Teacher Perceptions of the Georgia Teacher Evaluation Instrument and its Impact on Teacher Professional Growth

A Dissertation submitted
to the Graduate School
Valdosta State University

in partial fulfillment of requirements
for the degree of

DOCTOR OF EDUCATION

in Educational Leadership

in the Department of Leadership, Technology, and Workforce Development of the James L. and Dorothy H. Dewar College of Education and Human Services

April 06, 2023

TONYA LESURE

EdS, Valdosta State University, 2009
MEd, Troy University 2006
BSEd, Valdosta State University, 2000
ASEd, South Georgia College, 1998

This dissertation, "Teacher Perceptions of the Georgia Teacher Evaluation Instrument and its Impact on Teacher Professional Growth" by Tonya Ann Lesure, is approved by:

Dissertation Committee Chair

James L. Pate, Ph.D.
Professor of Curriculum, Leadership, and Workforce Development

Committee Members

Lantry L. Brockmeier, Ph.D., Researcher
Professor of Curriculum, Leadership, and Workforce Development

Jahn D. Lairsey, Ed.D
Assistant Professor of Teacher Education

Associate Provost For Graduate Studies and Research

Becky K. da Cruz, Ph.D., J.D.
Professor of Criminal Justice

Defense Date April 6, 2023

FAIR USE

DUPLICATION

I authorize the Head of Interlibrary Loan or the Head of Archives at the Odum Library at Valdosta State University to arrange for duplication of this dissertation for educational or scholarly purposes when so requested by a library user. The duplication shall be at the user's expense.

Signature 

I refuse permission for this dissertation to be duplicated in whole or in part.

Signature 

ABSTRACT

Due to the required amount of time, financial resources, and the potential effects of the evaluation on teacher performance and student learning outcomes, teacher perceptions and experiences with the Georgia TKES need to be examined. This quantitative non-experimental design study aimed to examine the perceptions and experiences of teachers with the Georgia TKES evaluation, considering factors such as time, financial resources, and potential effects on teacher performance and student learning outcomes. The purpose was to determine whether teachers have buy-in for a standards-based evaluation instrument that could potentially change their behavior. The study also explored whether there were any differences in perceptions based on teacher characteristics such as experience, grade level taught, and level of education. Data was collected from 347 respondents using the Teacher Survey of the Georgia TKES. The findings of the study indicated that gender had a significant effect on both perceived accuracy and perceived level of creativity, while grade level had a significant effect on perceived influence and perceived level of creativity. The interaction between the number of years certified as a teacher and educational level also had a significant effect on perceived level of creativity. Additionally, there were significant differences by the ratings for the TKES, TAPS, and SGP training on all three dependent variables. This study has the potential to provide new insights and findings that were not previously considered, which could lead to further opportunities for improving teacher performance.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not be possible without the support of my committee chair, Dr. James Pate. There were many times I did not see myself completing this journey of obtaining my doctoral degree. Dr. Pate you gave me my second wind when I was at the point of throwing in the towel. Thank you Dr. Pate for supporting and encouraging me to see the value and opportunities that would be so worth me completing my doctoral degree. My gratitude extends to Dr. Lantry Brockmeier and Dr. John Lairsey for taking time out of their busy schedules to read my dissertation and provide much appreciated feedback.

I would also like to thank my forever-encouraging family. Thank you John, my dear husband for telling me constantly "you are not going to quit. Tonya you have come to far and you are going to finish". My Uncle Marvin McIver and my father, John McIver for being my bank. Thanks Mom, Elizabeth Pryce for making me want more out of life. Thank you to my brothers, Eric McIver, John McIver, and Marshall Robinson for being so proud of me and encouraging me. Thank you to my favorite sister, Tracy McIver for being proud of me and supporting me. Thank you to my staff at Coffee County Educational Academy for supporting me and listening to my woes. Thank you God for giving me the strength to continue this journey because you had me all the way. And thank you to my children Arianna, Amyra, Asher, Jayla, and Sasha LeSure for being so proud of your momma. Thank you to my special nice Zy'reyia Canady for telling me "Auntie you gotta keep going and believe in yourself". I dedicate this dissertation to you all.

Chapter I

INTRODUCTION

Teacher evaluation systems have been used for decades to measure teacher quality (Marzano, 2012). Unfortunately, some teacher evaluation systems have failed for at least two reasons: They have not differentiated between effective and ineffective teaching, and they have not contributed to the development of highly skilled teachers (Kraft & Gilmour, 2017; Polikoff & Porter, 2014; U.S. Department of Education, 2009; Weisberg et al., 2009). A result of this problem was that federal legislation began to focus on U.S. educators by encouraging school districts to reform teacher evaluation instruments to address areas vital to ensuring student academic achievement.

The focus of improving the American educational system began in 1957 when the Soviet Union accomplished its mission of launching Sputnik I into outer space. The Soviet Union was making progress by producing scientists and engineers, which created the image of America lagging behind in technology after World War II. American policymakers focused on improving academic achievement and educational standards to ensure the American educational system could produce students with the necessary skills and knowledge to compete in a global economy.

The launching of Sputnik further ignited the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983), proposed by President Ronald Reagan. This publication drew the nation's focus to teacher accountability, which became a major focus of the reformation of the U.S. educational system. The publication identified areas of concern with American public education. The report noted the need to increase student academic achievement in all U.S. public schools by initiating

educational reform. Among other things, the report stated teachers were not qualified to teach subjects they were being asked to teach.

Years later, political policymakers and President George Bush remained concerned about American public schools lagging behind other nations academically. In 1989, as a response to the concern, a coalition of state governors led by President Bush proposed Goals 2000 as a solution to this problem. The educational goals of the program were set to be accomplished by the Year 2000. The program established a framework for carrying out Goals 2000 and provided rewards for states complying with the goal requirements (Mathison & Ross, 2013). By 1994, Goals 2000 became an official program with eight national education goals.

Reports before the deadline of 2000 indicated teacher quality had not improved (Hanlin, 2014). Goals 2000 stated that teachers would be provided with professional development programs to improve their professional skills to impact instruction and academic needs of all U.S. students. These efforts appeared to have been unsuccessful, and student achievement was not significantly impacted (Hanlin, 2014).

Goals 2000 was succeeded by the No Child Left Behind Act (NCLB) signed into law by President George W. Bush in early 2002. This law reauthorized the Elementary and Secondary Education Act, which was initially passed by Congress in 1965. The new law ensured that all students would receive a quality education. NCLB (2002) mandated state systemic changes for educational systems and held schools accountable for student performance by requiring schools to implement accountability systems that utilized student test-score data from statewide math and reading assessments. The law further required states to disaggregate student test-score data for subgroups based on ethnicity,

economic status, limited proficiency in English, and special education (Domina, 2014). States under NCLB were required to place sanctions on school systems that failed to meet Adequate Yearly Progress requirements for exam proficiency. These requirements placed pressure on school systems to address teacher quality, which could impact student academic achievement. NCLB was initiated to ensure that all U.S. students were taught by highly effective teachers.

Eight years later, Race to the Top was initiated. This initiative rewarded school systems for implementing school reform that could increase student academic achievement. These initiatives led to a greater focus on student academic achievement and the impact of teacher quality (Weems & Rogers, 2010). The Race to the Top initiative provided states with grants for reforming education to focus more on student performance outcomes (Aguilar & Richerme, 2014). Race to the Top drove states to require rigorous educational standards, improve the teaching and learning process, utilize classroom data, and implement new strategies for struggling schools (Gagnon et al., 2017). The reformation of education left educational stakeholders with the responsibility of determining whether or not teacher evaluations were impacting teacher practices in the school setting (Callahan & Sadeghi, 2015). School districts receiving funding from the program grants aligned their teacher evaluation systems to Race to the Top requirements. Teacher evaluations were used to assess whether or not instructional strategies teachers implement in the classroom address student learning needs and ensure student academic growth (Stronge, 2006).

The demands of globalization and global competition also serve as driving forces for school leaders, teachers, and policymakers to reform education (Spring, 2014). Public

school educators are held accountable for providing educational opportunities that will produce students capable of competing in the global economy (Conceição, 2016). School district administrators and principals who wish to improve effectiveness often use evaluation systems to evaluate and improve teacher effectiveness (Harris, Ingle, & Rutledge, 2014).

Statement of the Problem

Although stakeholders in the educational system have worked diligently to reform teacher evaluation to impact teacher performance, research is scarce on teachers' perceptions of how specific evaluation instruments improve instructional practice to increase student learning outcomes. Examining the attitudes and perceptions of the teachers involved in the process is vital to the educational reform movement (Shakman et al., 2012).

Evaluations are an important strategy to assess teacher quality; therefore, there is a growing consensus about the need for evaluation systems that could yield higher quality information to improve teacher performance (Taylor & Tyler, 2012). Policymakers began to enforce initiatives to establish American schools with highly qualified competent teachers to impact student learning outcomes. The Georgia teacher-evaluation instrument is the Georgia Teacher Keys Effectiveness System (TKES), which focuses on specific indicators of classroom practices used by teachers. Due to the required amount of time, financial resources, and the potential effects of the evaluation on teacher performance and student learning outcomes, teacher perceptions and experiences with the Georgia TKES need to be examined. Teacher perceptions of the impact of the Georgia TKES are unknown, and this study is designed to determine whether teachers perceive the

evaluation instrument to improve professional growth, instructional strategies, and student learning outcomes.

Purpose of the Study

The purpose of this study is to identify teacher perceptions of the Georgia TKES evaluation. Specifically, the study is designed to identify whether teachers perceive the Georgia TKES to improve their instructional practice, including any differences in perceptions by teacher characteristics of experience, grade level taught, and level of education.

The Georgia TKES was fully implemented during the 2012-2013 school year. This evaluation instrument is relatively new for teachers and administrators. Research is needed on perceived effectiveness of the TKES evaluation tool. This study can provide educational stakeholders with data on the effectiveness of the teacher-evaluation tool being used in Georgia. Results from the study can provide information on teachers' perception of the impact of the evaluation tool on teacher professional growth and instructional planning methods. This study also may identify professional learning opportunities that could address areas of concern for instructional planning.

Research Questions

There are six independent variables for this study. Respondents will be analyzed in groups by gender (male or female), race and ethnicity (Asian, Black or African American, Latino/Hispanic, White, Native Hawaiian or other Pacific Islander, and American Indian or Alaska Native), grade level (pre-kindergarten to 12th grade), education attainment level (bachelor's degree, master's degree, Educational Specialist degree, and doctorate), years of teaching experience (1–2 years, 3–5 years, 6–10 years,

and more than 10 years), and subject taught (mathematics, science, language arts, and electives). The dependent variables for this study are changes in teaching practices; agreement with assessment score; allowed teacher creativity; and effectiveness of training in TKES, Teacher Assessment Performance Standards (TAPS), and Student Growth Performance (SGP).

1. Is there a significant difference by gender and race or ethnicity on selected dependent variables? (factorial ANOVA)
    a. Is there a significant difference by gender and race or ethnicity on teachers' perceived influence of the TKES on teaching practices?
    b. Is there a significant difference by gender and race or ethnicity on teachers' perceived accuracy of the TKES assessment?
    c. Is there a significant difference by gender and race or ethnicity on the perceived level of creativity allowed by the TKES?
2. Is there a significant difference by grade level taught and level of education on selected dependent variables? (factorial ANOVA)
    a. Is there a significant difference by grade level taught and level of education on teachers' perceived influence of the TKES on teaching practices?
    b. Is there a significant difference by grade level taught and level of education on teachers' perceived accuracy of the TKES assessment?
    c. Is there a significant difference by grade level taught and level of education on the perceived level of creativity allowed by the TKES?

3. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on selected dependent variables? (factorial ANOVA)
    a. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on teachers' perceived influence of the TKES on teaching practices?
    b. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on teachers' perceived accuracy of the TKES assessment?
    c. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on the perceived level of creativity allowed by the TKES?
4. Is there a significant difference by content area taught on the selected dependent variables? (Simple ANOVA)
    a. Is there a significant difference by content area taught on teachers' perceived influence of the TKES on teaching practices?
    b. Is there a significant difference by content area taught on teachers' perceived accuracy of the TKES assessment?
    c. Is there a significant difference by content area taught on teachers' perceived level of creativity allowed by the TKES?
5. Is there a significant difference between levels taught of the effectiveness of training on the TKES on teaching practices?

6. Is there a significant difference between levels taught of the effectiveness of training on the TAPS on teaching practices?
7. Is there a significant difference between levels taught of the effectiveness of training on the SGP on teaching practices?

Summary of Methodology

The purpose of this descriptive, quantitative study is to evaluate how, if at all, teachers perceive that the Georgia TKES evaluation tool impacts their instructional strategies and professional growth. Data will be gathered through a survey using a 5-point scale. Results will be analyzed based on respondents' gender, race and ethnicity, grade level taught, education attainment level, years of teaching experience, and subject taught. Three survey questions assess teachers' perceptions of the quality of training received on the TKES components, ten survey questions assess perceptions of extent of influence of the TKES on teaching practices, ten survey questions assess the perceived agreement with accuracy of TKES scoring, and ten survey questions assess teacher's perceptions of the extent to which the TKES allows teacher creativity (see Appendix A).

Significance of the Study

The study is designed to identify the perceptions teachers have about the Georgia TKES evaluation instrument. The TKES is used to evaluate all teachers in the state, and research is needed to determine whether feedback based on the TKES improves teaching practices. Teacher behaviors can be identified that may improve student academic performance in selected academic courses (Master, 2013). It is important to understand whether teachers share buy-in of a standards-based evaluation instrument that could change teacher behavior. The process of changing teacher behavior through the use of the

TKES instrument could be hindered if teachers do not believe the process is a means to improve student academic achievement. This study will examine teacher perceptions of the Georgia TKES and determine whether teacher characteristics of experience, grade level taught, and level of education impact those perceptions.

Conceptual Framework of the Study

The conceptual framework for this study is based on the idea that teacher evaluation systems can be used as a mechanism to improve a teacher's professionalism by providing (a) useful feedback and support that address specific teachers' needs at their current level of practice and (b) relevant professional growth opportunities. Danielson's (1996) Framework for Teaching provides a basis for this study. Danielson (1996) identified and classified the behaviors of teachers who improve student achievement into 22 indicators and four domains. The four domains of teacher responsibilities are (a) planning and preparation, (b) the classroom environment, (c) instruction, and (d) professional responsibilities. Table 1 presents the four domains and relevant competencies.

