time. Although this was something that made the focus group think quite a bit, it was not related to her personal preparedness to conduct an effective evaluation. When comments were made about how the principal led his or her administrative team toward improvement in this area, those data descriptions were included.

Summary of Findings

As noted at the outset of this analysis, the purpose of this qualitative, program evaluation study was to determine how principals perceive their level of preparedness to effectively evaluate teachers. Specifically, the study examined how the training provided for the use of the

Georgia Teacher Keys Effectiveness System (TKES) will lead to teacher growth and, ultimately, student achievement. The study conducted a needs assessment that addressed further training needs of principals in effectively using the evaluation instrument so future professional development can be planned.

Data analysis revealed three consensus themes: inter-rater reliability, training, and experience, which were discussed in this chapter. The three sub-themes for inter-rater reliability were: standards, feedback, and interpreting surveys. The four sub-themes for training were: platform, professional learning plans (PLPs), process, state/district, and surveys. The four sub-themes for experiences were: years, use of other instruments, previous TKES training, and giving feedback.

Following a discussion of the three themes, excerpts from participant interviews were used to illustrate their attitudes and perceptions of implementing the instrument and future preparation needs as described in their own voice and through the researcher's analysis. The analysis indicates the ways in which the data collected underlie the organization and content for future principal preparation in implementing a teacher evaluation instrument.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

The goal and purpose of chapter five is to present a summary and discussion of the research findings as well as recommendations for further study. The chapter will speculate on the meaning of the data reported in chapter four, identify and signify the study's findings in the context of the literature reviewed in chapter two, describe how the findings of the study challenge, contradict, and extend the theoretical and conceptual knowledge of teacher evaluation, and state the significance of the study to the field of education.

Chapter five discusses whether the research design and methods used to answer the questions of the study was adequate. In recognizing what was and was not successful with the study, the researcher will explain how the research design might be re-organized or changed if it were to be completed again. The chapter is organized into five distinct, equally important sections. After this introduction, there is a recapitulation of the study to this point, which will provide a brief summary to the readers. Next, the researcher's interpretation, meaning, and significance of the study findings is presented in relation to what the findings mean in the current context of knowledge on the topic, in the context of the literature on the topic and of the current theories about the topic. Following is a meta-reflection on the research design and methods of the study citing the validity and usefulness of the specific research design and method used to conduct the study and then there is a section on the practical and policy implications of the study to address what the research will do to inform practitioners and policy-makers in the education field. The concluding sections include possible next steps in research on the topic, including what line of inquiry might come from the research and what possible directions might be taken

and a statement from the researcher on what she has learned about herself and her profession while doing this study.

Recapitulation of the Study

The purpose of this qualitative study was to determine the perceived level of self-preparedness of building level principals to conduct effective teacher evaluations and to identify the future professional development needs while using the Georgia Teacher Keys Effectiveness System.

In an effort to put the subject of teacher evaluation into context, a review of the literature was conducted. The section examined the philosophical background of teacher evaluation and the purpose of teacher evaluation by navigating through information from over a hundred years ago and reviewed writers and researchers such as: Taylor, Cubberly, Dewey, Hunter, Stronge, Danielson, Whitehead, and Marshall. It also reported the federal and state education reform acts starting with A Nation at Risk and reviewed the different models of evaluation and multiple components of teacher evaluation. Finally, there was specific information included on Georgia's Teacher Keys Effectiveness System and included a definition and example of use for a needs assessment.

To generate data that described the future training needs for principal preparation, a qualitative study based on participant interviews was completed. The study was conducted with nine different principals in a school district in Georgia that was in their first year of TKES implementation. Research questions provided the direction of the interview questions and the focus of the interview questions in this case study was to develop a Needs Assessment to answer the following:

1. What are the benefits and challenges of using the TKES instrument?

2. What is the level of ease in navigating the instrument in comparison to those used in the past?
3. How is effective feedback given while using the evaluation instrument?
4. What future training needs do principals report they need to use the instrument and give effective evaluations?

The results taken from each question contributed to the educational knowledge base by exploring the preparedness of evaluators in conducting teacher evaluations and their future training needs. Teacher evaluation is a reflection of human performance in organizational settings (Stronge, 1991) and is a central component to accountability in education. "Most authors identify the fundamental purposes of teacher evaluation as improving performance and documenting accountability," (Tucker & Stronge, 2005, p.6). Reforming the evaluation system through enhancing the skills building principals need to be effective evaluators of teachers is crucial if we are to meet the current performance goals.

Interpretation, Meaning and Significance of Study Findings

A focus of interpretation for this research that is directly related to why principal preparation is essential for the implementation of evaluation instruments is the theory of education accountability – a phenomenon derived from the decline in education due to specific areas in need of improvement. Teacher evaluation is one area in need of improvement and is only as effective as the evaluators using them.

Over the past decades the federal and state governments have made various attempts at educational reform by taking an advisory role, but it has had little impact. Education reform has followed changes in the purpose of teacher evaluation by adjusting the systems put in place and the policies that guide them. Reforming the evaluation system through enhancing the skills

building principals need to be effective evaluators of teachers is crucial if we are to meet the current performance goals. "Most authors identify the fundamental purposes of teacher evaluation as improving performance and documenting accountability," (Tucker & Stronge, 2005, p. 6). Accountability is the focus of many political agendas and social reform policies today and new evaluation tools, such as TKES, have been used to hold teachers, administrators, schools, and districts accountable.

Accountability is at the core of education policy and political debate. It is the driving force behind changes in education and teacher evaluation is a major topic in dictating how teachers are rated in accordance to how they perform the tasks in which they are accountable. In 1983 the National Commission on Excellence in Education (NCEE) reported in A Nation At Risk that educational institutions were grossly underachieving. The report emphasized the low skills as evidenced through test scores in American youth and the findings showed that the teaching profession needed to be improved and, thus, a new system of evaluation to show improvement was needed. The report concluded that the Nation was falling behind globally, and that our country was not as competitive as it once was (NCEE, 1983). The decline in education had a negative effect on society both socially and economically. The findings of this research, though more than forty years later, found that principal preparation to conduct teacher evaluation may still be contributing to the effect teachers have on education in terms of whether the instrument used to evaluate their performance is having an impact on their teaching ability.

In late 1999, Georgia introduced the evaluation of student performance on statewide assessments into the teacher effectiveness debate (Eady & Zepeda, 2007). State lawmakers introduced this concept with the passage of the A Plus Education Reform Act of 2000. This legislation required administrators to consider student achievement to a minimal degree in the

annual appraisal of teachers. With the passage of the Act, teacher expectations changed and incorporated "The role of the teaching in meeting the school's student achievement goals, including the academic gains of the students assigned to the teacher" (OCGA 20-2-210, 2000).

The debate about how to identify quality teachers continues today, but it was impacted greatly when the George W. Bush administration passed legislation known as the No Child Left Behind Act (NCLB) (2001). The Act placed the federal government in a more prominent role in American Education and increased accountability among schools and teachers. Accountability, in more recent years, has been directed toward changes made in teacher evaluation. In 2009 the Obama administration provided financial incentives to encourage educational growth. Through the American Recovery and Reinvestment Act (ARRA), over four billion dollars in competitive grants were invested to improve education at the state and local level. The "reward" focused on reform in four areas: Rigorous standards and assessments that would prepare students for college and the workforce; data systems that effectively measure student growth and success while informing principals and teachers how they can improve instruction; recruit, develop, reward, and retain the best teachers and principals; and turn around low-achieving schools. Race To The Top (RTTT) grants were awarded to states that tied teacher evaluation ratings to students' performance on state assessments. State and local districts developed instruments to align with RTTT guidelines and used criteria and categories, rating scales, response modes, comment sections, and formats based on concepts from the literature on effective teaching practice. Evaluation was not defined under RTTT, but instead describes the components of evaluation and reference what it can be used for (U.S. Department of Education [USDOE], 2009). Therefore, state legislatures passed bills requiring school districts to use a multi-level tier rating system instead of the old satisfactory/unsatisfactory systems.

As a result of RTTT changes, teacher evaluation has transitioned from focusing on teaching to focusing on student learning. Johnson (1997) wrote "By providing the public, administrators, and teachers with data on teachers' skills and performance, a teacher-evaluation system will increase the instructional productivity of teachers, enhance student learning, and ultimately improve the quality of the educational system" (p. 70). This change requires building principals to have a new set of skills that will lead to teacher self-reflection, growth and improved student achievement. As part of Georgia's Race To The Top grant guidelines, the school district studied in this research strived to have clear and precise indicators and resources to guide teachers and leaders to improved performance, but the results of this study point to future needs to accomplish the intent of this policy. As the district looks toward continued principal preparation, the requirement to use TKES should include further training and development of evaluator skills to improve their performance.