Table 1

Domains and Competencies of Danielson's Framework for Teaching

Domain	Competencies
Planning and preparation	Knowledge of content and pedagogy Knowledge of students Instructional goals for diverse students Knowledge of resources Coherent instruction Assessment of student learning
Classroom environment	Creating an environment of mutual respect and rapport Establishing a culture for learning Managing classroom procedures Managing student behavior Organizing physical space
Instruction	Clear and accurate communication Questioning and discussion techniques Student engagement in learning Feedback to students Flexibility and responsiveness
Professional responsibilities	Reflecting Maintaining accurate records Communicating with families Contributing to the school and district Growing and developing professionally Showing professionalism through advocacy and decision-making

Note. Source: *Enhancing Professional Practice: A Framework for Teaching,* by C. Danielson, 1996, Arlington, VA: Association for Supervision and Curriculum Development, 1996.

Danielson's (1996) theoretical framework is appropriate for the current study for several reasons. First, Danielson (1996) explicitly communicated standards and expectations needed for teachers to be effective. Second, the framework provides specific guidelines about what a teacher does while providing instruction. Third, Danielson's framework identifies teacher behaviors that improve student achievement. The Figure is a graphic of how Danielson's Framework for Teaching applies to teacher evaluation systems.

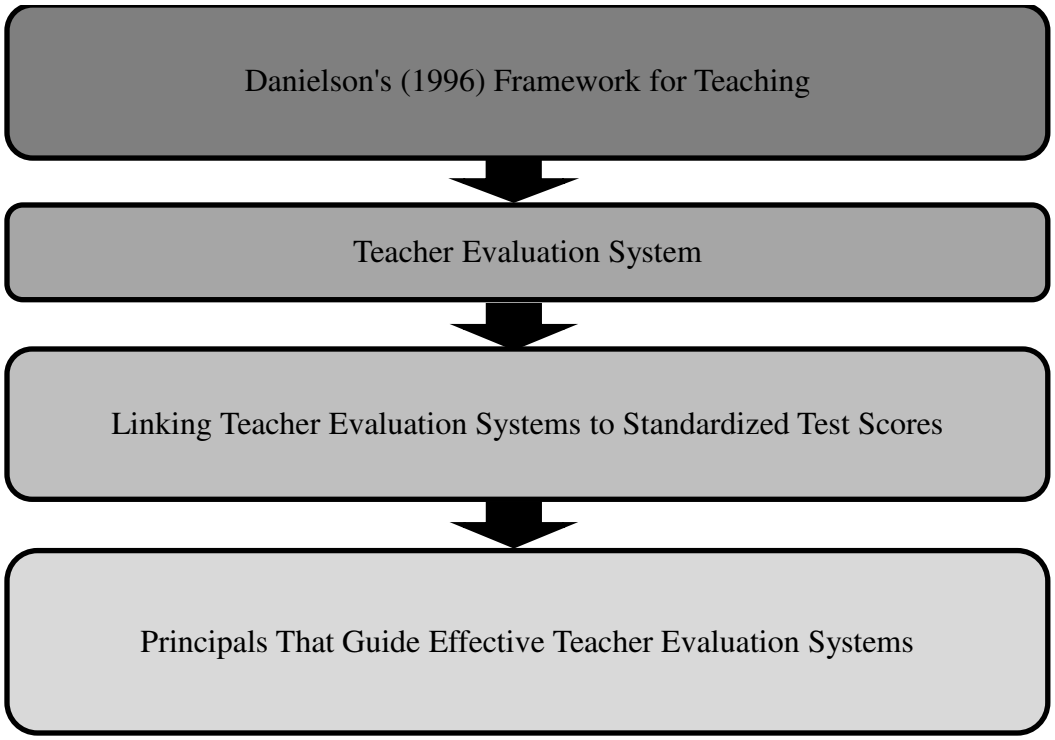

Figure. How Danielson's (1996) Framework for teaching applies to teacher evaluation systems.

Limitations

The design of this study is self-report; the anonymity of responses should encourage teachers to respond honestly. The study will compare survey results among a group of teachers and will be limited by population and sample size. The study could be further limited by a low participation rate. The methodology of the study does not allow for determining association and causation.

Definition of Terms

Formative evaluation serves the purpose of improving teaching methods by providing teachers with feedback (Shinkfield & Stufflebeam, 2012).

The No Child Left Behind Act (NCLB) was a federal law passed by Congress that increased education accountability in public schools in order to increase student academic achievement (NCLB, 2002).

Race to the Top is a national educational grant that rewarded states for implementing educational reforms (U.S. Department of Education, 2009).

Summative evaluation uses various data to inform personnel decisions, such as a teacher's worth to a school (Shinkfield & Stufflebeam, 2012).

Teacher effectiveness refers to the ability of the teacher to increase student learning by implementing effective learning strategies (Georgia Department of Education, 2018).

Georgia's Teacher Keys Effectiveness System (TKES) is a teacher-evaluation system designed to provide teachers feedback to increase teacher effectiveness and thus student learning (Georgia Department of Education, 2018). The system involves observations of three components: Teacher Assessment on Performance Standards, Professional Growth, and Student Growth (Georgia Department of Education, 2018).

Organization of the Study

This dissertation consists of five chapters. An introduction to this descriptive quantitative study of teacher perceptions of the Georgia TKES and its impact on instructional planning is presented in Chapter 1. Chapter 1 also includes an introduction to the study, statement of the problem, purpose of the study, research questions, significance of the study, conceptual framework, a short description of the methodology, and definition of terms. Chapter 2 will provide a review and analysis of the literature relevant to this study. Topics include a brief history of teacher evaluations, reforming teacher evaluation, effective teaching, teacher evaluations and professional development, teacher evaluation and student academic achievement, and the Georgia TKES. Chapter 3 will discuss the research design, instrumentation used to collect data, the population, and

the methods used to collect and analyze data. Chapter 4 will describe the results. Chapter 5 will include a discussion of the results of this descriptive quantitative study, including suggestions and recommendations for school districts and further research in this area.

Chapter II

Literature Review

Although teacher evaluation systems have been used for decades to measure effectiveness, they have not necessarily contributed to the development of highly skilled teachers (Kraft & Gilmour, 2017; Polikoff & Porter, 2014). However, teacher evaluations systems are continuously used in school districts throughout the nation to evaluate the effectiveness of teachers (Kraft & Gilmour, 2017; Polikoff & Porter, 2014). Croft et al. (2015) reported that federal reform initiatives such as NCLB, Race to the Top, and the Common Core State Standards were adopted to improve the academic performance of all K-12 students, especially those who are from racially diverse and low-socioeconomic backgrounds. Although such reform initiatives have been implemented to improve student academic achievement, they have not been favorable regarding teachers, who increasingly have been held accountable for improving learner outcomes as measured by high-stakes testing and teacher evaluations (Croft et al., 2015). Croft et al. described the combination of teacher evaluations and high-stakes testing as a "perfect storm that is eroding the bedrock of public education" (p. 70) and asserted that consequently, student performance is at an all-time low. The researchers suggested educational reform initiatives have been grounded in politics, profit for test manufacturers, and evaluation systems, which have done very little to improve teaching and learning. Ravitch (2016) added that current educational reform initiatives put more emphasis on teacher accountability than on student learning. Additionally, accountability measures have lessened the focus of evaluating student learning and increased focus on teacher performance and quality.

The Georgia TKES was incorporated in 2012-2013 (Georgia Department of Education, 2013). Teachers are assessed formatively and summatively on annual evaluations using TKES (Georgia Department of Education, 2018). Teachers self-assess prior to the beginning of each school year. Throughout the year, teachers are assessed formatively through observations and documentation, including a midyear conference. At the end of the year, teachers receive a summative performance evaluation (Georgia Department of Education, 2018). However, teacher perceptions of the impact of the Georgia TKES are not known.

The History of Teacher Evaluations

According to Jewel (2015), prior to 1965, methods for evaluating teachers were primarily left up to community leaders where local schools were established. For example, during the early period of American colonialism, teachers were evaluated by community and church leaders, who commonly held positions on a governing body called the trustee board. Unlike today, teachers were not evaluated based upon student performance, but rather on how well they could maintain order in the classroom, how well the physical premise of the campus was kept, and how well they were able to implement the political and religious views of the community. Rather than being trained and provided with feedback to help improve teaching and learning, teachers who did not meet expectations were typically dismissed by the trustee board.

Taylor and Tyler (2012) added that schools remained under local control until the Industrial Revolution of 1820–1860. During the Industrial Revolution, U.S. teachers had limited resources, held education that was just above that of their students, and relied on textbook recitation as the primary method of learning (Taylor & Tyler, 2012). As the

number of schools increased, the curriculum began to shift from religious subject matter to more academic subject matter, which required better educated teachers, who were called *principal teachers* (Taylor & Tyler, 2012). An influx of immigrant children led to an increase in the number of and the different levels of principal teachers, whose titles transitioned to administrators. As the number of administrators began to increase, so did their tasks, which included conducting formal teacher evaluations.

Since 1930, teacher evaluation have been rooted in the Hawthorne studies, which proposed that workplace efficiency improves when employees are observed and provided with feedback about their performance (Taylor & Tyler, 2012). The Hawthorne-based teacher evaluation model involved evaluations one to three times a year, mostly based on a checklist of observable classroom behaviors. Results from the evaluations were kept in a file in human resources departments and used for documenting years of teaching experience and for hiring and terminating teachers. However, the teacher evaluation did not focus on how the results could be used to improve teacher effectiveness (Taylor & Tyler, 2012).

Croft et al. (2015) added that during the late 1950s, the federal government increased its role in public education, especially after Russia launched Sputnik in 1957. Technical advances in other nations led to concern about U.S. student performance. Consequently, the federal government began to provide federal funding to improve student performance in reading, math, and science and set the stage for eventual federal involvement in K-12 education.

The Hawthorne model was used until the 1960s, and teacher evaluations continued to be used to determine the best way to improve instruction (Jewel, 2015).

Jewel (2015) added that standardized test scores began to be used to evaluate instructional quality. The 1965 Elementary and Secondary Education Act began to impact how teachers were evaluated (Gamson et al., 2015). The law's purpose was to improve educational outcomes among students from racially diverse and low-socioeconomic backgrounds (Gamson et al., 2015). The legislation was also strongly connected to the Civil Rights movement and designed to give underprivileged children access to a quality education. Teacher evaluations were a way to ensure the populace that underserved students received instruction from effective teachers.

Donaldson and Papay (2014) wrote that beginning in the 1970s, teacher evaluations were incorporated as a result of educational reform and used for teacher accountability relating to student performance. The model became known clinical supervision, which was the process whereby the school principal conducted pre-observations, observations, and post-observations of teachers in the classroom and met with teachers to discuss their professional practice in order to improve overall teacher quality (Donaldson & Papay, 2014). Compared to earlier models, during clinical supervision, principals played a more significant role. By 1980, clinical supervision was widely used to evaluate new teachers.

In 1979, the U.S. Department of Education was founded, which provided the federal government with further engagement in K-12 education. However, conservative legislators viewed the U.S. Department of Education as interference in state and local issues, including teacher evaluations (Ravitch, 2016).

Shaw (2016) wrote that during the 1980s, teacher evaluations were influenced by Madeline Hunter's model of mastery learning. Hunter developed a seven-step model that

included scripts teachers followed, along with observations and feedback from principals. Ultimately, teachers were evaluated based on their ability to follow the model. The assumption was that student achievement would increase when the model was properly implemented.

After 1983, teacher evaluations became a major component of federal educational reform. The federal government's concern with teacher evaluations emerged as a result of *A Nation at Risk*, a report published by former President Ronald Reagan's National Commission on Excellence in Education (1983). The report informed the nation that U.S. public school students' academic performance was inferior to that of students from other industrialized nations because of the nation's failing educational system. The document warned that the America's inferior educational system would eventually lead to an economic crisis and a threat to national security. The National Commission on Excellence in Education recommended more rigorous standards and more professional development to improve teacher effectiveness. In 1988, the Personnel Evaluation Standards published by the Joint Committee of Standards set forth evaluation models that focused on teachers' performance standards (Bergsmann et al., 2015).

Stecher and Garet (2014) reported the RAND Corporation conducted a study during the 1980s of 32 school districts throughout America. The purpose of the study was to identify effective teacher evaluation practices. The teacher participants, who were surveyed during the study, indicated the need for evaluation systems conducted by competent principals who would review teachers' performance accurately and provide meaningful feedback. As a result of the study, the RAND Corporation suggested districts provide resources to create evaluation systems addressing specific problems of schools,

evaluator training, and the incorporation of feedback received during observations. Both the RAND Study (as cited in Stecher & Garet, 2014) and *A Nation at Risk* (National Commission on Excellence in Education, 1983) documented the need for effective teaching practices, which in turn put pressure on state departments of education to improve teacher quality, as measured by student performance on standardized tests. Other teacher evaluation models used during the 1980s focused on career development for teachers based on their age, their years of experience, and developmental level. Such evaluation methods continued through the 1990s.

The 1990s brought about increased school accountability at the federal level and mandates to increase teaching and academic standards, often attached to grant money (Jewel, 2015). Teacher evaluations began to reflect the work offered by Charlotte Danielson in 1996. Danielson's teacher-evaluation model, which was more comprehensive than previous models, focused on 76 specific areas of teacher performance, which were classified into four domains: (a) planning and preparation, (b) the classroom environment, (c) instruction and delivery, and (d) professional development. The Danielson model also included an extensive ranking system for teachers: unsatisfactory, basic, proficient, and distinguished (Danielson, 2016).

Political influence and the initiative to design an evaluation tool that is unbiased and objective lies at the foundation of the reformation of teacher evaluation (Cochran-Smith et al., 2013). For example, in 2002, former President George W. Bush signed into law NCLB. The law proposed that every student in K-12 schools across the United States would be proficient in reading and in math by 2020 (Cochran-Smith et al., 2013). According to NCLB (2002), each K-12 teacher was to be deemed highly qualified. To be

highly qualified, the teacher must have held a bachelor's degree, a state-issued teaching certificate or license, and proof of qualification to teach a specific content area through college course hours or standardized tests such as the Praxis. The law also sought to provide equal educational opportunities to all K-12 students by tracking standardized-test performance by gender, race, socioeconomic status, and English language proficiency. NCLB held states, local educational agencies, and teachers accountable for student performance.

NCLB yielded mixed results for student outcomes; consequently, the federal government began to highlight the need for more teacher accountability and for changes to teacher evaluation systems (Dee & Jacob, 2011; Deming et al., 2016). As a result, in 2011, President Barack Obama waived the mandates set forth by NCLB and set new standards for teacher evaluations and student achievement (McGuinn, 2016). Race to the Top gave monetary incentives to educational systems that created evaluation processes to improve teacher and principal effectiveness based on student performance on state standardized assessments. Under the Race to the Top Initiative, state agencies were encouraged to design their own teacher evaluations. Teachers whose students consistently earned high test scores received financial rewards. This met opposition from educators, who were limited by the law to use collective bargaining to set evaluation practices. Additionally, 20 states integrated teacher evaluations in which student test scores counted for 30–50% of the teacher's rating. The method, called *value added*, also connected teacher evaluations to consequences, such as dismissal (Goldhaber, 2015).

In 2013, state evaluation systems varied widely. The variations in teacher evaluations were mostly due to the impact of Race to the Top, which encouraged school

districts to transition from older teacher evaluation models to models that evaluated teachers based on student achievement (McGuinn, 2014). Likewise, the Measures of Effective Teaching Project, a study funded by the Bill & Melinda Gates Foundation, also impacted teacher evaluation methods (Kane et al., 2012). The purpose of the Measures of Effective Teaching Project was to investigate ways to identify and foster more effective teachers. For 3 years, researchers from Bill & Melinda Gates Foundation studied the teaching practices of roughly 3,000 K-12 teachers who taught math, English/language arts, algebra, and biology and who were employed in districts in Charlotte-Mecklenburg, Dallas, Denver, Hillsborough, New York City, Memphis, and Pittsburgh (Bill & Melinda Gates Foundation, 2012; Kane et al., 2012). To collect data for the study, the researchers conducted classroom observations, surveyed students, and measured achievement gains. As a result of the study, Kane et al. (2012) reported suggestions concerning teacher evaluation systems. First, they suggested teacher evaluations include a measure of student achievement, or value-added, as an indication of teacher effectiveness. However, the report did not indicate how much weight should be attached to value-added measures. Second, the teacher evaluation system should weigh benchmarks against student achievement. Third, the evaluation system should include feedback on specific aspects of a teacher's practice to support teacher growth and development (Kane et al., 2012).

By late 2015, 42 states had waivers from the NCLB requirements, but many districts continued to use the value-added assessment models as a part of teacher evaluations (Akers, 2016). These 43 states transitioned to new student testing systems aligned with college- and career-readiness standards and aligned with the implementation of teacher effectiveness policy (Doherty & Jacobs, 2015). The states that have adopted

the new evaluation systems are utilizing a value-added model, which is used to estimate the effects of individual teachers on student learning outcomes (Koedel, Mihaly, & Rockoff, 2015).

Georgia's teacher evaluation system is similar to those found in 43 states of the United States (Doherty & Jacobs, 2015). Good and Lavigne (2017) described much educational research and confusion among researchers, the public, and politicians concerning the pros and cons of the value-added model and teachers' effects on student achievement. However, little, if any, qualitative research exists on the aspects of the pedagogical influences of the value-added model of evaluation from the perspectives of teachers after 3 years of evaluation implementation (Bogart, 2013; Taylor & Tyler, 2012).

Effective Teaching

Teacher effectiveness has become the focal point of reforming the teaching profession. Policymakers are beginning to enforce initiatives to provide students with teachers who can provide high-quality instruction that will impact student learning outcomes (Welsh, 2011). School district leaders and stakeholders recognized the importance of designing a teacher evaluation system that produced valid results when measuring a teacher's effectiveness (Darling-Hammond, 2014). The validity of an evaluation instrument lies in the tool accurately depicting an individual teacher's practices that contribute to student learning outcomes (Goe et al., 2008).

Current studies indicate that teacher effectiveness has long-term effects on student learning outcomes (Chetty et al., 2014; Jackson, 2012). Despite the prevailing theme in the literature that teacher effectiveness has a significant impact on student learning

outcomes, reformed teacher evaluations have missed the mark by not being able to measure the difference between effective and ineffective teachers (Marzano, 2012). Characteristics of an effective teacher are difficult to gauge (Rockoff, 2004). There is no consensus among educators about how to measure effective teaching.

Rockoff et al. (2011) revealed several themes that define effective teaching practices: high expectations for student learning, the use of higher order thinking and questioning, instructional differentiation, and a classroom environment conducive to learning. Coe et al. (2014) indicated teachers communicate high expectations for student academic achievement in an effective classroom in two ways, by ensuring students (a) are aware of learning goals and (b) understand how the learning goals will be assessed. Students should be given tasks developed with appropriate rigor. High expectations in effective learning environments include classroom management of student behavior. Coe et al. suggested that effective teachers have fewer classroom disruptions. Classroom management ensures instruction and learning can occur and are the focus (Wong, Wong, Rogers, & Brooks, 2012). In an effective teacher's classroom, students are held accountable for successfully completing tasks and demonstrating appropriate behavior that will not disrupt the learning environment. Teachers' high expectations for student academic success can lead to the effectiveness of teachers and enhanced student learning outcomes (Coe et al., 2014).

Higher order thinking and questioning are associated with effective teaching (Tofade et al., 2013). Effective teachers motivate students to think on higher levels instead of only recalling and memorizing information taught. Ramnarain (2011) suggested teachers who ask questions requiring students to synthesize and apply

knowledge and skills at all stages of learning help students to attain greater independence from the teacher. Effective teaching requires students to have confidence to learn on their own instead of depending on the teacher. Designing differentiated instructional strategies to meet the learning abilities, learning styles, and personalities of students has been proven to be successful for effective teaching. Reis et al. (2011) found that differentiated instruction strategies implemented in a middle school reading program for student's oral reading fluency and comprehension were more effective on student learning outcomes than teaching to the whole group of students.

A lack of consensus exists about whether to access teacher effectiveness using teaching outcomes, teacher perceptions, the teaching methods, or a mixture of these (Stigler & Hiebert, 1999). Stigler and Hiebert (1999) stated that teaching, not teachers, is an essential factor, which suggest instructional systems used to teach students are more influential on student academic achievement than are teaching strategies, professional development, assessments, curriculum, and other qualities related to the teaching profession. Effectiveness is a challenging idea to measure when considering the complex factors of education and the varied conditions in which educators work.

Reforming Teacher Evaluation

Historically, the link between teacher effectiveness, teacher influence, teacher evaluation, and student achievement has been confusing (Stronge, 2018). Teacher evaluations were commonly used to make decisions about employment. Evaluations provided teachers with feedback and guidance that could be used to improve a teacher's professional performance. The foundation of teacher evaluations focused on classroom observations and the act of teaching (Good & Lavigne, 2017). This focus on a small

segment of evaluating teachers limited the view of the quality of a teacher's instruction. Weisberg et al. (2009) stated that teacher effectiveness is a critical factor in increasing student academic outcomes. Weisberg et al. stated the evaluation systems used did not clearly distinguish teachers' instructional quality; therefore, there was a lack of information about the instructional quality of teachers and their impact on promoting student achievement.

In response to states receiving federal grants from Race to the Top, teacher and principal evaluation systems implemented throughout the nation have been revised to include student performance data to reflect instructional strategies used by teachers. NCLB (2002) initially pressured states to revise evaluations. These educational reform efforts are a response to a concern that depletion in educational quality also will affect economic growth in America (Hanushek, 2009). This concern was initially brought to the forefront in 1983 with *A Nation at Risk*, which identified flaws in the American public educational system (National Commission on Excellence in Education, 1983).

The reformation of teacher evaluation systems is one of the most common components of reforming public education (Taylor & Tyler, 2012). Teacher evaluation tools often serve the purpose of measuring a teacher's competency and professional growth (Weems & Rogers, 2010). A comprehensive teacher evaluation system consists of both formative and summative evaluations. Shinkfield and Stufflebeam (2012) differentiated between formative and summative evaluation as follows: Formative evaluations focus on providing teachers with feedback that could improve teacher effectiveness, and summative evaluations focus on a teacher's worth to the school system. Between 2009 and 2012, nearly two thirds of U.S. school districts revised their

teacher evaluation systems to include specific performance standards to measure teacher effectiveness and practices (Jerald, 2012).

To gain an understanding of how the teacher evaluation can impact a teacher's effectiveness on student learning outcomes, researchers and stakeholders need a better understanding of teacher perceptions of the implementation of a performance-based teacher evaluation instrument (Darling-Hammond et al., 2012). Limited research has focused on teacher perspectives about the implementation of new evaluation systems and how new systems measure teacher effectiveness. According to Strunk et al. (2014), little evidence has supported the benefits of implementation of the newly revised teacher evaluations regarding measuring a teacher's ability to provide quality instruction and increase student academic achievement.

Perceptions of Teacher Evaluation Systems

Teacher evaluations are used to provide information about teacher effectiveness and can be used to inform teachers about their use of instructional practices (Wiliam & Thompson, 2007). Teachers increasingly are held accountable for student achievement gains through standardized test scores and student grades (Ladd, 2016). Decades of educational reform have spurred legislators to support more rigorous and specific evaluation systems (Ladd, 2016). Teacher perception of educator evaluations may have a direct impact on the classroom instruction, and in turn, on student learning (DeWitt, 2018). Because teacher evaluation systems are used as measures of accountability and to improve instruction and student learning, understanding teachers' perceptions of teacher evaluation systems is important (Ladd, 2016).

In a recent study, Finster and Milanowki (2018) investigated teacher perceptions of the impact of a new performance evaluation system on their practice. The researchers concluded teachers need to see links between the components of an evaluation system and impact on student learning. In another study on teacher perceptions of an evaluation system on classroom practice, Donahue and Vogel (2018) interviewed 30 teachers in one school district across subject areas. The researchers gleaned five themes from the data indicating components of an evaluation system that benefited teachers: feedback, quality of relationships, the evaluation rubric, modeling, personal integrity, and self-reflection.

Torbert (2014) conducted a quantitative descriptive study to determine how 126 teachers and administrators from an urban school district in Georgia perceived the effectiveness of the TKES. The purpose of the study was to analyze teacher and administrator perceptions of the effectiveness of the TKES. Findings indicated no statistically significant difference in participants' perceptions of the TKES by years of teaching experience, gender, or school level for teachers and for administrators. Like Torbert (2014), Griffith (2017) also conducted a study to determine teachers' perceptions of the TKES. Griffith explored high school teachers' experiences with the Georgia TKES. The phenomenological study examined 30 teachers' use of evaluative feedback from TKES to inform and impact classroom effectiveness. Findings from the study suggested that feedback from the TKES has the potential to support a positive change in the classroom, provided appropriate time and resources are dedicated to implementing the evaluative process with fidelity.

Jaffurs (2017) studied the correlation among teacher evaluation, self-reflection, and their roles in improving teacher quality. Jaffurs investigated tenured teachers'

perceptions of the effect of their teacher evaluation tool on teacher quality and other factors that contribute to a teacher's improvement of instructional performance over time. The study included 1,420 K-12 teachers employed in a rural/suburban school system. Findings from the study indicated that the majority of teachers viewed their local teacher evaluation system as being neutral to satisfactory as a tool for building a teacher's effectiveness over time. Additionally, the majority of teachers reported that the feedback they received during conferences positively impacted teacher quality. The majority of the teachers also stated teacher evaluation systems should include self-reflection activities such as completing a self-reflective checklist and reviewing taped lessons. However, Jaffurs also found that formal observations did not translate into higher evaluation ratings.

Duran (2018) conducted a phenomenological study to understand the role of the Texas Teacher Evaluation and Support System on the effectiveness of four public school teachers. The purpose of the study was to determine how each teacher perceived the Texas Teacher Evaluation and Support System. The research question that guided the study was the following: How do Texas teachers make sense of organizational influences that impact their practice, particularly when experiencing an evaluation system focused at improving student achievement?" Two sub-questions were asked: In what ways do different concepts of teacher effectiveness influence teachers' teaching practices, and what motivates teachers to improve practice for teacher evaluation? NVivo coding, attribute coding, and descriptive coding were used to analyze the data collected from the semi-structured interviews. Key findings from the study suggested that teachers used more planned-out lessons on evaluation days and that higher scores on the evaluation

system meant being accepted as a professional. Duran admitted that the teachers knew how to manipulate the evaluation tool for their benefit. Further, nice administrators made the evaluation tool less stressful, and each evaluator brought a personal perspective to the evaluation tool. As a result of the study, Duran recommended that future research focus on how teachers make meaning of teacher evaluation systems.

Redman (2018) conducted a quantitative study to investigate teacher perceptions of the Texas Teacher Evaluation and Support System as implemented during the 2016–2017 school year. The purpose of the study was to explore teachers' perceptions regarding the professional evaluation system used in their Texas schools to determine the relevance the system had in influencing teachers' current practices. Two research questions guided the study: What are the perceptions of teachers in Texas on the evaluation system they participated in during the 2016–2017 school year? To what extent are teacher perceptions related to gender, years of experience, and grade level taught as measured by the Teacher Evaluation Profile? The Teacher Evaluation Profile survey instrument developed by Stiggins and Duke (1988) was used to survey the 490 K-12 teacher participants of the study. Results indicated that 67% of the participants reported their evaluation had an average to above average impact on their professional practice (Redman, 2018). Further, 61% reported feedback from evaluators contributed significantly to the evaluation of their professional practice. Only 10% responded that the teacher evaluation process had a strong impact on their professional practice.

Frasier (2017) used a mixed methods approach to examine how a state-wide standardized evaluation policy in North Carolina impacted the performance of 45 high school teachers. Two research questions guided the study: What, if any, role do reported

school evaluation conditions and school evaluation status play in shaping teacher motivation, experiences with feedback, and work decisions related to teacher evaluation; and what individual-teacher-level factors are associated with differences in teacher motivation, experiences with feedback, and work decisions related to teacher evaluation? Frasier used the Teacher Working Conditions Survey and semi-structured interviews to collect data for the study. The survey was available online through Qualtrics. The teachers surveyed and interviewed indicated that they made few changes in their practices due to teacher evaluation systems and that even when changes were made, they were superficial adjustments rather than sustained changes. Moreover, feedback from evaluators about formal evaluations was not a motivating force for improving their practice. Teachers also had negative views about observation (Frasier, 2017).