Evaluations, as defined by Stronge (1991), are a reflection of human performance in organizational settings. Guidelines included under RTTT were designed to reform education. In RTTT, evaluation is not defined, only the components of evaluation and what it can be used for are addressed (U.S. Department of Education, 2009). The goal of using evaluation as a reflection of one's performance remains the same, but the objectives have changed over time. As principals continue to evaluate teachers, the results should be a central component of instructional improvement by being a rigorous performance measurement that provides useful feedback (Shakman et al., 2012). In order to do so, principals need the required skills. Principal preparedness impacts accountability through their job performance. This study uses the lens of principals to determine which resources and ongoing training they report they need to become effective evaluators of teachers. Krein (1990) wrote, "I propose that the problem is not

an inability to determine what the components of a good program should be, but the failure to implement them properly," (p.1). As the inevitable change in teacher evaluation continues, principals need to clearly understand the way in which it is packaged, delivered and explained to teachers. It is a professional imperative that they self-educate, self-motivate, and self-advocate in order to gain the knowledge needed in terms of what is expected in helping their teachers through the feedback they give.

The principals in this study reported that their experiences helped them prepare to implement TKES, noting that practices and specific routines they put in place prior to TKES being mandated helped for a smooth transition and management of their time. The experiences of evaluators and the way in which they understand the theory behind evaluations is a focus of interpretation for this study. The principals in this study recognize why teachers need to be evaluated and found that the more experience they have with the curriculum being taught and the manner in which it is taught, the more they can effectively evaluate the teachers in their building. In addition, the experience of having used similar instruments in the past and making connection with how the semantics of evaluation is still the same, just with different requirements from the instrument, was found through the data collection in this study. There was an indication that all instruments can be implemented in a similar manner if particular language is used, which can be developed by principals and all evaluators after experiencing different instruments and the environments in which they are used.

Over the past thirty years, research has consistently concluded that effective evaluations should include student data, multiple measures, and evaluator training. In this study, the researcher specifically focused on the evaluator's ability to reflect on how the training prepared him or her effectively. As TKES has been implemented, principals need to decipher whether

they are equipped in their knowledge to discern effective teaching and give meaningful feedback. The participants in this study responded in a way that concluded they need additional training in the area of inter-rater reliability. Due to the fact that the credentialing for TKES was considered a minimal competency assessment that measured participants' understanding of the information and practice provided during training, it was only a bridge to ongoing learning as evaluators worked with TKES. However, there was an expectation that evaluators would continue to familiarize themselves with the TKES process, but no monitoring to determine if they had was implemented.

Over the past twenty-seven years, Jon Saphier has looked at data on how administrators have evaluated teachers. He found that, "The incoming skills of administrators tend to be quite low. Administrators have little or no preparation for analyzing instruction when they are hired, and almost never had to show they could do so to get their job" (2008, p. 2). The research in this study examined the principals' perceptions of their own readiness to conduct effective evaluations on their teachers and the appropriateness of the professional development provided by the local districts to enhance their skills. Investigating the participants' perceptions of their readiness to conduct effective teacher evaluation and what they feel they need to develop their skills on individual components of TKES is another focus of interpretation of this study.

Another focus of interpretation for this research is the principal preparation in terms of the components of teacher evaluation. Although the details vary, there are several components of teacher evaluation that are included in all teacher evaluation models. "A good teacher evaluation system must define the teaching task and provide a mechanism for judging the teacher (Wise et al., 1985, p. 65). There must be a clear understanding of what good teaching is, what the evidence of good teaching is, and how it will be measured. TKES incorporates multiple

measures of effectiveness to determine a composite score that is then used to differentiate teacher effectiveness using four rating categories (USDOE, 2009). Although there are other potential uses of the teacher evaluation process, such as personnel decisions, the teacher evaluation process is most often cited as a way to enable schools to assist teachers in the improvement of their skills by assessing teachers' strengths and weaknesses to determine areas needed for staff development and professional growth. The expectation of teacher evaluation is that it will recognize good teaching and enhance the quality of teaching and lead to more effective student learning and achievement.

The data collected in this study addressed the principal preparedness in multiple components of TKES. There are both formative and summative evaluations conducted through TKES. Popham (1985) stated that teacher evaluation has two purposes: The improvement of teachers' skills so that they can become more effective (formative evaluation) and evaluation focused on changes in status such as granting tenure, placement on probation, dismissal actions (summative evaluation). Weber (1987) maintained that it is a mistake to think that one purely formative or summative system can serve the purpose of growth, accountability, school improvement, and personnel decisions and most school districts today are moving toward changing their evaluation instruments to be both formative and summative. The principals in this study concurred that both types of evaluation are beneficial, but they have challenges as well. The time formative evaluation takes, as reported by all principals, is the biggest obstacle.

Observations and walkthroughs are combined to create two formative assessments, which are then used for the summative assessment. Administrator Carolyn Downey developed the walkthrough to be more visible in the classroom and get a big picture of the learning environment. In addition, they allowed her to show her staff she cared about them and she

"came to realize that walkthroughs, along with meaningful dialogue, was an effective approach to help the teacher grow professionally," (Downey, Steffy, English, Frase & Poston, 2004, p. 10). The Downey walkthrough model was created from the information and professional development she did. "The idea behind walkthroughs is that firsthand observations can paint a picture to inform improvement efforts," (David, 2008, p. 81). Research on walkthroughs is limited, but available studies reveal "Wide variation in their usefulness and effects" (David, 2008, p. 82). Frase and Hetzel (1990) published School Management by Wandering Around and followed business models to compare the benefits of having the principal wander throughout the school. Skretta (2007) defines as effective walkthrough as feedback that gives teachers "relevant, real time data on their instruction" (p. 18). Walkthroughs can be crucial to improvement, however "Only when districts make their purpose clear and carry them out in a climate of trust" (David, 2008, p. 89). The principals in this study felt that the confinement of needing to document every walkthrough on the platform actually impeded their ability to absorb all of what happens in a classroom during a visit.

The new evaluation systems that are in place due to the mandates of RTTT require numerous walkthroughs and communication between the evaluator and teacher. The age of accountability is being tested with these new systems and there is not proper support in place to be effective. The opinion of this author is if walkthrough observation quotas are met, the workload of evaluators and stress level for teachers will become unbelievable. The principals in this study agreed with this conclusion. The walkthroughs and observations take a great deal more time than instruments that were used in the past and each teacher receives at least four walkthroughs that are between ten to fifteen minutes each. In addition, there are two observations that last between thirty to forty-five minutes. The amount of time the components

of the instrument take has led to a feeling of “checking off” the requirements instead of focusing on the results and improving the quality of what they are actually evaluating. The frustration felt by the principals due to the time the instrument takes to complete is not grounded in the idea that they have more to do, but the realization that the mandates are creating a feeling of sterilization and institutionalization.

Surveys are another component of TKES that impacts the focus of interpretation for this study. The surveys provide a means for collecting student perception data in grades 3-5, 6-8, and 9-12. Among the advantages of using this survey design are the rapid turnaround in data collection, the limited cost in gathering the data, and the ability to infer perceptions of a larger population from smaller groups of individuals. In the Teacher Keys Effectiveness System, the student Surveys of Instructional Practice is used as a measure of teacher effectiveness and documentation in that they provide support for four of the standards that reflect the direct experience of students in classrooms. Data from student surveys must be used to inform the rating of a teacher’s performance on these standards on the Formative Assessment (when data is available) and on the Summative Assessment. Universities routinely survey students on how professors are performing, but until recently, K-12 students have not been given the chance to evaluate their teachers. This is because, although students spend hundreds more hours in classrooms than any administrator, people doubt that students can provide valid, reliable, and stable responses about the quality of teaching (Ferguson, 2012). The participants in this study agreed that they needed more training to make the surveys beneficial. They mentioned the surveys in terms of their experiences and how they could make them better, but also where they could benefit from future trainings. The principals’ did not agree with Ferguson (2012) who said, “Students know good instruction when they experience it as well as when they do not.”

These above mentioned components, much like the evaluation instruments developed by Taylor and Cubberly during the scientific method era of teacher evaluation, this instrument uses the formative evaluations to determine the summative. It is a process that emphasizes measurement and analysis of data to ensure that teachers and schools are productive. However, it is apparent that some of the participants in this study believe it is too scientific and not "human" enough – missing the interaction and relationship building due to the amount of documentation that is required. The principals spoke about how they love being in the classroom – but the instrument makes it constraining by having to fill the requirements and it is just too much right now without any flexibility.

The components of TKES discussed by the participants in this study recognized several that Cashin (1996) recommended for a successful evaluation system. However, it was the consensus and conclusion derived from the data that some needed further development.

1. Clarify Institutional Goals: While the goals of TKES have been identified by the district, a better understanding of the outcome of being rated at a certain level is needed.
2. Decide on the purposes of the data to be collected: The survey data is not collected in a manner that serves each teacher equally. In addition, the inter-rater reliability of all data collected on teachers was reported to be weak.
3. Use pilot programs when appropriate: TKES was implemented in some school districts prior to the one reported in the study.
4. Foster open communication: The communication between the evaluators and teachers is not consistent between schools. Again, the participants noted a desire to have more inter-rater reliability in this area.