Kappler (2017) used three research questions to guide a study of teacher perceptions of the Mississippi Teacher Evaluation System. The purpose of the study was to discover how Mississippi teachers who participated for at least 3 years perceived the teacher evaluation system and to determine whether the evaluation tool was useful in supporting and developing effective teaching practices and professional growth. The three research questions were the following: How do Mississippi teachers perceive the usefulness of the Mississippi Teacher Evaluation System? How do Mississippi teachers perceive that the Mississippi Teacher Evaluation System develops their effective teaching practices? How do Mississippi teachers perceive that the Mississippi Teacher Evaluation System supports teachers' professional growth? Kappler used a qualitative multiple-case study approach to provide descriptive, narrative, and percentage responses to the research questions. Seventy teachers completed the survey, and 20 participated in focus groups. Of

the 70 teachers surveyed, 91.4% perceived that the Mississippi Teacher Evaluation System was useful to them in reflecting and learning about their educator performance level, and 92.8% reported classroom observations had some degree of usefulness. About 80% of the participants perceived walk-throughs as having some degree of usefulness. Post-observation conferences were perceived as useful by 71.4% of the teachers. Pre-observation conferences were perceived as useful by 67.3% of teachers (Kappler, 2017).

Adkins (2017) investigated teacher perceptions of the teacher evaluation system in Florida, including its effectiveness and how it relates to improving teacher pedagogy and student achievement. The purpose of the qualitative case study was to understand the perceptions of elementary school teachers regarding their current evaluation system. The study included 12 participants at a K-5 elementary school in a rural Florida school district. The researcher used semi-structured open-ended questions to ensure that the interviews were focused, yet flexible; responses were digitally recorded. Adkins noted results of such a study would allow educational leaders to have an awareness and understanding of teacher perceptions, which in turn could help district leaders restructure their teacher evaluation system into a more effective tool to evaluate teachers and enhance or improve teacher pedagogy and student achievement.

DeWitt (2018) examined teacher perceptions of the teacher evaluation system used in Missouri to determine whether classroom instruction improved due to evaluation feedback. Three research questions guided the study: What is the difference in the teacher perception of the impact of the evaluation process on classroom instruction based on the involvement of the teachers in the designing of evaluation requirements? What is the difference in perception of the impact of the evaluation process between teachers based

on which type of evaluation tool is being used? What is the difference in teacher perception of the impact of the evaluation process based on the size of the school being surveyed? DeWitt surveyed 841 K-12 teachers. DeWitt created the survey to collect demographic information, perceptions of teachers on the teacher evaluation system in Missouri, and teacher perceptions on professional development practices. An analysis of variance (ANOVA) and the Pearson's *r* value were used to establish the significance of the differences in the mean of teacher perceptions regarding the teacher evaluation process. Through the analysis of the survey data, the findings showed teachers had varying perceptions regarding some aspects of the teacher evaluation systems. Teachers noted the teacher evaluation system encouraged feedback regarding teacher performance so teachers could look for ways to improve their classroom practice. Collaboration was also noted to be embraced in some school districts but not in others, due to district size and professional development practices involved. As a result of the findings from the study, DeWitt suggested future research focus on understanding how teacher perceptions regarding the impact of the teacher evaluation process on classroom practice vary with time.

Bradley-Levine et al. (2017) examined how teacher perceptions of the teacher evaluation process were influenced by the involvement of teacher leaders as evaluators. The study utilized a mixed methods online survey design with 148 teachers and teacher leaders in the district. Three sets of scaled items from Angelle and Dehart's (2011) Teacher Leadership Inventory were used. The three sets of items measure these concepts: Sharing Expertise, Sharing Leadership, and Supra-Practitioner. The five main variables tested in this study were Sharing Expertise Scale, Sharing Leadership Scale, Supra-

Practitioner Scale, Overall Teacher Leadership Scale, and Teacher Evaluation Scale. Findings from the study indicated that teachers appreciated the opportunity to be evaluated by colleagues rather than only by principals and that a culture of shared leadership should extend beyond teacher leaders to teachers themselves. As a result of the study, Bradley-Levine et al. suggested further studies be conducted across districts that have implemented similar evaluation models to determine how teacher evaluation systems influence teaching and learning outcomes.

Norris et al. (2017) conducted a study to develop a better understanding of 22 physical education teachers' perceptions of teacher evaluation systems. The researchers used a mixed methods approach and two sources of data, a short survey of physical education teachers and formal semi-structured interviews. Interviews revealed three common themes: (a) Physical education is valued, but not prioritized; (b) teacher evaluation in physical education is greatly needed, yet not transparent; and (c) physical educators are not confident in their evaluators. Results from the Norris et al. study indicated that (a) physical education remains a marginalized subject with low priority; (b) teacher evaluation systems are not tailored toward noncore subjects; (c) and physical education teachers, like classroom teachers, are not confident that their evaluators can give a fair and valid assessment.

The Georgia Department of Education Teacher Keys Effectiveness System

Teacher evaluation systems in Georgia have mirrored those across the country that rely on student achievement to determine teacher effectiveness (Rogers, 2019). During the 2012-2013 legislative session, the Georgia Legislature mandated the use of a teacher evaluation system. According to the Georgia Department of

Education (2013), the Georgia TKES was fully implemented as the official evaluation tool for Georgia public schools during the 2012-2013 academic year. The TKES has two major goals: (a) to improve student achievement and (b) to increase teacher effectiveness (Georgia Department of Education, 2013, 2018).

The TKES consists of a rubric used annually to measure teacher effectiveness through summative and formative evaluations. According to the Georgia Department of Education (2018), the TKES measured three components: (a) Teacher Assessment on Performance Standards (50% of overall score); (b) Professional Growth (20%), and (c) Student Growth (30% of overall score). The Assessment on Performance Standards is a qualitative, rubrics-based evaluation method that evaluators use to measure teacher quality on 10 performance standards. Professional Growth is measured by teachers' attainment of Professional Growth Goals (Georgia Department of Education, 2018). Student Growth is measured by Student Growth Percentile Measures based on state assessment data (Georgia Department of Education, 2018).

For the Teacher Assessment on Performance Standards part of the TKES (Georgia Department of Education, 2013, 2018), a teacher is rated based on 10 Performance Standards grouped into five domains. These standards are shown in Tables 2–6.

Table 2

Georgia Teacher Keys Effectiveness System Performance Standards for the Domain of Planning

Performance Standard	Description
1. Professional Knowledge	Demonstrates content knowledge and pedagogy: • Addresses curriculum standards • Facilitates student higher level thinking • Links 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 and best practices research • Has high expectations for all students and a clear understanding of the curriculum • Understands the intellectual, social, emotional, and physical development of the age group
2. Instructional Planning	Uses standards, strategies, resources, and data: • Analyzes and uses student learning data to inform planning. • Develops clear, sequential plans integrated across the curriculum • Plans instruction effectively for content mastery, pacing • Plans for instruction to meet the needs of all students. • Aligns lesson objectives to state and district curricula and standards and student learning needs. • Develops appropriate course, unit, and daily plans, and can adapt plans when needed

Note. Source: *Georgia's Teacher Keys Effectiveness System Implementation Handbook,* by Georgia Department of Education, 2018, Atlanta, GA: Author.

Table 3

Georgia Teacher Keys Effectiveness System Performance Standards for the Domain of Instructional Delivery

Performance Standard	Description
3. Instructional Strategies	Uses research-based strategies to engage students in active learning: • Engages students in active learning • Builds on students' existing knowledge and skills • Reinforces learning goals consistently throughout the lesson • Uses a variety of research-based strategies and resources • Effectively uses appropriate instructional technology • 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	• Challenges and supports learning of diverse students • Differentiates the instructional content, process, product, and learning environment • Provides remediation, enrichment, and acceleration • Uses flexible grouping strategies • Uses diagnostic, formative, and summative assessment data • Develops critical and creative thinking by providing activities at the appropriate level of challenge • Demonstrates high learning expectations for all students commensurate with their developmental levels

Note. Source: *Georgia's Teacher Keys Effectiveness System Implementation Handbook,* by Georgia Department of Education, 2018, Atlanta, GA: Author.

Table 4

Georgia Teacher Keys Effectiveness System Performance Standards for the Domain of Assessment of and for Learning

Performance Standard	Description
5. Assessment Strategies	Uses diagnostic, formative, and summative assessments: • Aligns assessment with 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 appropriate for the developmental level of students • Collaborates to develop common assessments
6. Assessment Uses	Uses data to inform instruction and provide feedback to students and parents: • Uses diagnostic assessment data to develop goals for students, differentiate instruction, and document learning • Plans formal and informal assessments to measure student mastery of learning objectives • Uses assessments for both formative and summative purposes to inform, guide, and adjust instruction • Systematically analyzes and uses data to measure student progress, design appropriate interventions, and 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 • Teaches students how to self-assess and to use metacognitive strategies in support of lifelong learning

Note. Source: *Georgia's Teacher Keys Effectiveness System Implementation Handbook,* by Georgia Department of Education, 2018, Atlanta, GA: Author.

Table 5

Georgia Teacher Keys Effectiveness System Performance Standards for the Domain of Learning Environment

Performance Standard	Description
7. Positive Learning Environment	Provides a well-managed, safe, respectful classroom: • Responds to disruptions in a timely, appropriate manner • Establishes clear expectations for classroom rules, routines, and procedures and enforces them consistently and appropriately • Models caring, fairness, respect, and enthusiasm for learning • Promotes a climate of trust and teamwork • Promotes respect for and understanding of students' diversity • Actively listens and pays attention to students' needs and responses • Creates a warm, attractive, inviting, and supportive classroom • Arranges materials and resources to facilitate group and individual activities
8. Academically Challenging Environment	Creates a student-centered academic environment: • Maximizes instructional time • Conveys that mistakes should be embraced as a valuable part of learning • Provides students appropriately challenging and relevant material and assignments • Provides transitions that minimize loss of instructional time • Communicates high, but reasonable, expectations for student learning • Provides academic rigor, encourages critical and creative thinking, and pushes students to achieve goals • Encourages students to explore new ideas and take academic risks

Note. Source: *Georgia's Teacher Keys Effectiveness System Implementation Handbook,* by Georgia Department of Education, 2018, Atlanta, GA: Author.

Table 6

Georgia Teacher Keys Effectiveness System Performance Standards for the Domain of Professionalism and Communication

Performance Standard	Description
9. Professionalism	Shows commitment to ethics and school mission and participates in professional development: • Follow federal and state laws; Code of Ethics; and state and local school board policies, regulations, and practices • Maintains professional demeanor and behavior • Respects and maintains confidentiality • Evaluates and identifies personal strengths and weaknesses related to professional skills and their impact on student learning and sets goals for improvement. • Participates in ongoing professional development based on identified areas for improvement and incorporates learning into classroom activities • Demonstrates flexibility in adapting to school change • Engages in activities outside the classroom intended for school and student enhancement
10. Communication	Communicates with students, parents, and other stakeholders: • Fosters positive interactions and promotes learning in the classroom and school • Engages in ongoing communication; shares instructional goals, expectations, and student progress with families in a timely and constructive manner • Collaborates and networks with colleagues and community to enhance and promote student learning • Uses precise language, correct vocabulary and grammar • Explains directions, concepts, and lesson content to students in a logical, sequential, and age-appropriate manner • Adheres to policies on communication of student information • Promotes accessibility to parents and students via a collaborative, approachable style • Listens and responds with cultural awareness, empathy, and understanding to stakeholders • Uses modes of communication appropriate to the situation

Note. Source: *Georgia's Teacher Keys Effectiveness System Implementation Handbook,* by Georgia Department of Education, 2018, Atlanta, GA: Author.

Teachers are graded on the TKES using performance appraisal rubrics (Georgia Department of Education, 2015, 2018). The rubric uses a scale that guides evaluators in assessing how well a standard is performed. The rubric states the measure of performance expected of teachers and provides a qualitative description of performance at each level: Level I, Level II, Level III, and Level IV. 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, and thus Level III is the expected level of performance.

Chapter III

Methodology

This chapter includes details of the methodology utilized for the current study. The rationale for choosing a nonexperimental survey design with group comparisons is discussed. A discussion of the target population, accessible population, and sample is also presented. Next, I discuss the procedures for study participation and data collection. I then present the data analysis procedures and statistical assumptions. A summary of the important details of the proposed methodology concludes the chapter. The research questions presented below guided the study:

1. Is there a significant difference by gender and race or ethnicity on selected dependent variables?

a. Is there a significant difference by gender and race or ethnicity on teachers' perceived influence of the TKES on teaching practices?

b. Is there a significant difference by gender and race or ethnicity on teachers' perceived accuracy of the TKES assessment?

c. Is there a significant difference by gender and race or ethnicity on the perceived level of creativity allowed by the TKES?

2. Is there a significant difference by grade level taught and level of education on selected dependent variables?

a. Is there a significant difference by grade level taught and level of education on teachers' perceived influence of the TKES on teaching practices?

b. Is there a significant difference by grade level taught and level of education on teachers' perceived accuracy of the TKES assessment?

c. Is there a significant difference by grade level taught and level of education on the perceived level of creativity allowed by the TKES?

3. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on selected dependent variables?
   a. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on teachers' perceived influence of the TKES on teaching practices?
   b. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on teachers' perceived accuracy of the TKES assessment?
   c. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on the perceived level of creativity allowed by the TKES?
4. Is there a significant difference by content area taught on the selected dependent variables?
   a. Is there a significant difference by content area taught on teachers' perceived influence of the TKES on teaching practices?
   b. Is there a significant difference by content area taught on teachers' perceived accuracy of the TKES assessment?
   c. Is there a significant difference by content area taught on teachers' perceived level of creativity allowed by the TKES?
5. Is there a significant difference between levels of the effectiveness of training on the TKES on teaching practices?

6. Is there a significant difference between levels of the effectiveness of training on the TAPS on teaching practices?
7. Is there a significant difference between levels of the effectiveness of training on the SGP on teaching practices?

Research Design

A nonexperimental survey design with group comparisons was employed for this study. A nonexperimental design is research that lacks the manipulation of an independent variable, random assignment of participants to conditions or orders of conditions, or both (Babones, 2016; Della Porta & Keating, 2008). Specifically, a survey design was followed to describe characteristics of a group of population (Johnson & Christensen, 2012). Researchers who follow the survey design use either self-developed or previously validated surveys to collect data from participants regarding the topic under consideration (Kelley-Quon, 2018). In survey design, a sample of the target population is recruited because including the entire population of interest in the study is impossible. Sampling procedures are used to ensure representativeness of the population. For this study, I used stratified random sampling to ensure representativeness of each group. Participants were asked to answer web-based surveys. Those who preferred a hard copy survey were provided one with postage paid by the researcher.

The nine independent variables included gender (male or female), race or ethnicity (Asian, Black or African American, Latino/Hispanic, White, Native Hawaiian or other Pacific Islander, and American Indian or Alaska Native), grade level (prekindergarten to 12th grade), education attainment level (bachelor's degree, master's degree, Educational Specialist's degree, and doctorate), years of teaching experience (1–2

years, 3–5 years, 6–10 years, and more than 10 years), subject taught (mathematics, science, language arts, and electives), extent of influence of training on TKES, extent of influence of training on TAPS, and extent of influence of training on SGP. Gender, race or ethnicity, grade level, educational attainment, and content area taught were measured on the nominal scale; years of experience were measured on an interval scale; and training was measured on a nominal scale.

The dependent variables for this study were influence on teaching practices, agreement with assessment score, and allows teacher creativity. These items were measured on the interval level using a 5-point rating scale, multiplied by the 10 items in each section to result in a total score of 10–50 for each variable. Ten items in Section B asked about influence on teaching practice, 10 items in Section C asked about agreement with assessment score, and 10 items in Section D asked about allowed teacher creativity. For each of these sections, a summed score of 10–50 was gathered from each participant (rating of 1–5 times 10 questions). These total variable scores were used to answer RQ1 through RQ4.

Participants

The target population of this study was elementary, middle school, and high school teachers in Georgia. The accessible population was selected from a southern George RESA district for certified teachers. There are 1,129 elementary teachers, 472 middle school teachers, and 532 high school teachers in the district. I used stratified random sampling to identify participants for the study.

The required minimum sample size was determined through conducting a power analysis using G*Power software (Faul et al., 2013). Five factors considered in the power

analysis were representativeness, significance level, effect size, power of test, and statistical test. Significance level refers to the probability of incorrectly rejecting a true null hypothesis, also commonly called type I error (Mascha & Vetter, 2017). The power of a test refers to the probability of rejecting a false null hypothesis (Mascha & Vetter, 2017). In most quantitative studies, significance level is set at 95%, and power of test is set at 80% (Koran, 2016). I used the same factors for this study: 95% significance level and 80% power of test. Lastly, I used factorial ANOVA and ANOVA to address the research questions and test the hypotheses. Using 95% significance level, 80% power of test, medium effect size (f^2 =. 25), and ANOVA with five groups (the greatest number of groups for the independent variables, representing race or ethnicity), the minimum required sample size was determined to be 400.

Stratified random sampling allows a researcher to reach a targeted sample in which each member of the identified smaller groups (or strata) that shares common attributes or characteristics has an equal probability of being chosen (Frey, 2018; Nickolas, 2020). Random samples are taken from the stratified groups in proportion to the population (Nickolas, 2020). A stratified sample can only be obtained if a researcher has a complete list of the population available, which was the case in this study. A representative sample is a subset of a population that accurately reflects the characteristics of the larger group. The accessible population was a little more than 2,100 teachers across elementary, middle, and high schools. With a 95% confidence interval and allowing for 5% margin of error, the sample should be around 400 teachers. Given the accessible population and considering the representativeness of the accessible population, the sample size necessary for this study was 400 teachers; specifically, I

sought to recruit 200 elementary, 100 middle, and 100 high school teachers. However, to account for nonresponse from teachers, the minimum sample was doubled to 800 teachers, including 400 elementary, 200 middle, and 200 high school teachers. Thus, the sample was representative in terms of grade level taught.

Instrumentation

Survey Description

I collected data for this study through the Teacher Survey of the Georgia TKES (see Appendix A). I adapted a survey developed in previous research by Battle-Edwards (2017). The survey gathers data on teachers' perceptions of the TKES. Each survey item is rated on a 5-point rating scale, from 1 = *not at all* to 5 = *to a very large extent*. Section A gathers data on perceptions of training on the TKES components, with an item rated on the 5-point scale yielding a rating of 1 to 5 for each type of training (TKES, TAPS, and SGP). Each type of training is rated by a single survey item. Section B gathers data on influence on teaching practices, with 10 items rated on the 5-point scale for a total dependent variable score of 10-50. Section C gathers data on perceived agreement with accuracy of TKES evaluation, with 10 items rated on the 5-point scale for a total dependent variable score of 10-50. Section D gathers data on creativity allowed, with 10 items rated on the 5-point scale for a total dependent variable score of 10-50. Section E gathers the following demographic information: experience as a Georgia certified teacher, highest level of education completed, ethnicity or race, gender, grade level teaching in, and content area.

A lengthy search of the literature revealed only one previous survey suitable for the current study. Battle-Edwards (2017) developed a survey to measure the influence of

the Georgia TKES on teachers' instructional practices and motivation. In Battle-Edwards' mixed-methods case study, the survey was completed by 41 teachers in an Atlanta Title I high school. The survey included the following various components of the TKES Teacher Assessment of Performance Standards: instructional planning, professional knowledge, instructional strategies, differentiated instruction, assessment strategies, assessment uses, positive learning environment, academically challenging environment, professionalism, and communication. I obtained permission from Battle-Edwards to use the survey and adapt it to the current research. Battle-Edwards noted the small amount of literature related to the Georgia TKES and recommended more research on the topic.

Validity

Validity indicates the extent that an instrument measures the construct or variable it is intended to measure (Pogrow, 2018). For content validity, instrument items should measure what they are intended to measure (Creswell & Plano Clark, 2018). Content validity means the items on an instrument represent the entire domain the instrument is designed to measure (Salkind, 2010). I adapted a survey developed by Battle-Edwards (2017) and altered survey items based on my research questions and variables examined in the study. Battle-Edwards did not indicate reliability or validity of the survey. Therefore, in an effort to provide content validity, a formal expert panel reviewed the survey. Expert review and validation is a technique to review instruments for clarity, technical quality of items, and content validity (Brockmeier et al., 2009, 2014; McNeill & Brockmeier, 2005). The expert panel included a university professor in education who helped design the Georgia TKES and a dissertation consultant and editor with 20 years of

experience in educational research. I invited four more experts to examine the survey: the district superintendent, assistant superintendent, a principal, and the assessment and accountability director from another district.

Reliability

The reliability of an instrument means the results will be consistent in multiple applications with the same sample of participants (Heale & Twycross, 2015). To test for the internal consistency of the self-developed survey, Cronbach's alpha will be used. A Cronbach's alpha value of higher than .70 is considered acceptable (Heale & Twycross, 2015), meaning that the questions in the scale are consistently measuring the intended variable.

After IRB approval, I pilot tested the adapted survey with a few teachers, who were not part of the final study sample. Prior to using the survey instrument, pilot testing was conducted to review and assess the questions to ensure that they were credible, relevant, and able to yield the required information from all participants. Pilot testing refers to the process of trying research instruments or methods with a small group of people with the same characteristics as the intended research sample (Lavrakas, 2008; McNeill & Brockmeier, 2005). The goals of a pilot test are to review the instrument wording and directions and test the feasibility of an instrument with a target population (Lavrakas, 2008).

Data Collection

Upon receiving Institutional Review Board (IRB) permission, approval from the superintendents of the southern Georgia RESA district was secured. I sent a letter to the superintendents of each school district in the southern Georgia RESA district. The letter

outlined the specifics of the study and provided a detailed explanation of how the survey would be conducted. The letter ensured the superintendents that teacher participation was voluntary and confidential and that their identities would be protected. The letter was followed up with an email, if necessary. I also sent a letter to the IRB of each school district. After securing the approval of the superintendents and any district IRBs, I sent out invitation emails to 400 elementary, 200 middle, and 200 high school teachers in the district. Teachers were informed that their responses would be kept anonymous and results would be reported in aggregate. A link to the Qualtrics online survey tool was distributed via e-mail to all teachers evaluated under the Georgia TKES. Using the Qualtrics link, teachers were first directed to an online consent form to electronically sign. Then, the Qualtrics link directed teachers to the survey online. The email noted that participants who preferred to complete a paper copy of the survey may contact me for a pdf version of the survey, which they could complete with pen and return in a postage-paid return envelope. Teachers did not receive both paper and online links initially to prevent the potential for teachers submitting both. Until the appropriate sample was achieved (200 elementary, 100 middle, and 100 high school teachers), I sent follow-up email reminders in 2-week intervals to encourage teachers who have not completed the survey to do so.

To increase survey response rates, I followed the recommendations of Saleh and Bista (2017). These researchers identified factors influencing online response rates in educational research. They recommended eliciting the aid of authority figures; in this instance, I included a sentence in the invitation email to teachers noting that the district superintendent had approved the study. The survey was also designed to be concise and

take little time to complete. The invitation letter noted the estimated time to complete the survey, which was determined during the pilot test. Participants were assured of anonymity, and the survey included no open-ended responses. Saleh and Bista recommended no more than three reminders; however, up to four reminders were sent to potential respondents. I had hoped that following these research-based recommendations from Saleh and Bista would increase the response rate.

Data Analysis

The data analysis for this study was performed using the Statistical Package for the Social Sciences (SPSS) for Windows to provide a range of descriptive as well as inferential statistics. Descriptive statistics are described in the next section, followed by a discussion of the inferential statistics. Statistical considerations and assumptions are addressed as well.

Descriptive Statistics

Descriptive analysis was conducted first to characterize the demographic characteristics of the participants. Specifically, the frequency and percentage of categorical or nominal demographic characteristics (gender, race or ethnicity, grade level, education level, and content area taught) were computed. The nine independent variables included gender (male or female), race or ethnicity (Asian, Black or African American, Latino/Hispanic, White, Native Hawaiian or other Pacific Islander, and American Indian or Alaska Native), grade level (pre-K to 12th grade), education attainment level (bachelor's degree, master's degree, Educational Specialist's degree, and doctorate degree), subject taught (mathematics, science, language arts, and electives), level of influence of training on TKES, level of influence of training on TAPS, and level of

influence on SGP. Frequencies and percentages were computed for the interval demographic characteristic of years of teaching experience (1-2 years, 3-5 years, 6-10 years, and more than 10 years).

Descriptive statistics in the form of frequencies were reported for every survey item, which are rating-scale items from 1-5. Survey Item 1 in Section A measures the extent to which the training respondents received for the TKES influenced their teaching practice. Survey Item 2 measures the extent to which the training respondents received for the TAPS influenced their teaching practice. Survey Item 3 measures the extent to which the training respondents received for the SGP influenced their teaching practice. Frequencies and percentages of each response were reported for each survey item.

For survey sections B, C, and D, I first reported item-level statistics, including frequencies and percentages, of each response for each survey item. Each section has 10 survey items. I then reported the total scores for each section. The summed responses to the 10 items in each section yielded a total scale score of 10-50. Mean, sample size, standard deviation, minimum value, maximum value, skew, and kurtosis were computed from total scores.

Inferential Statistics

The main part of the data analysis included inferential analyses. Specifically, one-way ANOVA and two-way factorial ANOVA were used to compare means on the dependent variables across one or more independent variables. The primary purpose of one-way ANOVA and two-way factorial ANOVA is to compare means to determine the potential effect of the independent variables on the dependent variable. The one-way

ANOVA and two-way factorial ANOVAs were conducted using a 95% confidence level and an alpha level of .05.

For RQ1, RQ2, and RQ3, two-way factorial ANOVAs were used. The dependent variables across all three research questions are teachers' perceptions of the state-level evaluation instrument impacting teaching practices, agreement with assessment score, and allowed teacher creativity. These items are measured on the interval level using a 5-point rating scale and summed over the 10 items to reach a total score for each variable of 10-50, also at the interval level of measurement. The independent variables for RQ1 were the nominal variables of gender and race or ethnicity. The independent variables for RQ2 were the nominal variables of grade level and teacher education level. The independent variables for RQ3 were years of experience and education level.

A two-way factorial ANOVA was computed for each pair of independent variables and dependent variable; thus, three two-way factorial ANOVAs were conducted for RQ1, RQ2, and RQ3. The table showing tests of between-subjects effects was reported to determine whether the two independent variables or their interaction were statistically significant. The effects table included type III sum of squares, degrees of freedom, mean square, *F*-statistic, and significance value. The *F*-statistic is simply a ratio of two variances. If the *F*-statistic was significant (*p*-value is less than .05), then it was concluded that there was a significant difference on the means across the groups as identified by the independent variables.

The two-way factorial ANOVA does not depict which of the means in the design are different or if they are different. In order to do this, post hoc tests are needed. If the results were statistically significant, Tukey post hoc test results were conducted for

different levels of the independent variables. A table was created to show the multiple comparisons, including the mean difference, standard error, significance level, and 95% confidence interval.

For RQ4, one-way ANOVA was used. The dependent variables were teachers' perceptions of the state-level evaluation instrument impacting teaching practices, agreement with accuracy of assessment, and teacher creativity. The independent variable was content area taught. The dependent variables were measured based on a scale score of 10-50, using an interval level of measurement, whereas the independent variable was measured in nominal (or categorical) form.

For RQ5, RQ6, and RQ7, one-way ANOVA was used. The dependent variable was teachers' perceptions of the state-level evaluation instrument impacting teaching practices. The independent variable was the level of extent to which teachers reported training influenced their teaching practice (for RQ5, training on TKES; for RQ6, training on TAPS; for RQ7, training on SGP). Responses to each item in Section A were treated as nominal variables of level of training effectiveness: (a) not at all, (b) to a small extent, (c) to some extent, (d) to a large extent, and (e) to a very large extent. Participants in each variable group were compared based on the dependent variable as well as total scores for Survey Section B: influence on teaching practice.

The one-way ANOVA results were reported to determine whether there was a statistically significant difference between the group means. The one-way ANOVA included sum of squares, degrees of freedom, mean square, *F*-statistic, and significance value for both between groups and within groups. The *F*-statistic is simply a ratio of two variances. If the *F*-statistic was significant (*p*-value is less than .05), then it was

concluded that there was a significant difference on the means across the groups as identified by the independent variable.