5. Obtain support from high level officials: State and local officials have done the training and will continue to be available to evaluators to explain and define components of the instrument.
6. Ensure that the system is flexible and legal: The system is legal, but there were many frustrations reported that prove it is not as flexible as some principals would like.
7. Define major and minor faculty responsibilities: The duties and responsibilities of faculty are defined under the standard of professionalism.
8. Use multiple data sources: Walkthroughs, observations, surveys and student data are used for formative and summative evaluations.
9. Define the criteria: The criteria are clearly defined in the TKES handbook, the problem lies in its size and ease of navigation. The participants noted that they rarely use the handbook as a reference guide.
10. Train the evaluators: Training was done by the State Department however, the evaluators in this study requested future training in the areas of the platform, process, and surveys so there was more consistency among evaluators.

Another focus of interpretation in this study is inter-rater reliability, or concordance. It is the degree of agreement among raters. It gives a score of how much homogeneity, or consensus, there is in the ratings given by those who judge something. In this case, the principals are the judges (i.e. evaluators). The inter-rater reliability in regards to TKES is the concurrence between ratings that are earned by the teacher. The participants in this study noted that there were several steps they have taken in controlling the inter-rater reliability in their building, but it is an area to focus on for future preparation. The principals stated throughout the data collection period, in many different ways, in which this is an area they would benefit from more consistency

throughout their levels, areas, and the entire district and state. The principals reported that they needed more consistency in rating and giving feedback to the teachers. There was little agreement on when feedback should be written or oral – or even given at all.

"Feedback can be said to describe any communication or procedure given to inform a learner of the accuracy of a response, usually to an instructional question," (Carter, 1984, p. 53). Feedback should be purposeful and deliberate towards the components of the evaluation tool that are being evaluated and should be ongoing in order to improve teacher and student performance. "Evaluations should provide all teachers with regular feedback that helps them grow as professionals, no matter how long they have been in the classroom" (The New Teacher Project, 2010, p. 3). The requirement for TKES are four ten-minute walkthroughs, two thirty-minute observations, two formative assessment and one summative assessment per year, per teacher, with feedback required on each; approximately 180 minutes per teacher. Therefore, the feedback given, as reported by the principals in this study, is not as productive and relevant as it should be due to time constraints.

Feedback occurs when outputs of a system are "fed back" as inputs as part of a chain of cause-and-effect that forms a circuit or loop. In this research, the system is inferred to be evaluation instrument and all components that comprise what a teacher is rated on within it. Simple causal reasoning about a feedback system is difficult because the first system influences the second and second system influences the first, leading to a circular argument. This makes reasoning based upon cause and effect tricky, and it is necessary to analyze the system as a whole. Specifically, with teacher evaluation, the feedback given to a teacher from one experience influences the outcome and reaction to the next and so forth. The participants in this

study noticed that their feedback is limited due to the amount of time it takes to upload comments and ratings onto the platform.

Feedback should be purposeful and deliberate towards the components of the evaluation tool that are being evaluated. Feedback should also be ongoing in order to improve teacher and student performance. The principals in this study noted that the pressure of doing four ten-minute walkthroughs, two thirty-minute observations, two formative assessments and one summative assessment per year, per teacher, with feedback required on each is daunting. Although the theory is ideal, the realism of implementation is troublesome in the minds of these principals.

Again, the main reason that it is troublesome is time. In order to differentiate on the basis of teacher understanding and level of competence, the principals realized that there are several areas where they can do so. The only conference that has to be in person is the summative assessment at the end of the year. During the other formative assessments, feedback can be written and loaded on the platform or given in a small or large group setting. As the interviews took place, it was clear that all principals gave feedback in a different manner and some inter-rater reliability in this area is needed to ensure the feedback that teachers need to improve their performance is being given.

The inter-rater reliability of how the standards being evaluated are decided is also a concern that was derived from this research and found to be a need for future principal preparation. During the formative evaluations, all of the principals mentioned rating the standards in a different manner; some looked for the same three standards while other principals stated that they changed how they gave feedback as the year progressed; some mentioned that the comments were unique for each teacher and others wrote scripted, similar comments for all.

The principals are not opposed to rating the standards in a consistent manner, but felt there was not guiding factors to implement. They understood the importance of giving feedback on the standards, but disagreed on which standards each evaluation should address. Regardless of how they felt the standards should be evaluated, if feedback is not given immediately or geared toward a specific goal, it might not be as productive. As Scheeler, et.al. (2004) stated, "Feedback should be immediate, specific, positive, and corrective." Studies show that learning takes place when feedback is given more frequently (Scheeler, 2004); however, as noted above, the time it takes to give the feedback is the most imposing factor. In addition, if different teachers are getting feedback on different standards the impact on the school as a coherent organization reaching toward a similar goal might be jeopardized.

Evaluators must be able to ask questions, tell the truth, and encourage feedback from members of the community, (Bolman & Deal, 2003). In the case of this study, the community is their teachers and the communication, in terms of the feedback given after an evaluation, is a critical part of the process. However, administrators may not feel equipped to give such purposeful feedback on evaluations "Evaluators often lack specific knowledge about the content areas in which they evaluate teachers, especially at the secondary level" (Donaldson, 2010, p. 54). Some of the high school principals in fact felt as though they do not have sufficient experience to evaluate certain departments. However, they all agreed that it isn't as much the content, but rather the delivery of the content, that is what should be evaluated. To this, they all responded by being very confident in their ability to evaluate – if given the time and professional development that will lead to developing the skills needed to speak to each of the ten standards evaluated through TKES in a consistent manner.

Training is another focus of interpretation for this study. All evaluators must be fully trained and credentialed by a state and/or district credentialed trainer before using Teacher Keys Effectiveness System (TKES). The credentialing assessment is a minimal competency assessment that measures participants' understanding of the information and practice provided during training. It includes both recall of specific TKES information and practice of all the TKES processes and is considered a bridge to ongoing learning as evaluators work with TKES. The expectation is that evaluators continue to familiarize themselves with the TKES process as they work within their districts, however, there is no mandate on how much (or how little) continued learning takes place. The training opportunities are not communicated directly through the state or the district, but rather, evaluators must seek them out. As one principal stated, "You don't know what you don't know," meaning – if you are not in attendance at the training, the information given is not going to impact your ability to implement the instrument effectively. One of the main factors that have motivated the nation's move to re-design teacher evaluation is that, historically, traditional evaluation systems have not differentiated among high and low-performing teachers (Sartain, et.al, 2011). Current evaluation systems and training for evaluators does not provide teachers with the information they need to make timely and effective improvements in their instructional practice (McLaughlin, 1990). It is of particular concern that most evaluation systems do not differentiate between strong and weak instruction or provide sufficient feedback and, as recognized by the participants in this study, the trainings to help them with this task are not easily available and/or communicated as a need to complete.

A second area where the research participants mentioned needing additional training was on implementing professional development plans. A professional learning plan is, "A set of purposeful, planned actions and the support system necessary to achieve the identified goals.

Effective professional learning programs are ongoing, coherent, and linked to student achievement," (Killion, 2008, p. 11). In TKES, these plans are referred to as Professional Learning Plans (PLP's) and one is created by the evaluator and approved by the principal within the electronic platform. It may be developed by the evaluator in collaboration with the teacher, coaches, mentors, or other qualified individuals and provides guidelines and timelines for specific, mandatory professional learning which supports immediate improvement of teacher practice and increased teacher effectiveness. The participants mentioned needing additional training on navigating the PLP process through the platform and accessing the resources that are available to help strengthen the plan. Another training topic that is aligned with the platform and its use is the over-all TKES process. The six steps for TKES implementation are outlined in the TKES Handbook, but, as referenced above, the handbook is not used consistently or frequently. The principals in this study noted that they would like future training in navigating the platform and implementing the process in a more effective, efficient manner.

Whether the training should be provided by the state or the local school district is important to the relevancy and purpose of this study. In order to report a clear directive as a needs assessment, the research provided must lend itself to proper use in planning future preparation for principals. The participants in the study reported the way they were supported through the state and local training on the above mentioned platform, process, and PLP components and most felt future training should come from the district because it could serve their needs better than the state. It has been their experience, especially through this process, that the district does a better job developing their people. Although they were not knocking the state, they recognized that Cobb knows their people and they do a great job identifying, differentiating, and addressing areas of need. All of the principals also felt that the training needed to be done

by level: elementary, middle, and high. They realize that the needs differ by level and what you are rating looks different as well. They felt the training should be differentiated by areas within the district as well because teachers in different areas are dealing with students and staff that have different struggles – so the professional development could be geared toward what those principals need to know and understand to help their own teachers. There was one principal that mentioned she thought the training should be done by the state because then it would be coming from one source and since it is a state instrument and they are evaluating from it, it would be helpful to know what they see as those being appropriate.

The research collected in this study will expand on previous studies on the topic of teacher evaluation. It is significant because it focused on the preparation of the building principal. Mitgang (2012) highlighted the importance of training and support for principals, after they are hired, that is tailored to both individual and district needs. This study specifically looked at the training and support in regards to the implementation of a new evaluation system and uncovered valuable information for future principal preparation in the school district studied. "Indeed, much of the published literature on the principal's role in evaluation consists of explanations of procedural and legal aspects of evaluation and calls for better training of principals" (Painter, 2000, p. 2). Prior studies have showed that principals are not adequately prepared to be effective evaluators and the participants in this study, for the most part, agree.

Other barriers that affect the evaluation process and its outcome include deficiencies in the instrument tool and in the organization itself (Stronge, 1991). Many studies have looked at the process and identified the most common mistakes that evaluators make (Freeberg, 1969; Johnson, 1997). In this study, the process was mentioned in terms of an area where future professional learning should be directed. The process was clearly interpreted in different ways

and it has caused a great deal of inconsistency among buildings. The principals overwhelmingly agreed that this consistency, inter-rater reliability, was needed to make the evaluation instrument more effective. Finally, many of the studies previously conducted show that evaluations done by administrators are often based on perception and influenced by factors such as the halo effect or leniency (Haefele, 1993; Wise et al., 1985). In this study, the principals stated that, although the new instrument mandates a lot of time, the number of formative assessments allow for a more thorough summative evaluation.

The data from this research has the potential to guide the regulations regarding the training principals receive in implementing an instrument and giving feedback during the evaluation process. It is through the findings of this study that practices can be implemented to ensure that principals have the skills and knowledge necessary to be effective. Ongoing supports and training are crucial for a transformational change in the way teachers are evaluated because almost all respondents, even those who believed that principals supported the teacher evaluation program, felt that principals lacked sufficient resolve and competence to evaluate accurately (Wise et al., 1985). The skills acquired by principals under the new system will help them facilitate teacher self-reflection, growth, and improved student achievement.

The findings of this study are relevant to the field of education and specifically teacher evaluation because it is crucial for districts and states to understand the level of preparedness their principals report to have to conduct effective teacher evaluations. Districts can use this data to determine if they have trained their principals sufficiently. Kimball and Milanowski (2009) found that principals who conducted more valid evaluations reported that district training helped build their evaluation skills. Therefore, the validity of evaluations will increase as principals become more prepared. If Georgia principals are going to take time away from other duties to

fulfill the vast observation and feedback requirements of TKES, proper evaluation training should be provided so it is not viewed as a waste of time and taken less seriously.

Teacher evaluation has intrinsic importance to multiple groups of people and this needs assessment will provide a concrete, concise outline of what current principals reported to need to become effective evaluators. This study covered a current reality that evaluation systems are changing and the level of preparedness of the principals needs to be addressed. The study took place in a real-life setting where the implementation of teacher evaluation systems was based on relatively new theories that have been widely accepted but minimally evaluated.

Meta-Reflection on Research Design and Methods of Study

The qualitative research model that was used for this study emerged from research efforts in the social sciences and was appropriate for this study because it is effective in understanding the total environment in which training takes place (Creswell, 1998). The four research questions acted as a guide for data collection and analysis and the researcher collected data from multiple sources to determine evaluators' perceptions on their preparedness to conduct effective evaluations using TKES. The research is particularly useful to school district officials who are planning the training for evaluators, especially principals, in implementing, or continuing the use of, a teacher evaluation instrument.

The individual interviews with the nine principal participants offered an opportunity to gain insight into the participants' perspective (Creswell, 2007). As predicted in the research proposal, the years of experience was a key component in preparedness to conduct effective evaluations. The research questions were open-ended, general, and focused on understanding the central phenomenon of the study: What is the level of preparedness principals report to have while effectively implementing a new evaluation instrument. Open-ended questions are recommended by qualitative researchers because these types of questions lead the participant to

discuss their experiences as they understand them (Hatch, 2002). The researcher found this to be true during the interviews and discovered that the principals' responses were genuine and true. The questions were intended to provide the most accurate picture of the lived experiences of the participants thus, as the responses varied, the questions were adjusted to fit the needs of the interviewee and researcher. The interviews were almost an hour long for each participant. In an ideal world, with no limitations, longer interviews or several follow up interviews could improve on the results of the study. However, in the essence of reality, time impacted the availability of not only the researcher, but the participants as well.

The researcher remained neutral in nature and respectful of points of perspective to be mindful of reliability within the study. The data was not easily coded due to the fact that the questions were open-ended, but the time it took to code the data led to several concrete, clear conclusions. Open-ended questions allow for a greater variety of responses from participants by are difficult to analyze statistically because the data must be coded or reduced in some manner. As the coding took place, the researcher was systematically aware of the similarities and connections between questions and responses. For example, during the questions about inter-rater reliability, the principals noted several different ways of conducting the formative evaluations, however, they all connected in the fact that they were all different ways. It wasn't the actual way in which the evaluation was being conducted that the researcher was focused on, but rather that they vary so drastically and are not consistent among evaluators. The responses from the questions asked during the study allowed the researcher to validate the purpose of the study and recognize the needs of future principal preparation.

After the individual interviews were completed, the researcher conducted a focus group interview to explore the topic in a different setting. According to Marshall and Rossman (2011)

the method of interviewing participants in focus groups comes largely from marketing research but has been widely adapted to include social science and applied research. The researcher created a supportive environment and asked focused questions to encourage discussion and the expression of differing opinions and points of view. During the focus group interview, conversations within the group were led by the participants and not the researcher. The strengths of this interview was that it was socially oriented, studying the responses of the participants in an atmosphere more natural than artificial experimental circumstances. The participants generated new ideas between each other that led to a better understanding and greater validity to the purpose of the study in understanding their needs for future preparation.

The primary ethical issues that arose from the interviews centered on the dynamics of power and influence. Due to the fact that the researcher is herself a principal in the school district being studied, there was a level of trust that needed to be established within the first minutes of the interview. The researcher made it a point to not appear over-powering in her manner of questioning and was sensitive to the dynamics of the group during the focus group. The interviews were also very time-consuming and exhausting – for both the researcher and the participants. "Undoubtedly, conducting interviews is taxing, especially for inexperienced researchers engaged in studies that require extensive interviewing," (p. 173). As mentioned earlier in this chapter, time is not something principals have an excess of, so that was consistently on the mind of the researcher. One way to ease the impact of the time constraints might have been to conduct the research during the summer when principals are not as busy.

The data collection phase of research presents ethical issues into informed consent procedures, deception or covert activities, confidentiality toward participants, benefits of research to participants over risks, and participant requests that go beyond the social norms. The

researcher protected the anonymity of the participants, explained the purpose of the study, and did not engage in deception about the nature of the study. Maintaining professional interactions and staying in compliance to the interview questions asked, helped maintain ethical behavior and the validity of the research results.

Creswell's (2013) explanation of a purposeful sampling approach was used to determine the specific principals to be interviewed during this study. In order to understand the differences between elementary, middle, and high school principal responses, three from each level were interviewed and their differences were analyzed when coding the data. Neither the personal identity nor the identity of the schools was reported. In order to accommodate a setting that was comfortable for the interview, the researcher offered a location that was easily accessible, usually being the principal's own building. An effort was made to schedule it during natural school breaks or after school hours, which were the times preferred by the participants. By scheduling in this way, the researcher believes the data collected is more accurate than if the school principal were faced with daily responsibilities while being interviewed. Follow up interviews were not needed with the principals. However, if conducting the study again, the researcher might find it useful to create a few further questions to address the three themes that emerged: inter-rater reliability, experience, and training to gain insight on more specific needs for future preparation.

The above mentioned, purposeful sample was chosen for this study and was consistent with the research questions being posed. A major strength of purposeful sampling is that it allowed the researcher to select participants that can provide insight into the question being studied (Creswell, 2007). In this study, only principals from Cobb County were considered because the researcher was looking specifically at principals who were in the first year of

participation using TKES. Although the researcher created sub-groups at the beginning of the study (elementary, middle and high school levels), once coded, it was concluded that the sub-group responses did not vary enough to report the data in that manner. In the future, a different study might include more principals at each level, as a larger sample might determine if there is a difference among sub groups. The intent of selecting this purposeful sample was to develop a further understanding of evaluator preparedness to conduct effective evaluations. The interviews of the selected participants provided a voice to evaluators who might otherwise be silent on the subject and responses were interpreted by the researcher to be valid and reliable.

One weakness of purposeful sampling, which the researcher found to be a challenge in this study, was that the number of participants might be considered a "convenience sample" (Hatch, 2002). If the researcher had varying responses then she might be inclined to agree with this, but, although there were only nine participants, those nine responded in a consistent manner. The relationship between the researcher and the participants is another weakness that posed itself to be a small challenge in the study. Being a principal and evaluator using the instrument herself, the researcher remained aware of the biases that could emerge from interviewing peers. It helped that she is in her first year as a principal, so the participants have seniority and more experience in this area. Being mindful of the hesitation that might present itself in answering the interview questions allowed her to clarify any misconceptions as the interviews were conducted. A lot of body language and eye contact was noted in the analytical notes on how the participants felt when responding to a particular question. At times they did not say anything, but their gestures indicated that they were either frustrated or unhappy with the topic they were discussing. The thoughts of the researcher were not reported, as they are merely that – assumptions from the researcher. If there were more time to form a stronger, more trusting

relationship, the researcher would have expanded on the assumptions that were spoken through such body language.