The one-way ANOVA result does not tell which of the means in the design are different or if they are different. In order to do this, post hoc tests are needed. Tukey post hoc test results were conducted to determine which groups differed from each other. The results included the mean difference, standard error, significance level, and 95% confidence interval.

Statistical Assumptions

One-way ANOVA and two-way factorial ANOVAs are considered parametric tests. Statistical considerations and assumptions must be met before ANOVA can be used. There are four assumptions of parametric tests: (a) normality, (b) homogeneity of variance, (c) independence, and (d) dependent variable in interval/ratio form (Sedgwick, 2015). A Shapiro-Wilks test was performed to detect whether the dependent variables complied with the normality assumption (Siddiqi, 2014). In addition, if further investigation is needed to ensure the normality assumption is met, stem-and-leaf pots, histograms, or Q-Q plots were generated and used to evaluate how well the distribution of the dataset matches a standard normal distribution. Furthermore, skewness and kurtosis values for each variable were calculated to assess the normality of the data. These values measure the shape of the distribution of each variable.

Outliers affect statistics, such as means and standard deviations (Kwak & Kim, 2017). The scores for the variable scales were converted to *z*-scores. The actual *z*-score depended on the number of respondents. A larger value might be used with a very large

sample size, while a smaller value might be used with a small sample size. For the outliers, I used the cut-off value of 3.29 by frequency distribution.

Levene's test was used to test for homogeneity of variance and whether there was a violation of this assumption (Sedgwick, 2015). The independence of observations was ensured by participant responses being independent of each other (Huber & Melly, 2015). An assumption is a dependent variable in interval/ratio form (Sedgwick, 2015). That assumption was met.

Summary

The purpose of this nonexperimental survey design with group comparisons was to examine teacher perceptions and effectiveness of TKES evaluation in improving teacher professional growth to impact student academic learning. The target population for this study was elementary, middle, and high school teachers in Georgia. A total of 800 teachers (400 elementary, 200 middle, and 200 high school teachers) were asked to participate in the study. A stratified random sampling was used to ensure a representative sample. Data were collected using a survey that teachers may fill out either online or via paper-and-pen (see Appendix A). The dependent variables were influence on teaching practices, agreement with assessment score, and allows teacher creativity. The independent variables were gender, race or ethnicity, grade level taught, education level, years of teaching experience, content area taught, and extent of influence of training (on TKES, TAPS, and SGP). One-way ANOVAs and two-way factorial ANOVAs were conducted to address the research questions.

Chapter IV

Results

Introduction

The main purpose of this quantitative research study was to determine how (a) gender, (b) race or ethnicity, (c) grade level taught, (d) level of education, (e) years of education as a certified teacher, (f) content area taught, (g) level of effective training for the TKES, (h) level of effective training for the TAPS, and (i) level of effective training for the SGP impact (1) teachers' perceived influence of the TKES on teaching practices, (2) teachers' perceived accuracy of the TKES assessment, and (3) the perceived level of creativity allowed by the TKES. The following research questions guided this study:

1. Is there a significant difference by gender and race or ethnicity on selected dependent variables?
    d. Is there a significant difference by gender and race or ethnicity on teachers' perceived influence of the TKES on teaching practices?
    e. Is there a significant difference by gender and race or ethnicity on teachers' perceived accuracy of the TKES assessment?
    f. Is there a significant difference by gender and race or ethnicity on the perceived level of creativity allowed by the TKES assessment?
2. Is there a significant difference by grade level taught and level of education on selected dependent variables?
    d. Is there a significant difference by grade level taught and level of education on teachers' perceived influence of the TKES on teaching practices?

e. Is there a significant difference by grade level taught and level of education on teachers' perceived accuracy of the TKES assessment?

f. Is there a significant difference by grade level taught and level of education on the perceived level of creativity allowed by the TKES assessment?

3. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on selected dependent variables?

   a. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on teachers' perceived influence of the TKES on teaching practices?

   b. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on teachers' perceived accuracy of the TKES assessment?

   c. Is there a significant difference by years of experience as a certified teacher in Georgia and education level on the perceived level of creativity allowed by the TKES assessment?

4. Is there a significant difference by content area taught on the selected dependent variables?

   a. Is there a significant difference by content area taught on teachers' perceived influence of the TKES on teaching practices?

   b. Is there a significant difference by content area taught on teachers' perceived accuracy of the TKES assessment?

c. Is there a significant difference by content area taught on teachers' perceived level of creativity allowed by the TKES assessment?

5. Is there a significant difference between level of effective training for the TKES on TKES teaching practices?
6. Is there a significant difference between level of effective training for the TAPS on TKES teaching practices?
7. Is there a significant difference between level of effective training for the SGP on TKES teaching practices?

The data analysis for this study will be performed using the Statistical Package for the Social Sciences (SPSS) for Windows, to provide a range of descriptive as well as inferential statistics. The data is first analyzed using frequency analysis to determine the distribution of the respondents into specific demographic groups. Multiple one-way and two-way ANOVAs were conducted to evaluate the research hypotheses. The underlying assumptions of each ANOVA procedure are discussed before presenting its results. The research questions were evaluated based on the results obtained from these analyses.

This chapter is organized as follows: First, descriptive statistics are provided followed by presenting the results from the reliability analysis of the survey instruments. After that, the results of evaluating the research questions and hypotheses are presented. Finally, the chapter concludes with a summary of the results.

Demographic Characteristics

The data was first analysed using frequency analysis to determine the distribution of the respondents into specific demographic groups. As shown below in Table 1, a total of 347 respondents participated in the study. Among these 347 participants, the majority

were female (280 out of 347, 80.7%). Likewise, a majority of the participants were reported to be of White or Caucasian ethnicity (305 out of 347, 87.9%). With regard to the grade level taught, the largest group of participants were high school teachers (94 out of 347, 27.1%), followed by middle school teachers (75 out of 347, 21.6%).

The majority of the participants were also post-graduate degree holders, with 38.9% (135 out of 347) reported to have a master's degree, and 32% (111 out of 347) reported to have an Educational specialist's degree. With regard to content area taught, the largest group included the respondents who taught general subjects in elementary grades (67 out of 347, 19.3%), followed by respondents who taught Mathematics (51 out 347, 14.7%). The respondents were also surveyed about the number of years they have been certified as a teacher in Georgia. The largest group of respondents have been certified for more than 21 years (114 out of 347, 32.9%).

Table 1

Frequencies and Percentages of Demographic Characteristics of Participants

	N	%
Gender		
Male	67	19.3
Female	280	80.7
Ethnicity		
Teachers of Color	42	12.1
White or Caucasian	305	87.9
Grade Level		
Elementary: Pre-K to K	24	6.9

	N	*%*
Elementary: Gr1-3	62	17.9
Elementary: Gr4-5	53	15.3
Middle School: Gr6-8	75	21.6
High School: Gr9-12	94	27.1
Other	39	11.2

Table 1

Frequencies and Percentages of Demographic Characteristics of Participants

	N	*%*
Educational Level		
Bachelor's degree	79	22.8
Master's degree	135	38.9
Educational Specialist's degree	111	32.0
Doctoral degree	22	6.3
Content Area		
Elementary grades, general	67	19.3
Special education	51	14.7
English Language Arts	54	15.6
Mathematics	51	14.7
Social Studies	24	6.9
Vocation, career, or technical education	21	6.1
Science	21	6.1
Other	58	16.7
Number of Years Certified as a Teacher		

1-5 years	55	15.9
6-10 years	43	12.4
11-15 years	61	17.6
16-20 years	74	21.3
21+ years	114	32.9

Frequency Analysis for Each Survey

The respondents were also surveyed regarding their perceptions of the extent that three specific training programs have influenced their teaching practices, specifically the TKES, the TAPS, and the SGP. Tables 2-5 contain the breakdown of responses for each instrument. As shown below, the respondents indicated that the TKES training has influenced their teaching practices 'moderate' (131 out of 347, 37.8%), and 'large' (123 out of 347, 35.4%; see Table 2). The TAPS training was reported to influence the respondents' teaching practices 'moderate' (137 out of 347, 39.5%; see Table 2), while the SGP was similarly reported to have influenced the respondents' teaching practices 'moderate' (141 out of 347, 40.6%; see Table 2).

Table 2

Frequencies and Percentages of the Perceived Influence on Teaching Effectiveness Based on TKES, TAPS, and SGP

	1	2	3	4	5	*Mdn*	*M*	*SD*
TKES	61 (17.6%)	131 (37.8%)	123 (35.4%)	32 (9.2%)	0 (0.0%)	3	3.36	0.88
TAPS	50 (14.4%)	46 (13.3%)	137 (39.5%)	91 (26.2%)	23 (6.6%)	3	2.97	1.11
SGP	31 (8.9%)	53 (15.3%)	141 (40.6%)	98 (28.2%)	24 (6.9%)	3	3.09	1.03

Note: 1 = Not at all, 2 = To a small extent, 3 = To some extent, 4 = To a large extent, 5 = To a very large extent

Table 3 presents the frequencies and percentages of responses of participants in each of the TKES items for influence on teaching practices. The data showed that for all items, the greatest number of participants responded 'moderate' in all of the items.

Table 3

Item Frequencies and Percentages for TKES – Influence

Item	1	2	3	4	5	*Mdn*	*M*	*SD*
Professional Knowledge	33 (9.5%)	70 (20.2%)	137 (39.5%)	84 (24.2%)	23 (6.6%)	3	2.98	1.05
Instructional Planning	32 (9.2%)	68 (19.6%)	140 (40.3%)	83 (23.9%)	24 (6.9%)	3	3.00	1.04
Instructional Strategies	34 (9.8%)	68 (19.6%)	135 (38.9%)	88 (25.4%)	22 (6.3%)	3	2.99	1.05
Differentiated Strategies	39 (11.2%)	55 (15.9%)	140 (40.3%)	89 (25.6%)	24 (6.9%)	3	3.01	1.07
Assessment Strategies	41 (11.8%)	65 (18.7%)	145 (41.8%)	77 (22.2%)	19 (5.5%)	3	2.91	1.05

Assessment	40	64	147	78	18	3	2.91	1.04
Use	(11.5%)	(18.4%)	(42.4%)	(22.5%)	(5.2%)			
Positive	44	63	140	70	30	3	2.94	1.11
Learning	(12.7%)	(18.2%)	(40.3%)	(20.2%)	(8.6%)			
Environment								
Academically	42	54	147	79	25	3	2.97	1.08
Challenging	(12.1%)	(15.6%)	(42.4%)	(22.8%)	(7.2%)			
Environment								
Professionalism	50	66	116	84	31	3	2.94	1.17
	(14.4%)	(19.0%)	(33.4%)	(24.2%)	(8.9%)			
Communication	50	70	130	71	26	3	2.86	1.13
	(14.4%)	(20.2%)	(37.5%)	(20.5%)	(7.5%)			

Note: 1 = Not at all, 2 = To a small extent, 3 = To some extent, 4 = To a large extent, 5 = To a very large extent

For professional knowledge (n = 107, 30.8%), instructional planning (n = 107, 30.8%), instructional strategies (n = 110, 31.7%), differentiated strategies (n = 113, 32.6%), and academically challenging environment (*n* = 104, 30.0%), there were more participants who responded either ‘large’ or ‘very large’ than “not at all” or ‘small’.For assessment strategies (*n* = 106, 30.5%), assessment use (*n* = 104, 30.0%), positive learning environment (*n* = 107, 30.8%), professionalism (*n* = 116, 33.4%), and communication (*n* = 120, 34.6%), there were more participants who responded “not at all” or ‘small’. The median score for all items was 3.

Table 4 presents the frequencies and percentages of participants’ responses on the accuracy items. The most frequently chosen answer for all items on the scale was “to some extent.” Thus, the mean score for all the items was about 3 with a standard deviation of 1. Additionally, the median score for all items was 3. Seven out of the 10 accuracy items have more responses for ‘large’ and ‘very large’. For assessment strategies (*n* = 103, 29.60%) and assessment use (*n* = 106, 30.50%), there were more participants who responded “not at all” and ‘small’.

Table 4

Item Frequencies and Percentages for TKES – Accuracy

Item	1	2	3	4	5	*Mdn*	*M*	*SD*
Professional Knowledge	28 (8.1%)	64 (18.4%)	138 (39.8%)	90 (25.9%)	27 (7.8%)	3	3.07	1.04
Instructional Planning	35 (10.1%)	64 (18.4%)	137 (39.5%)	90 (25.9%)	21 (6.1%)	3	2.99	1.05
Instructional Strategies	31 (8.9%)	66 (19.0%)	138 (39.8%)	89 (25.6%)	23 (6.6%)	3	3.02	1.04
Differentiated Strategies	30 (8.6%)	63 (18.2%)	140 (40.3%)	91 (26.2%)	23 (6.6%)	3	3.04	1.03

Table 4

Item Frequencies and Percentages for TKES – Accuracy

Item	1	2	3	4	5	*Mdn*	*M*	*SD*
Assessment Strategies	30 (8.6%)	73 (21.0%)	144 (41.5%)	82 (23.6%)	18 (5.2%)	3	2.96	1.00
Assessment Use	31 (8.9%)	75 (21.6%)	144 (41.5%)	78 (22.5%)	19 (5.5%)	3	2.94	1.01
Positive Learning Environment	29 (8.4%)	62 (17.9%)	139 (40.1%)	97 (28.0%)	20 (5.8%)	3	3.05	1.01

Academically Challenging Environment	29 (8.4%)	60 (17.3%)	149 (42.9%)	89 (25.6%)	20 (5.8%)	3	3.03	1.00
Professionalism	36 (10.4%)	65 (18.7%)	145 (41.8%)	81 (23.3%)	20 (5.8%)	3	2.95	1.03
Communication	39 (11.2%)	58 (16.7%)	153 (44.1%)	78 (22.5%)	19 (5.5%)	3	2.94	1.03

Note: 1 = Not at all, 2 = To a small extent, 3 = To some extent, 4 = To a large extent, 5 = To a very large extent

Table 5 presents the frequencies and percentages of participants' responses on the TKES – creativity items. The most frequently chosen answer for each item on the scale was 'moderate' and "to a large extent." For all items, professional knowledge (n = 152, 43.8%), instructional planning (n = 155, 44.7%), instructional strategies (n = 157, 45.2%), differentiated strategies (n = 135, 38.9%), assessment strategies (n = 137, 39.5%), assessment use (n = 133, 38.3%), positive learning environment (n = 162, 46.7%), academically challenging environment (n = 144, 41.5%), and professionalism (n = 165, 47.5%), communication (n = 151, 43.5%) there were more participants who responded either 'large' or 'very large' than "not at all" or 'small'.