After the data was compiled the needs of the principals were reported. As a systematic process, this needs assessment determined and will hopefully address the needs or gaps between current conditions and desired conditions with TKES usage by the principals in this school district. The discrepancy between the current conditions and the wanted condition was qualitatively measured to appropriately identify the needs. As a needs assessment, the results can be part of a planning process to effectively support the positive attributes of the evaluation system and identify the areas of weakness. By clearly identifying the problem and evaluating the process, the resources being used to train and prepare evaluators can be directed toward developing and implementing a feasible solution (Fulgham & Shaughnessy, 2008). The desired results will be articulated and the distance between the results are the actual need is what the researcher focused upon. In relation to TKES and what evaluators need, a training needs assessment was appropriate to use because it is a systematic inquiry of training needs within an organization for the purposes of identifying priorities and making decisions, and allocating finite resources in a manner consistent with identified program goals and objectives (Triner, Greenberry & Watkins, 1996).

The qualitative research presented in this study meets rigorous standards that strive for "understanding" a deep structure of knowledge that comes from visiting personally with participants, spending extensive time in the field, and probing to obtain detailed meanings (Creswell, 2013). In conducting a needs assessment for principal training to give meaningful feedback, the interview process was the most efficient and reliable method to use to attain the amount of detail necessary to answer the research questions. In order to meet the standards of

rigor for this study, the researcher developed interview questions that described in detail the lived experiences of the principals interviewed. Creswell (2013) summarized validation in qualitative research to be an attempt to assess the accuracy of the findings, as best described by the researcher and the participants (p. 249-250). The strength of qualitative research is the extensive time spent in the field, the detailed thick description, and the closeness of the researcher to participants in the study all add to the value or accuracy of the study. This study included the following standards of rigor: Triangulation, clarifying researcher bias, member checking, and thick, rich descriptions as explained in detail in chapter four.

Practical and Policy Implications of the Study

The study informs educators in philosophical and practical manners. The body of knowledge on teacher evaluation and the impact it has on principals has expanded as a result of this study. Multiple groups can benefit from the findings of this research study. These groups include teachers, principals, lawmakers, district level administrators, and university level education professors.

Teachers need to involve themselves in their own evaluation by having open communication with their evaluators and recognizing where the troubles in conducting effective evaluation might lie. The teachers could impact how the principals rate them by self-assessing their performance and having open communication with the principal. It would help to clarify expectations. The models magnify the evidence of policy makers, researchers, and the evaluators, but the voice of the teacher is missing in teacher evaluation.

The study informs principals of the shared challenges that they might face when implementing a new evaluation instrument. The job requirements of principals need to be examined if they are to effectively use TKES. In addition to the increased workloads, principals are risking a decline in positive interactions with their teachers due to the nature of the

instrument. District level officials can learn from this study and develop the support needed for future preparation principals might need to implement the instrument effectively. District level administrators can influence the success of teacher evaluations in their district by giving the time and resources needed for implementation.

Lawmakers can benefit from the findings of this study. It would be beneficial for lawmakers to read studies such as this before making new laws or changing education policy. The interviews from the nine principals would have had more meaning if their responses were heard by state lawmakers who hold positions of great importance and influence on educational change. The history of educational policy, in which the field is being held accountable through, is lacking input from the very people who evaluate and are evaluated. This study responded to the identified research problem by challenging political support of teacher evaluation reform with the philosophical support of the principals who have been impacted by the changes.

The study also informs practitioners who create teacher evaluation instruments on the feelings toward components of this specific instrument. Professionals working at the university level who are involved in principal preparation programs can also benefit from this study. By reading it, faculty at the higher levels of education will understand the expectations of principals and recognize the changes that have happened. This change impacts how principal preparation programs are designed, implemented, and evaluated. The university's education department is ultimately evaluated by the quality of principals they graduate.

Data from research question number one focuses on what the benefits and challenges of using the TKES instrument are to the participants in this study. The main challenge an evaluator might face was reported by all participants, overwhelmingly, to be time. The time it takes to implement the formative assessments, give the feedback, and document performance is

significant when considering the numerous other duties a principal has. Another challenge that appeared through the data was consistency. The inter-rater reliability was seen as an issue across the board in all major data themes. There was no consistency in implementation or continuation of using the instrument. Through investigating answers for research question three, how effective feedback is given while using the evaluation instrument, it was found that there was not consistency in how feedback is given either.

Research question number three specifically targeted what future training needs principals' report they need to use the instrument to give effective evaluations and district level officials can take the data delivered and construct informative, time-worthy staff development for their evaluators. The principals mentioned that they did not feel there was adequate training on the platform, the process, or the student surveys that are mandated. The district reported through the study can benefit the most as the sample was from principals all in that district. The future preparation can be individualized for the levels (high, middle, elementary) or based on principal experiences (whether they have already used the instrument or not).

Responses from research question number two explain the level of ease in navigating the instrument in comparison to those used in the past and are important in terms of how the study educates practitioners and scholars who design teacher evaluation by presenting them with additional information to add to the knowledge they have on the theory of evaluation. The process is ever-changing and those who develop new instruments can look for input from evaluators to make them more effective and efficient. Researchers in the field of teacher evaluation can use the feedback given by these principals to modify current instruments. The components of an evaluation instrument tend to change with the current education policy. Depending on the measures needed to be successful, practitioners could recognize the issue of

time and understanding and create a way of addressing those concerns within the instrument itself. It was found in this study that TKES jeopardizes how realistic it is for evaluators to utilize it to the fullest potential and, although new policy directs evaluation instruments to have high standards, it remains to be seen whether the concerns principals have are legitimate warnings of what will happen with these changes in procedure and routine for teacher evaluation.

Future Research on the Topic

Teacher evaluations are here to stay and it can be expected that they will continue to change in format, method, and instrument. In addition, principals will continue to be responsible and held accountable to implement the instruments in an effective manner. There are several possibilities for further research. The following are ones in which the researcher feels have the most potential and would be the most impactful.

The findings of this study indicate that the new teacher evaluation instrument that is being used in Georgia, TKES, is not being implemented as effectively as it has potential to be, and principals need future preparation through professional development. While the research in this study begins to examine the impact of the new instrument and the ways in which principals desire more training, more data should be collected to derive a conclusive understanding of specific areas where growth can lead to implementing the instrument in the most effective manner. A school-to-school comparison of principal perception of their preparedness in local school districts or throughout the state could be conducted to help achieve this goal.

Another worthy expansion of this study would be to determine if the principal's perceptions of the new model is consistent with their general belief of how teachers should be evaluated. A comparative study that describes the perceptions of principals in a follow-up study would be worthwhile because it would help to discover if perceptions change over time as the new instruments move beyond the first year. There is a possibility that, with experience and

training, attitudes will change. A comparative study that looks at the before and after perceptions of principals would provide a more complete narrative of the impact teacher evaluation is making in the lives of principals in Georgia.

Another recommended study is one more specific to the requirements and demands placed on principals, in addition to, being teacher evaluators. Through the course of this study and reflection on the data results, principals are impacted just as much, and possibly more, that teachers with changes in teacher evaluation and fulfilling the accountability measures. In what became an ironic development throughout the study, instead of the new instrument helping to generate support for the teachers in areas of need, the results indicated that there is less time to effectively achieve that goal due to the managerial requirements of the instrument. Given that principals are in jeopardy of being overburdened, a study of their daily routine and duties would be helpful in understanding what could be realistic expectations of a teacher evaluation instrument. The job of a principal is too important to our students, teachers and communities to have them consistently feeling overwhelmed and under-supported.

Self-Reflection

During qualitative research, there never seems to be enough time or end to adding more data. It was hard to limit the study to the four research questions because each interview question connected to the research questions could have numerous follow-up questions. Being able to focus on the immediate questions and directly correlated responses was not as simple as it sounded to the researcher. As far as time, it was enlightening to see how the study can be limited in terms of ease to complete. The data is based on others' responses and that requires the researcher to depend on their schedule for availability. Patience was a lesson learned through this and many other aspects of writing the paper.

Staying aware of researcher bias was a huge lesson during the writing of this paper and the collection of the data. As a principal herself, the researcher is also conducting evaluations using TKES and has strong opinions on the instrument and how feedback is given to teachers. The challenge was met by conducting mock interviews with a couple of principals not involved in the study and reflecting on the tone and content of the interviews. Separating herself from the data helped the researcher become aware of the creation of the overall themes without personal bias. In addition, as she heard others talk about the instrument, she changed some of the feelings she had about it.

An unanticipated finding of this study that the researcher has reflected on was the complexity of teacher evaluations and the impact it has on principals. Going into this study, the perception was that teacher evaluation was just the process to determine how good a teacher is at teaching. The study, however, enlightened the researcher on many issues circling around the very notion and purpose of teacher evaluation. There is a real disconnect between state policy makers and educators. There needs to be more of a partnership where the legislators listen to the very voices that implement the instruments in which they mandate.

In addition, a reflection that manifested from this study was that most evaluators do not know the very instrument they are using. It would be beneficial if all principals could invest the time and energy the researcher did to understand the instrument itself. The data collected before, during and after the interviews helped her understand the different components of the instrument. Navigation of the platform is not easy and without a need to do so, the researcher probably would not have. Although there is a training manual and there are multiple emails sent regarding components of the instrument, like many of the participants, the researcher would not have looked at them if the study hadn't needed her to do so.

Qualitative study is a way to understand the lives and experiences of people. It is a daily challenge for people in all walks of life to make sense of their life and find meaning and purposes in not only what they do, but how they do it. Most educators, by nature, are seekers of further knowledge and understanding. Therefore, it is a challenge to them when they are doing something that does not fit in their philosophical understanding of what is "right". At the conclusion of this study, evaluation, especially TKES, seems to have presented itself as one of those things that will add to the challenge of doing something in which they are not comfortable. The message is clear that common sense and input from the experts *in the field which is being evaluated* should be given to opportunity to communicate their wants, needs, and opinions about the evaluation instrument in which they use. We must be realistic of principal's time and allow for their professional judgment to guide the process. The new teacher evaluations can be successful without incorporating the traditional, top-down, power arrangement with the principal as a boss reporting to "big brother" at the state department. We must have better awareness of critical factors that lead to a more successful transition toward teacher evaluation reform.

The twenty-first century is an era of major educational change as a result of federal accountability measures such as Race to the Top. In a review of teacher evaluation, it is clear there are many gaps that need to be filled with literature that offer the perspectives of the teachers and principals whom are directly affected by these changes. More studies are needed to evaluate educational decision makers who offer the "cure" for our educational problems and concerns. Research needs to continue to move into schools and view the phenomenon through the lens of the principals who will be conducting the evaluations and making the decisions that impact their students and future achievement.

REFERENCES

Acheson, K. A. & Gall, M. D. (2003). *Clinical supervision and teacher development: Pre-service and in-service applications*. USA: John Wiley and Sons.

Badertscher, N. (2014, Jan 7). New evaluation pilot 'skewed,' with too few unsatisfactory teachers. *The Atlanta Journal Constitution*, Retrieved from; http://www.ajc.com/news/news/new-evaluation-pilot-skewed-with-too-few-unsatisfa/nTpKN/

Ballard, K. & Bates, A. (2008). Making a connection between student achievement, teacher accountability, and quality classroom instruction. *The Qualitative Report 13(4)*, 560-580.

Benedict, A.E., Thomas, R.A., Kimering, J., Leko, C. (2013). Trends in teacher evaluation: What every special education teacher should know. *Teaching Exceptional Children, 45(5)*, 60-68.

Bolman, L.G. & Deal, T.E. (2008). Reframing organization: Artistry, choice, and leadership. San Francisco, CA: Jossey-Bass.

Boyce, A.C. (1915). *Measures for measuring teacher efficiency*. Chicago, IL: The Chicago Press.

Brandt, C., Thomas, J., & Burke, M. (2008). State policies on teacher evaluation practices in the Midwest Region. (REL Technical Brief, REL 2008 – No. 004). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

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

Burke, P., Lind, K.G. (1987) Performance assessment techniques for teacher career ladder advancement. *NASSP Bulletin* 71: pp. 27-35.

Carnegie Forum on Education & the Economy, Task Force on Teaching as a Profession. (1986) *A nation prepared: Teachers for the 21st century*. New York: Carnegie Forum on Education and the Economy.

Cashin, W.E. (1996). Developing an effective faculty evaluation system. IDEA Paper No. 33. Kansas State University, Manhattan, KS.

Clark, D. (1993). Teacher evaluation: A review of the literature with implications for educators. Unpublished manuscript, California State University at Long Beach, CA.

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

Corbin J. & Strauss, A. (2008). *Basics of qualitative research: techniques and procedures for developing grounded theory. Third ed*. Thousand Oaks: Sage Publications.

Creswell, J.W. (2013). *Qualitative Inquiry and Research Design Choosing Among Five Approaches*. Nebraska, Lincoln: SAGE Publications, Inc.

Creswell, J.W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches, 2nd ed*. Thousand Oaks, CA: Sage.

Creswell, J.W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson/Merrill Education.

Creswell, J.W. (2007). *Qualitative Inquiry and research design: Choosing among five approaches. Second ed*. Thousand Oaks: Sage Publications.

Creswell, J. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.

Cubberly, E.P. (1909). *Changing conceptions of education.* Cambridge, MA: The Riverside Press.

Cubberly, E.P. (1929). *Public school administration*. Boston, MA: Houghton-Mifflin.

Danielson, C. (1996). Enhancing professional practice: A framework for teaching. Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2007). Enhancing professional practice: A framework for teaching (2nd ed.). Alexandria, VA: ASCD.

Danielson, C. (2011). *The framework for teaching evaluation instrument.* Princeton, NJ: The Danielson Group.

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

Darling-Hammond, L. (2004). Standards, Accountability, and school reform. Teachers College Record, 106(6), 1047-1085.

Darling-Hammond, L. & McLaughlin, M. (2011). Policies that support professional development in an era of reform. *Kappan Magazine, 92(6)*, 81-92.

David, J. (2008). Classroom walkthroughs. Educational Leadership, 65(4), 81-82.

Denham, C.H. (1987). A perspective on the major purposes and basic procedures for teacher evaluation. *Journal of Personnel Evaluation in Education*, 1, 29-32.

Dinan, S. (2009). Obama wants teacher 'accountability'. Washington Times. Retrieved from: http://washingtontimes.com/news/2009/mar/10/obama-calls-accountability-education/.

Donaldson, L. (2010). Collaborative communities of firms: Purpose, process, and design. New York: Springer.

Downey, C.J., Steffy, B.E., English, F.W., Frase, L.E. & Poston, W.K. (2004). *The three-minute classroom walk-through: Changing school supervisory practice one teacher at a time*. Thousand Oaks, CA: Corwin Press.

Drago-Severson, E. (2009). Leading adult learning: supporting adult development in our schools. Thousand Oaks: Corwin Press.

Dryer, K. & Renn, M.T. (2010). School leaders need specialized development. *Getting an Education, 29(6)*, 3-7.

Dunkleberger, G. (1982). *Classroom Observations--What Should Principals Look For?* NASSP Bulletin, 66(458), 9-15. [EJ 274 276].

Elmore, R.F. (2000). Building a new structure for school leadership. The Albert Shanker Institute: Washington, D.C.

Feeney, E.J. (2007). Quality feedback: The essential ingredient for teacher success. *Clearing House: A Journal of Educational Strategies, Issues, and Ideas, 80(4)*, 191-197.

Fitzpatrick, J.L., Sanders, J.R., Worthen, B.R. (2011). Program Evaluation: Alternative approaches and practical guidelines. Boston: Pearson.

Frase, L., Hetzel, R. (1990). *School Management by Wandering Around*. Lancaster, PA: Technomic Publishing Company.

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

Freire, P. (1970). *Pedagogy of the oppressed.* New York: Seabury.

Fulgham, S.M. & Shaughnessy, M. (2008). Q and A with education technology leaders: Interview with Roger Kaufman. *Educational Technology*, 49–52.

Fullan, M. (1993). *Changing forces: Probing the depth of educational reform.* Bristol, PA: Falmer.

Georgia Department of Education. (2012). Teacher keys evaluation system trainer manual. Retrieved from: http://legisweb.state.wy.us/InterimCommittee/2012/TKESHandbook.pdf

Georgia Department of Education. (2012). Georgia teacher keys effectiveness system implementation procedures. Retrieved from: http://www.gadoe.org

Georgia's race to the top plan. Retrieved from: http://www.gadoe.org.

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

Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. Washington, DC: National Comprehensive Center for Teacher Quality.

Goodwin, B. (2010). Changing the odds for student success: What matters most. Denver, CO: Mid-continent Research for Education and Learning (McREL).

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

Hatch, J. (2002). *Doing qualitative research in education settings*. Albany, NY: State University of New York Press.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hazi, H. (2012). Expert judgment: A concept for teacher evaluation in a postmodern world. A paper presented at the annual meeting of the American Educational Research Association. Vancouver, British Columbia.

Hobbs, D.L. (1990). A training appropriations process. *Training & Development Journal*. 44(5), 109-115.

Holifield, M. & Cline, D. (1997). How is instructional supervision done and who should do it? *NASSP Bulletin*, 81(590), 109-113.

Howard, B. B., & McColskey, W. H. (2001). Evaluating experienced teachers. *Educational Leadership*, 58, 48–51.

Hunter, M. (1994). Enhancing teaching. Upper Saddle River, NJ: Pearson Education.

Jacob, A. (2012). Examining the relationship between student achievement and observable teacher characteristics: Implications for school leaders. *International Journal of Educational Leadership Preparation, 7(3)*.

Johnson, B.L. (1997). An organizational analysis of multiple perspectives of effective teaching: Implications for teaching evaluation. *Journal of Personnel Evaluation in Education, 11*, 69-87.