Table 5

Item Frequencies and Percentages for TKES – Creativity

Item	1	2	3	4	5	*Mdn*	*M*	*SD*
Professional Knowledge	21 (6.1%)	49 (14.1%)	125 (36.0%)	125 (36.0%)	27 (7.8%)	3	3.25	1.00

Instructional Planning	19 (5.5%)	43 (12.4%)	130 (37.5%)	129 (37.2%)	26 (7.5%)	3	3.29	0.97
Instructional Strategies	17 (4.9%)	47 (13.5%)	126 (36.3%)	130 (37.5%)	27 (7.8%)	4	3.30	0.97
Differentiated Strategies	21 (6.1%)	55 (15.9%)	136 (39.2%)	109 (31.4%)	26 (7.5%)	3	3.18	0.99
Assessment Strategies	21 (6.1%)	44 (12.7%)	145 (41.8%)	113 (32.6%)	24 (6.9%)	3	3.22	0.96
Assessment Use	16 (4.6%)	57 (16.4%)	141 (40.6%)	110 (31.7%)	23 (6.6%)	3	3.19	0.95
Positive Learning Environment	15 (4.3%)	49 (14.1%)	121 (34.9%)	128 (36.9%)	34 (9.8%)	4	3.34	0.98
Academically Challenging Environment	18 (5.2%)	50 (14.4%)	135 (38.9%)	115 (33.1%)	29 (8.4%)	3	3.25	0.98
Professionalism	16 (4.6%)	47 (13.5%)	119 (34.3%)	129 (37.2%)	36 (10.4%)	4	3.35	0.99
Communication	17 (4.9%)	49 (14.1%)	130 (37.5%)	119 (34.3%)	32 (9.2%)	3	3.29	0.98

Note: 1 = Not at all, 2 = To a small extent, 3 = To some extent, 4 = To a large extent, 5 = To a very large extent

For professional knowledge, instructional planning, differentiated strategies, assessment strategies, assessment use, academically challenging environment, and communication the median score was 3. For instructional strategies, positive learning, and professionalism the median score was 4.

Analyzing the Correlations Among the Independent and Dependent Variables

A Spearman's correlation analysis was conducted to determine whether there are correlations between demographic characteristics, TKES, TAPS, SGP, perceived influence, perceived accuracy, and perceived level of creativity (see Table 6). Correlations among the variables ranged from -.203 between content area and gender to .688 between SGP and perceived influence. There is a significant negative correlation between gender and grade level ($r_s(345) = -.188$, $p < .01$), and a significant positive correlation between gender and perceived accuracy ($r_s(345) = .163$, $p < .01$). The strength of both these associations is weak. There are positive significant correlations between the TKES and TAPS ($r_s(345) = .667$, $p < .01$) and SGP ($r_s(345) = .638$, $p < .01$). The magnitude of these associations is strong. A moderate significant correlation exists between the TKES and perceived influence ($r_s(345) = .588$, $p < .01$). There is a weak to moderate significant positive association between the TKES and perceived accuracy ($r_s(345) = .393$, $p < .01$) and perceived level of creativity ($r_s(345) = .372$, $p < .01$). There is a strong significant positive correlation between the TAPS and SGP ($r_s(345) = .636$, $p < .01$). There is a moderate significant positive correlation between the TAPS and perceived influence ($r_s(345) = .589$, $p < .01$). There is also a weak to moderate significant positive association between the TAPS and perceived accuracy ($r_s(345) = .345$, $p < .01$), and perceived level of creativity ($r_s(345) = .334$, $p < .01$). The SGP is significantly

positively correlated with perceived influence ($r_s(345) = .688$, $p < .01$). The magnitude of this association is strong. There is a moderate significant positive correlation between the SGP and perceived accuracy ($r_s(345) = .481$, $p < .01$), and a weak to moderate significant positive association between the SGP and perceived level of creativity ($r_s(345) = .386$, $p < .01$). Moreover, the perceived influence is positively correlated with perceived accuracy ($r_s(345) = .523$, $p < .01$) and perceived level of creativity ($r_s(345) = .473$, $p < .01$). The magnitude of these associations is moderate. There is also a strong significant positive relationship between perceived accuracy and perceived level of creativity ($r_s(345) = .652$, $p < .01$).

Table 6

Spearman's Rho Correlations among Independent and Dependent Variables

	1	2	3	4	5	6	7	8	9	10	11	12
1. Gender	-											
2. Ethnicity	.042	-										
3. Grade Level	-.188**	-.062	-									
4. Educational Level	-.045	-.079	.101	-								
5. Number of Years Certified as a Teacher	.039	.018	.038	.372**	-							
6. Content Area	-.203**	.074	.300**	.062	.058	-						
7. TKES	.061	-.098	-.040	.127*	.147**	-.033	-					
8. TAPS	.035	-.080	-.136*	.113*	.034	-.097	.667**	-				
9. SGP	.078	-.105	-.082	.139**	.100	-.099	.638**	.636**	-			
10. Influence	.065	-.041	-.030	.016	.011	-.103	.588**	.589**	.688**	-		
11. Accuracy	.163**	.039	-.125*	-.035	-.068	-.060	.393**	.345**	.481**	.523**	-	
12. Creativity	.110*	-.009	-.090	-.040	-.030	-.064	.372**	.334**	.386**	.473**	.652**	-

Note. $^{*}p < .05$, $^{**}p < .01$.

Reliability Analysis

Cronbach's alpha was used to conduct a reliability analysis for Perceived Influence, Perceived Accuracy, Perceived Level of Creativity, Total Scale (i.e., a summed total for perceived influence, perceived accuracy, and perceived level of creativity), and Total Scale plus TKES, TAPS, and SGP (33 items total). The results revealed that each section of the survey (a = .97, .98, and .98), the whole survey (a = .97), and the survey plus TKES, TAPS, and SGP (a = .97) were reliable as the reliability coefficients for all these scales were well above the acceptable level of .70.

Evaluation of the Research Questions

This section presents the results of evaluating the research questions and their corresponding hypotheses. Multiple two-way ANOVAs were conducted to evaluate Research Questions 1 through 3. In addition, a series of one-way ANOVAs were performed to evaluate Research Questions 4 through 7. Before presenting these results from these analyses, their underlying statistical assumptions were assessed. These assumptions are:

1) There are no missing observations in the data.
2) There should be no outliers in the data.
3) The data should be normally distributed for each level of the independent variables.
4) The variances of the dependent variable should be equal across the levels of the independent variables. This assumption is referred to as the homogeneity of variances assumption.

5) There should be independence of observations.

6) The dependent variables should be measured on an interval or ratio scale

Before running the analyses, the researcher ensured that the data did not contain missing values. Thus, the first assumption was valid. In addition, the fifth assumption was considered valid based on the sampling design utilized in this study. Assumption 6 was also deemed valid as all three dependent variables of perceived accuracy, perceived influence, and perceived creativity allowed by the TKES assessment were assumed to be measured on interval scales. Assumptions 2 through 4 were assessed using different statistical procedures described as follows.

Then, *z*-scores were also calculated to identify significant outliers. Values greater than 3.29 or less than -3.29 were considered as outliers and were checked for accuracy. The normality assumption was evaluated by conducting a series of Shapiro-Wilk tests for each category of the independent variables. Levene's test of equality of variances was employed to examine the fourth assumption.

Research Question 1

RQ1: Is there a significant difference by gender and race or ethnicity on selected dependent variables?

RQ1A: Is there a significant difference by gender and race or ethnicity on teachers' perceived influence of the TKES on teaching practices?

The two independent variables in this question are gender (male and female) and race (White and teachers of color). The dependent variable is perceived level of influence of the TKES on teaching practices. Descriptive statistics were calculated on both independent variables and the overall sample. Regarding gender, males had a range from

10 to 50 with a mean of 28.24 (*SD* = 9.21) and females had a range from 10 to 50 with a mean of 29.83 (*SD* = 9.77). Regarding race, Whites had a range from 10 to 50 with a mean of 29.37 (*SD* = 9.69), and teachers of color had a range from 10 to 49 with a mean of 30.60 (*SD* = 9.55). The overall sample had a range from 10 to 50 with a mean of 29.52 (*SD* = 9.67). Descriptive statistics were also calculated for the interaction between gender and race. Male teachers of color had a range from 12 to 46 with a mean of 28.10 (*SD* = 11.955). On the other hand, female teachers of color had a range from 10 to 49 with a mean of 31.38 (*SD* = 8.732).

Table 7

Descriptive Statistics for Perceived Influence of the TKES on Teaching Practices by Gender and Race

Gender	Race	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Male	Teachers of Color	10	28.10	11.99	12	46	0.17	-1.46
	White	57	28.26	8.77	10	50	0.11	0.13
	Total	67	28.24	9.21	10	50	0.12	-0.25
Female	Teachers of Color	32	31.38	8.73	10	49	-0.53	0.06
	White	248	29.63	9.89	10	50	-0.19	-0.31
	Total	280	29.83	9.77	10	50	-0.23	-0.29
Total	Teachers of Color	42	30.60	9.55	10	49	-0.37	-0.59
	White	305	29.37	9.69	10	50	-0.14	-0.27
	Total	347	29.52	9.67	10	50	-0.16	-0.32

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived influence of the TKES on teaching practices between males and females, between Whites and teachers of color, and between any of the combinations of these groups. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality,

independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived influence of the TKES on teaching practices for each category of gender and race to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. The results of this test indicated evidence of normal distribution for male teachers of color ($SW(10) = .939$, $p = .547$), White males ($SW(57) = .974$, $p = .247$), and female teachers of color ($SW(32) = .969$, $p = .471$). However, there was evidence that the normality assumption was violated for White females as the Shapiro-Wilk test showed significant results for this group of participants ($SW(248) = .966$, $p < .001$).

Due to the normality assumption being violated for one out of the four subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. It was found that the data were rendered closest to a normal distribution, as determined based on the Shapiro-Wilk test, for a λ of 0.9 in the Box-Cox transformation method.

Based on the results of the Shapiro-Wilk test for the transformed data, no significant departures from normality were indicated for male teachers of color ($SW(10) = .939$, $p = .543$), White males ($SW(57) = .971$, $p = .178$), and female teachers of color ($SW(32) = .975$, $p = .660$). However, the results of these tests were significant for White females ($SW(248) = .969$, $p < .001$). The results from Levene's tests of equality of variances indicated that the homogeneity of variances assumption of the two-factor ANOVA was satisfied, $F(3,343) = 1.083$, $p = .356$.

The results of the two-factor ANOVA conducted based on the transformed data revealed that there was no significant interaction between gender and race ($F(1,343) = 0.204$, $p = .652$). In addition, the main effect of race ($F(1,343) = 0.188$, $p = .665$) or gender ($F(1,343) = 1.437$, $p = .231$) on perceived influence of the TKES on teaching practices were not statistically significant. These results indicated that there were no statistically significant differences in mean perceived influence of the TKES on teaching practices between males and females, between Whites and teachers of color, and between any combinations of these categorical variables. Hence the answer to RQ1a is that there is no difference by gender and race or race on teachers' perceived influence of the TKES on teaching practices.

RQ1B: Is there a significant difference by gender and race or ethnicity on teachers' perceived accuracy of the TKES assessment?

The two independent variables in this question are gender (male and female) and race (White and teachers of color). The dependent variable is perceived accuracy of the TKES assessment. Descriptive statistics were calculated on both independent variables and the overall sample. Regarding gender, males had a range from 10 to 50 with a mean of 29.72 ($SD = 9.260$) and females had a range from 10 to 50 with a mean of 33.36 ($SD = 8.623$). Regarding race, Whites had a range from 10 to 50 with a mean of 32.79 ($SD = 8.742$), and teachers of color had a range from 10 to 50 with a mean of 31.74 ($SD = 9.690$). The overall sample had a range from 10 to 50 with a mean of 32.66 ($SD = 8.854$). Descriptive statistics were also calculated for the interaction between gender and race. Male teachers of color had a range from 10 to 44 with a mean of 29.10 ($SD = 12.041$). On

the other hand, White females had a range from 10 to 50 with a mean of 33.47 (*SD* = 8.600).

Table 8

Descriptive Statistics for Perceived Accuracy of the TKES Assessment by Gender and Race

Gender	Race	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Male	Teachers of Color	10	29.10	12.041	10	44	-0.490	-1.125
	White	57	29.82	8.814	10	50	0.033	-0.161
	Total	67	29.72	9.260	10	50	-0.117	-0.340
Female	Teachers of Color	32	32.56	8.897	10	50	-0.343	0.337
	White	248	33.47	8.600	10	50	-0.384	0.159
	Total	280	33.36	8.623	10	50	-0.378	0.152
Total	Teachers of Color	42	31.74	9.690	10	50	-0.495	-0.029
	White	305	32.79	8.742	10	50	-0.304	-0.023
	Total	347	32.66	8.854	10	50	-0.339	-0.016

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived accuracy of the TKES assessment between males and females, between Whites and teachers of color, and between any of the combinations of these groups. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived accuracy of the TKES assessment for each category of gender and race to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for male teachers of color (*SW*(10) =

.922, $p = .372$), White males ($SW(57) = .974$, $p = .254$), and female teachers of color ($SW(32) = .963$, $p = .336$). On the other hand, the Shapiro-Wilk test showed significant results for White females ($SW(248) = .956$, $p < .001$).

Due to the normality assumption being violated for one out of the four subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. It was found that the data were rendered closest to a normal distribution, as determined based on the Shapiro-Wilk test, for a λ of 0.8 in the Box-Cox transformation method.

Based on the results of the Shapiro-Wilk test for the transformed data, no significant departures from normality were indicated for male teachers of color ($SW(10) = .933$, $p = .477$), White males ($SW(57) = .967$, $p = .115$), and female teachers of color ($SW(32) = .968$, $p = .448$). However, the results of these tests were significant for White females ($SW(248) = .96$, $p < .001$). The results from Levene's tests of equality of variances indicated that the homogeneity of variances assumption of the two-factor ANOVA was satisfied for the transformed data, $F(3,343) = 0.769$, $p = .512$.