Kennedy, M. (2005). Inside *teaching: How classroom life undermines reform.* Cambridge, MA: Harvard University Press.

Kersten, T.A. & Israel, M.S. (2005). Teacher evaluation: Principals' insights and suggestions for improvement. *Planning and Changing, 36(1&2),*47-67.

Kimball, S., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly, 45(34),* 34-70.

Knowles, M. (1998). *The Adult Learner*. Houston: Gulf Publishing.

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

Laguardia, A. & Pearl, A. (2009). Necessary educational reform for the 21st century: The future of public schools in our democracy. *Urban Review, 41*, 352-368.

Lane, M. (2000). HPT: A concept in search of analysis. *News and Notes,* 1 & 3.

Lashway, L. (2003). Role of the school leader. *Clearinghouse on Educational Management*. Retrieved from http://cepm.uoregon.edu/trends_issues/rolelead/index.html#developing

Lincoln, Y. & Guba, E. (1985). *Naturalistic Inquiry*. Beverly Hill, CA: Sage.

Lynch, J.M. (2012). Responsibilities of today's principal: Implications for principal preparation programs and principal certification policies. *Rural Special Education Quarterly, 31(2),* 40-47.

Mager, R.F. (1972). *Goal analysis*. Belmont, CA: Fearon Publications.

Marshall, K. (2012). Fine tuning teacher evaluation. *Educational Leadership, 70(3).*

Marshall, K. (2005). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan, 86(10),* 727-735.

Marshall, C., & Rossman, G. B. (2011). Designing qualitative research. (5th ed). Thousand Oaks, CA: Sage Publications.

Marzano, R.J. (2012). The two purposes of teacher evaluation. *Educational Leadership, 70(3).*

Mathers, C., Oliva, M., with Laine, S. W. M. (2008). Improving instruction through effective teacher evaluation: Options for states and districts. Research and Policy Brief. Washington, DC: National Comprehensive Center for Teacher Quality.

McLaughlin, M. (1990). Conscious versus unconscious learning. *TESOL Quarterly*, 24, pp. 617- 634.

McLaughlin, M. W. & Oberman, I. (Eds). (1996). *Teacher learning: New policies, new practices.* New York: Teachers College Press.

McManus, T. (2011). Students will grade teachers in new evaluation method. Augusta Chronicle.

McMillan, J. H., & Schumacher, S. (2006). *Research in education: Evidence-Based Inquiry*. New York: Pearson Education, Inc.

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

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San Francisco: Jossey-Bass.

Mielke, P. & Frontier, T. (2012). Keeping improvement in mind: Comprehensive teaching frameworks can help schools empower teachers, not just judge them. *Educational Leadership 70(3).*

Miller, R. (2000). Greater expectations national panel: A summary of efforts in school reform since 1983. Retrieved from: http://www.greaterexpectations.org .

Millman, J. & Darling-Hammond, L., Eds. (1990). The new handbook of teacher evaluation: Assessing elementary and secondary school teachers. Thousand Oaks: Corwin Press.

Mitgang, L. (2012). The making of the principal: Five lessons in leadership training. New York, NY: Wallace Foundation. Retrieved from: http://www.wallacefoundation.org/.

National commission on excellence in education. (1983). A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education (Research report). Retrieved from U.S. Department of Education website: http://ww2ed.gov/pubs/NatAtRisk/.

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

Ostovar Namaghi, S.A. (2010). A data-driven conceptualization of teacherevaluation. *The Qualitative Report, 15(6),* 1504-1522.

Painter, S. R. (2000). Principals' efficacy beliefs about teacher evaluation. *Journal of Educational Administration, 38(4),* 368-378.

Patton, M. (2002). Qualitative research and evaluation methods, 3rd ed. Thousand Oaks, CA: Sage.

Patton, M. (1990). *Qualitative evaluation and research methods*. (2nd ed.) Newbury Park, CA: Sage.

Peterson, K.D. (2000). Teacher evaluation: A comprehensive guide to new directions and new practices (2nd ed). Thousand Oaks: Corwin Press.

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

Pitler, H., & Goodwin, B. (2008). Classroom walkthroughs: Learning to see the trees and the forest. *Changing Schools*, 9-11.

Politics and Policy Project. (2014). University of West Georgia. Retrieved from: http://politicsandpolicyproject.wiki.westga.edu/Teacher+Evaluation+in+Georgia

Popham, W.J. (1987). *The merits of measurement-driven instruction.* Phi Delta Kappan. May, pp. 679-682.

Race to the top. Retrieved from: http://en.wikipedia.org/wiki/Race_to_the_top.

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

Rummler, G. & Brache, A.P. (1995). *Improving performance: How to manage the white space on the organization chart.* San Francisco: Jossey-Bass.

Saphier, J. (2008). The skillful teacher: Building you teaching strengths. Acton, MA. Research for Better Teaching.

Sartain, L., Stoelinga, S.R., & Brown, E.R. (2011). Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation. Consortium on Chicago School Research, Chicago, IL.

Scheeler, M.C., Ruhl, K.L., McAfee, J.K. (2004). Providing performance feedback to teachers: A review. Teacher Education and Special Education, 27(4), 396-407.

Schmoker, M. (2006). Results *now: How we can achieve unprecedented improvements in teaching and learning.* Alexandria: Association for Supervision and Curriculum Development.

Schmoker, M. (2006). The New fundamentals of leadership. SEDL letter, 17(2), 3-7.

Schmoker, M. (1999). *Results: The key to continuous school improvement.* Alexandria, VA: Association of Supervision and Curriculum Development.

Scriven, M. (1988). Philosophical inquiry methods in education. In R. Jaeger, Ed. Complementary Methods, pp. 131-183. American Educational Research Association.

Sergiovanni, T. J., & Starratt, R. J. (2007). Supervision: A redefinition (8th Ed.). New York: McGraw Hill.

Shakman, K., Breslow, N., Kochanek, J., Riordan, J., & Haferd, T. (2012). Changing cultures and building capacity: An exploration of district strategies for implementation of teacher evaluation systems. Retrieved from: http://ltd.edc.org/sites/ltd.edc.org/files/.

Shakman, K., Riordan, J., Sanchez, M.T., Cook, K.M., Fournier, R. & Brett, J. (2012) An Examination of Performance-Based Teacher Evaluation Systems in Five States. Regional Educational Laboratory Northeast & Islands (129).

Shearer, L. (2013). Georgia should wait on new teacher evaluation system. *Online Athens.*

Sisson, G.R. & Swanson, R.A. (1990). Improving work performance. *Instructional Technology.*

Spring, J. (2008). *The American school from the puritans to no child left behind*. New York, NY: McGraw-Hill.

State of the Union, 2009.

Strauss, V. (2012). Georgia professors blast teacher evaluation system. The Washington Post. Washington, D.C.

Stronge, J.H. (2007). Qualities of effective teachers. Alexandria: *Association for Supervision & Curriculum Development.*

Stronge, J. H., & Tucker, P. D. (2003). Handbook on teacher evaluation: Assessing and improving performance. Larchmont, NY: Eye On Education.

Stronge, J. Ed. (1997). Evaluating teaching: A guide to current thinking and best practice. Thousand Oaks: Corwin Press.

Stronge, J.H. (2007). *Qualities of Effective Teachers.* Alexandria, VA*:* ASCD.

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

Stronge, J. H., & Tucker, P. D. (2003). Handbook on teacher evaluation: Assessing and improving performance. Larchmont, NY: Eye On Education.

Sullivan, S. & Glanz, J. (2006). Building effective learning communities: Strategies for leadership, learning and collaboration. Thousand Oaks, CA: Corwin Press.

Sullivan, S. & Glanz, J. (2000). *Supervision that improves teaching: Strategies and techniques*. Thousand Oaks, CA: Corwin Press, Inc.

Sweeney, J., & Manatt, R. (1986). Teacher evaluation. In R. Berk (Ed.), *Performance assessment: Methods and applications.* Baltimore: John Hopkins University Press.

Taylor, F.W. (1911). *The Principles of Scientific Management.* New York, NY: Harper & Brothers.

Teacher Keys Effectiveness System Handbook (2013). Retrieved from: http://www.nctq.org/docs/Georgia_TKES_Handbook_7-18-2012.pdf

Thomas, D., Holdaway, E. & Ward, K. (2000). Policies and practices involved in the evaluation of school principals. *Journal of Personnel Evaluation in Education*, 14(3), 215–240.

Thorson, J.R., Miller, K.M. & Bellon, J.J. (1987). Instructional improvement through personnel Evaluation. *Educational Leadership* 35(5), 52-54.

Toch, T. (2008). Fixing teacher evaluation. *Educational Leadership, 66*(2), 32-37.

Triner, D., Greenberry, A., & Watkins, R. (1996). Training needs assessment: A contradiction in terms. Educational Technology, 36(6), 51-55.

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

U.S. Department of Education, (2014). Developing effective teacher evaluation systems: A conversation with Charlotte Danielson. Retrieved from: http://www.ed.gov/.

Varlas, L. (2009). Highly effective teachers: Defining, rewarding, supporting, and expanding their roles. Infobrief, 15(3).

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

Whitehead, A. N. (1929). *The Aims of Education and Other Essays.* New York: Macmillan.

Wikoff, M. (1994). The quality movement meets performance technology. *P & I, 33(8),* 42-45.

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement*, 45, 401-405.

Witters-Churchill, L. (1991). University preparation of school principals. *School Organization, 11(3),* 339-345.

Wright, P. & Geroy, G. (1992). Needs analysis theory and the effectiveness of large-scale government sponsored training programs: A case study. *Journal of Management Development, 11(5).*

Zepeda, S. J. & Ponticell, J. A. (1998). At cross-purposes: What do teachers need, want, and get from supervision? *Journal of Curriculum and Supervision,* 14(1), 68-87.

APPENDIX A: CONSENT FOR NONMEDICAL INTERVIEW STUDY

AAHRPP DOCUMENT #193

UNIVERSITY OF ALABAMA
HUMAN RESEARCH PROTECTION PROGRAM

CONSENT FORM FOR NONMEDICAL INTERVIEW STUDY

UNIVERSITY OF ALABAMA

Individual's Consent to be in a Research Study

You are being asked to be in a research study. This study is called "Implementing Effective Teacher Evaluations: A Needs Assessment for Future Principal Preparation". This study is being done by Mrs. Tricia Patterson. She is a doctoral student in the College of Education Leadership, Policy, and Technology Studies at the University of Alabama.

What is this study about?
The study will attempt to determine how principals perceive their level of preparedness to give beneficial feedback to teachers upon completing an evaluation. Specifically, the study will examine the feedback tools and training provided for implementation of the Georgia Teacher Keys Effectiveness System (TKES) in a way that will lead to teacher growth and, ultimately, student achievement. The study will conduct a program evaluation and produce a needs assessment that addresses further training needs of principals in the area of providing teachers effective feedback so future professional development can be planned.

Why is this study important—What good will the results do?
As schools work toward meeting the higher standards of new teacher evaluation instruments, finding key areas where principals need further training in using the instrument is important. Hence, the questions in this study address how building principals report they are prepared to give effective feedback while using a new evaluation instrument and, moreover, what the feelings toward the instrument are in contrast to others used, what the benefits and challenges of giving feedback might be, and what additional professional development principals report they need to become effective evaluators.

Why have I been asked to take part in this study?
You are a Principal with at least two years' experience and are currently implementing a new evaluation instrument. You verbally agreed to participation at the last Principal district meeting.

How many other people will be in this study?
The investigator hopes to interview 9 Principals from Cobb County School District.

What will I be asked to do in this study?
If you agree to be in this study, Mrs. Patterson will interview you at your school after or before school hours or at a place of your own choosing about your experiences with the evaluation instrument and your preparedness to give feedback. The interviewer would like to audio record the interview to be sure that all your words are captured accurately. However, if you do not want to be taped, simply mark the box at the end of this letter. A focus group interview with the eight other principals will take place as well at a convenient location during non-school hours. The focus group interview will be video recorded. A follow up individual interview will be conducted in the same manner as the first to ensure all information has been communicated.

How much time will I spend being in this study?
The interviews should last about 45-60 minutes each, depending on how much information about your experiences you choose to share.

Will being in this study cost me anything?

The only cost to you from this study is your time.

Will I be compensated for being in this study?
In appreciation of your time, I will bring you a coffee of your choosing during the first interview and refreshments will be provided during the focus group interview.

What are the risks (problems or dangers) from being this study?
The chief risk to you is that you may find the discussion of your experiences to be frustrating or stressful. You can control this possibility by not being in the study, by refusing to answer a particular question, or by not telling us things you find to be frustrating or stressful.

What are the benefits of being in this study?
There are no direct benefits to you unless you find it pleasant or helpful to describe your experiences with implementing a new evaluation instrument and eventually discovering future professional development needs of principals.

How will my privacy be protected?
You are free to decide where the interviews will take place so we can talk without being overheard and your name will be changed in the results.

How will my confidentiality be protected?
The only place where your name appears in connection with this study is on this informed consent. The consent forms will be kept in a locked file drawer in Mrs. Patterson's home, which is locked when she is not there. A name-number list will not be used so there is no way to link a consent form to an interview. When recording the interview, your name will not be used, so no one will know who you are on the tape. Once back in her office, the researcher will listen to the tape and type out the interview. At the end of the study the tapes will be destroyed. This should occur within one month of the interview. You may also refuse to be audiotaped, in which case the interviewer will take handwritten notes.

What are the alternatives to being in this study?
The only alternative is not to participate.

What are my rights as a participant?
Being in this study is totally voluntary. It is your free choice. You may choose not to be in it at all. If you start the study, you can stop at any time. Not participating or stopping participation will have no effect on your relationships with the Cobb County School District.

The University of Alabama Institutional Review Board is a committee that looks out for the ethical treatment of people in research studies. They may review the study records if they wish. This is to be sure that people in research studies are being treated fairly and that the study is being carried out as planned.

Who do I call if I have questions or problems?
If you have questions about this study right now, please ask them. If you have questions later on, please call Tricia Patterson at 404-786-7086. If you have questions or complaints about your rights as a research participant, call the Research Compliance Office at the University at 1-877-820-3066.

You may also ask questions, make a suggestion, or file complaints and concerns through the IRB Outreach Website at http://osp.ua.edu/site/PRCO_Welcome.html . After you participate, you are encouraged to complete the survey for research participants which is online there. You may also e-mail us at participantoutreach@bama.ua.edu.

I have read this consent form. I have had a chance to ask questions.

_____

Signature of Research Participant Date

_____

Signature of Investigator Date

____ Check this box if you AGREE to having the interviews audio and video taped

APPENDIX B: INTERVIEW PROTOCOL

Interview Protocol

Time of Interview:

Date:

Place:

Interviewer: Tricia Patterson

Interviewee:

Position of Interviewee:

Questions:

1. What experiences prepared you to implement TKES?

2. How do you differentiate the feedback you give to teachers?

3. How is the feedback given?

4. What other instruments have you used? How to they compare to TKES?

5. What additional professional development do you feel you need to use the instrument effectively?

Analytical Notes:

APPENDIX C: TKES HANDBOOK

TKES HANDBOOK

APPENDIX D: IRB APPROVAL LETTER

Office for Research

Institutional Review Board for the Protection of Human Subjects

THE UNIVERSITY OF ALABAMA RESEARCH

January 14, 2015

Tricia Patterson
ELPTS
College of Education
Box 870302

Re: IRB # 15-OR-009, "Implementing Effective Teacher Evaluations: A Needs Assessment for Future Principal Preparation"

Dear Ms. Patterson:

The University of Alabama Institutional Review Board has granted approval for your proposed research.

Your application has been given expedited approval according to 45 CFR part 46. Approval has been given under expedited review category 7 as outlined below:

(7) Research on individual or group characteristics or behavior (including but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Your application will expire on January 13, 2016. If your research will continue beyond this date, please complete the relevant portions of the IRB Renewal Application. If you wish to modify the application, please complete the Modification of an Approved Protocol form. Changes in this study cannot be initiated without IRB approval, except when necessary to eliminate apparent immediate hazards to participants. When the study closes, please complete the Request for Study Closure form.

Please use reproductions of the IRB approved stamped consent forms to obtain consent from your participants.

Should you need to submit any further correspondence regarding this proposal, please include the above application number.

Good luck with your research.

Sincerely,

Carpantato L. Hayes, MSW, LCSW, CIP
Director of Research Compliance Officer
Office for Research Compliance
The University of Alabama

358 Rose Administration Building
Box 870127
Tuscaloosa, Alabama 35487-0127
(205) 348-5960
(205) 348-7189
(205) 929-1266

5

APPENDIX E: COBB COUNTY SCHOOL DISTRICT RESEARCH APPROVAL

COBB COUNTY
SCHOOL DISTRICT

Empowering Dreams for the Future

P.O. Box 1088
Marietta, GA 30061
Telephone: (770) 426-3300
www.cobbk12.org

February 17, 2015

Ms. Tricia Patterson
1411 Lafayette Lane
Marietta, GA 30068

Dear Ms. Patterson:

Your research project titled, **Implementing Effective Teacher Evaluations: A Needs Assessment for Future Principal Preparation**, has been approved. Listed below are the schools where approval to conduct the research is complete. Please work with the school administrator to schedule administration of instruments or conduct interviews.

School
Dickerson Middle School
Hightower Trail Middle School
Mt. Bethel Elementary
Murdock Elementary
Pope High School
Sedalia Park Elementary
Simpson Middle School
South Cobb High School
Wheeler High School

Should modifications or changes in research procedures become necessary during the research project, changes must be submitted in writing to the Office of Accountability & Research prior to implementation. At the conclusion of your research project, you are expected to submit a copy of your results to this office. Results cannot reference the Cobb County School District or any District schools or departments.

Research files are not considered complete until results are received. If you have any questions regarding the process, contact our office at 770-420-4918.

Sincerely,

Dr. Ehsan Kattoula
Director of Accountability & Research

nf