The results of the two-factor ANOVA conducted based on the transformed data revealed that there was no significant interaction between gender and race ($F(1,343) = 0.012$, $p = .913$). In addition, the main effect of race on perceived accuracy of the TKES assessment was not statistically significant ($F(1,343) = 0.171$, $p = .679$). These results indicated that there were no statistically significant differences in mean perceived accuracy of the TKES assessment between Whites and teachers of color, and between any combinations of race and gender. On the other hand, effect of gender on perceived accuracy of the TKES assessment was found to be statistically significant, $F(1,343) =$

3.928, $p = .048$, $\eta^2 = .011$. It can be concluded from these results that males reported significantly lower scores on perceived accuracy of the TKES assessment than females. The partial eta squared value of .011 indicated that the effect of gender on the outcome variable was very small. Hence the answer to RQ1b is that there was a significant difference in teachers' perceived accuracy of the TKES assessment on teaching practices by gender with males reporting significantly higher scores than females, but there was no significant difference by race or the interaction between race and gender.

RQ1C: Is there a significant difference by gender and race or ethnicity on the perceived level of creativity allowed by the TKES assessment?

The two independent variables in this question are gender (male and female) and race (White and teachers of color). The dependent variable is perceived level of creativity allowed by the TKES assessment. Descriptive statistics were calculated on both independent variables and the overall sample. Regarding gender, males had a range from 10 to 50 with a mean of 27.87 ($SD = 9.571$) and females had a range from 10 to 50 with a mean of 30.51 ($SD = 9.370$). Regarding race, Whites had a range from 10 to 50 with a mean of 29.96 ($SD = 9.465$), and teachers of color had a range from 10 to 50 with a mean of 30.24 ($SD = 9.471$). The overall sample had a range from 10 to 50 with a mean of 30.00 ($SD = 9.453$). Descriptive statistics were also calculated for the interaction between gender and race. Male teachers of color had a range from 10 to 40 with a mean of 25.60 ($SD = 12.249$) while female teachers of color had a range from 10 to 50 with a mean of 31.69 ($SD = 8.122$).

Table 9

Descriptive Statistics for Perceived Level of Creativity Allowed by the TKES Assessment by Gender and Race

Gender	Race	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Male	Teachers of Color	10	25.60	12.249	10	40	0.118	-1.947
	White	57	28.26	9.098	10	50	-0.008	-0.035
	Total	67	27.87	9.571	10	50	-0.045	-0.455
Female	Teachers of Color	32	31.69	8.122	10	50	-0.413	0.727
	White	248	30.35	9.523	10	50	-0.226	-0.141
	Total	280	30.51	9.370	10	50	-0.251	-0.083
Total	Teachers of Color	42	30.24	9.471	10	50	-0.473	-0.304
	White	305	29.96	9.465	10	50	-0.179	-0.167
	Total	347	30.00	9.453	10	50	-0.212	-0.199

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived level of creativity allowed by the TKES assessment between males and females, between Whites and teachers of color, and between any of the combinations of these groups. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived level of creativity allowed by the TKES assessment for each category of gender and race to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for male teachers of color($SW(10) = .853$, $p = .063$), and female teachers of color ($SW(32) = .961$, $p = .286$). On the other hand, the Shapiro-Wilk test showed

significant results for White males ($SW(57) = .956$, $p = .038$) and White females ($SW(248) = .961$, $p < .001$). The results from Levene's tests of equality of variances indicated that the homogeneity of variances assumption of the two-factor ANOVA was met, $F(3,343) = 1.460$, $p = .225$.

Due to the normality assumption being violated for two out of the four subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. However, it was found that no improvements could be achieved in the normality of the data using this transformation method. Hence, this question was evaluated using the raw data.

The results of the two-factor ANOVA revealed that there was no significant interaction between gender and race ($F(1,343) = 1.178$, $p = .279$). In addition, the main effect of race on perceived level of creativity allowed by the TKES assessment was not statistically significant ($F(1,343) = 0.131$, $p = .718$). These results indicated that there were no statistically significant differences in the mean perceived level of creativity allowed by the TKES assessment between Whites and teachers of color and between any combinations of race and gender. On the other hand, it was found that the effect of gender on perceived level of creativity allowed by the TKES assessment was statistically significant, $F(1,343) = 4.934$, $p < .05$, $\eta^2 = .014$. It can be concluded from these results that males reported significantly lower scores on perceived level of creativity allowed by the TKES assessment than females. The partial eta squared value of .014 indicated that the effect of gender on the outcome variable was very small. Hence the answer to RQ1c is that there was a significant difference in teachers' perceived level of creativity allowed by the TKES assessment by gender with males reporting significantly lower scores than

females, but there was no significant difference by race or the interaction between race and gender.

Research Question 2

RQ2: Is there a significant difference by grade level taught and level of education on selected dependent variables?

RQ2A: Is there a significant difference by grade level taught and level of education on teachers' perceived influence of the TKES on teaching practices?

The two independent variables in this question are grade level taught (pre-K to K, grades 1-3, grades 4-5, grades 6-8, grades 9-12, and other) and level of education (bachelor's degree, master's degree, educational specialist's degree, and doctoral degree). The dependent variable is perceived level of influence of the TKES on teaching practices. Descriptive statistics were calculated on both independent variables and the overall sample.

Regarding grade level, the scores for pre-K to K had a range from 18 to 50 with a mean of 33.92 (*SD* = 8.797), for grades 1-3 ranged from 10 to 50 with a mean of 29.27 (*SD* = 8.945), for grades 4-5 ranged from 14 to 50 with a mean of 30.21 (*SD* = 10.674), for grades 6-8 had a range from 10 to 50 with a mean of 27.87 (*SD* = 9.977), for grades 9-12 ranged from 10 to 49 with a mean of 27.63 (*SD* = 8.735), and for other grades ranged from 10 to 50 with a mean of 34.03 (*SD* = 9.593). Regarding educational level, the scores for bachelor's degree ranged from 10 to 50 with a mean of 29.72 (*SD* = 9.527), for master's degree ranged from 10 to 50 with a mean of 29.30 (*SD* = 9.641), for educational specialist's degree ranged from 10 to 50 with a mean of 29.22 (*SD* = 9.783), and for doctoral degree ranged from 14 to 50 with a mean of 31.73 (*SD* = 10.194).

The scores for pre-K to K and bachelor's degree ranged from 27 to 50 with a mean of 35.67 (*SD* = 9.605). The scores for grades 1-3 and master's degree ranged from 11 to 44 with a mean of 31 (*SD* = 8.024). The scores for grades 4-5 and master's degree ranged from 10 to 50 with a mean of 30.67 (*SD* = 10.924). The scores for grades 6-8 and doctorate degree ranged from 21 to 43 with a mean of 32.13 (*SD* = 8.476). The scores for grades 9- 12 and doctorate degree ranged from 20 to 38 with a mean of 29.5 (*SD* = 7.416). The scores for other grades and educational specialist's degree ranged from 29 to 50 with a mean of 37.33 (*SD* = 5.327).

Table 10

Descriptive Statistics for Perceived Influence of the TKES on Teaching Practices by Grade Level and Educational Level

Grade Level	Educational Level	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Elementary: Pre-K to K	Bachelor's degree	9	35.67	9.605	27	50	0.678	-1.295
	Master's degree	10	33.80	7.642	41	47	0.051	-0.205
	Educational Specialist's degree	5	31.00	10.559	18	47	0.652	1.488
	Total	24	33.92	8.797	18	50	0.380	-0.564
Elementary: Gr1-3	Bachelor's degree	17	28.12	10.428	10	48	-0.114	-0.445
	Master's degree	22	31.00	8.024	11	44	-1.069	1.562
	Educational Specialist's degree	22	28.41	8.953	11	50	0.415	0.451
	Doctorate degree	1	30.00	.	30	30	.	.
	Total	62	29.27	8.945	10	50	-0.237	-0.025
Elementary: Gr4-5	Bachelor's degree	6	28.67	9.585	10	38	-1.917	4.465
	Master's degree	21	30.67	10.924	10	50	-0.109	-0.213
	Educational Specialist's degree	23	30.43	10.268	10	49	-0.315	-0.371
	Doctorate degree	3	28.33	19.088	14	50	1.474	.
	Total	53	30.21	10.674	10	50	-0.158	-0.418

Table 10

Descriptive Statistics for Perceived Influence of the TKES on Teaching Practices by Grade Level and Educational Level

Grade Level	Educational Level	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Middle School: Gr6-8	Bachelor's degree	20	28.55	9.545	11	49	0.068	-0.083
	Master's degree	25	28.60	9.456	10	50	-0.253	0.879
	Educational Specialist's degree	22	24.86	11.171	10	46	0.265	-0.94
	Doctorate degree	8	32.13	8.476	21	43	0.21	-1.728
	Total	75	27.87	9.977	10	50	-0.068	-0.406
High School: Gr9-12	Bachelor's degree	21	29.10	8.921	10	49	-0.184	0.966
	Master's degree	45	26.93	9.233	10	40	-0.392	-0.867
	Educational Specialist's degree	24	27.33	8.090	10	41	-0.328	-0.262
	Doctorate degree	4	29.50	7.416	20	38	-0.392	1.233
	Total	94	27.63	8.735	10	49	-0.335	-0.362
Other	Bachelor's degree	6	32.50	8.597	22	45	0.442	-0.927
	Master's degree	12	30.33	12.471	10	50	-0.161	-1.091
	Educational Specialist's degree	15	37.33	5.327	29	50	0.518	1.141
	Doctorate degree	6	34.67	11.622	20	46	-0.387	-2.159
	Total	39	34.03	9.593	10	50	-0.581	-0.135
Total	Bachelor's degree	79	29.72	9.527	10	50	-0.074	0.037
	Master's degree	135	29.30	9.641	10	50	-0.295	-0.23
	Educational Specialist's degree	111	29.22	9.783	10	50	-0.15	-0.503
	Doctorate degree	22	31.73	10.194	14	50	0.161	-1.065
	Total	347	29.52	9.672	10	50	-0.163	-0.324

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived influence of the TKES on teaching practices by grade level, educational level, and any of the combinations of grade level and educational level. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality,

independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived influence of the TKES on teaching practices for each category of grade level and educational level to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for pre-K to K and master's degree ($SW(10) = .983$, $p = .978$), pre-K to K and educational Specialist's degree ($SW(5) = .963$, $p = .828$), grades 1-3 and bachelor's degree ($SW(17) = .975$, $p = .903$), grades 1-3 and educational Specialist's degree ($SW(22) = .954$, $p = .379$), grades 4-5 and master's degree ($SW(21) = .967$, $p = .661$), grades 4-5 and educational Specialist's degree ($SW(23) = .977$, $p = .845$), grades 4-5 and doctorate degree ($SW(3) = .889$, $p = .352$), grades 6-8 and bachelor's degree ($SW(20) = .976$, $p = .875$), grades 6-8 and educational Specialist's degree ($SW(22) = .941$, $p = .205$), grades 6-8 and doctorate degree ($SW(8) = .901$, $p = .293$), grades 9-12 and bachelor's degree ($SW(21) = .964$, $p = .609$), grades 9-12 and educational Specialist's degree ($SW(24) = .975$, $p = .779$), grades 9-12 and doctorate degree ($SW(4) = .978$, $p = .89$), other grades and bachelor's degree ($SW(6) = .948$, $p = .726$), other grades and master's degree ($SW(12) = .948$, $p = .611$), other grades and educational Specialist's degree ($SW(15) = .95$, $p = .529$), and other grades and doctorate degree ($SW(6) = .864$, $p = .205$). On the other hand the results of these tests were significant for pre-K to K and bachelor's degree ($SW(9) = .815$, $p = .03$), grades 1-3 and master's degree ($SW(22) = .888$, $p = .018$), grades 4-5 and bachelor's degree ($SW(6) = .75$, $p = .02$), grades 6-8 and master's degree ($SW(25) = .918$, $p = .046$), and grades 9-12 and master's degree ($SW(45) = .935$, $p = .014$).

Due to the normality assumption being violated for five out of the 22 subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. It was found that the data were rendered closest to a normal distribution, as determined based on the Shapiro-Wilk test, for a λ of 0.9 in the Box-Cox transformation method.

Based on the results of the Shapiro-Wilk test for the transformed data, no significant departures from normality were indicated for pre-K to K and master's degree ($SW(10) = .981$, $p = .972$), pre-K to K and educational specialist's degree ($SW(5) = .953$, $p = .758$), grades 1-3 and bachelor's degree ($SW(17) = .977$, $p = .928$), grades 1-5 and educational specialist's degree ($SW(22) = .946$, $p = .265$), grades 4-6 and master's degree ($SW(21) = .967$, $p = .656$), grades 4-7 and educational specialist's degree ($SW(23) = .982$, $p = .938$)grades 4-8 and doctoral degree ($SW(3) = .879$, $p = .322$), and grades 6-8 and bachelor's degree ($SW(20) = .974$, $p = .836$). Moreover, no significant deviations from normality were observed for grades 6-9 and master's degree ($SW(25) = .921$, $p = .053$), grades 6-10 and educational specialist's degree ($SW(22) = .94$, $p = .195$), grades 6-11 and doctoral degree ($SW(8) = .898$, $p = .275$), grades 9-12 and bachelor's degree ($SW(21) = .964$, $p = .608$), grades 9-14 and educational specialist's degree ($SW(24) = .977$, $p = .842$), and grades 9-15 and doctoral degree ($SW(4) = .98$, $p = .903$). In addition, there was no evidence suggesting violations of normality for other grades and bachelor's degree ($SW(6) = .944$, $p = .695$), other grades and master's degree ($SW(12) = .953$, $p = .676$), other grades and educational specialist's degree ($SW(15) = .938$, $p = .36$), and other grades and doctoral degree ($SW(6) = .868$, $p = .217$). However, the results of these tests were significant for pre-K to K and bachelor's degree ($SW(9) = .81$, $p = .026$), grades 1-4

and master's degree ($SW(22) = .9$, $p = .03$), grades 4-5 and bachelor's degree ($SW(6) = .761$, $p = .025$), and grades 9-13 and master's degree ($SW(45) = .939$, $p = .020$). The results from Levene's test of equality of variances indicated that the homogeneity of variances assumption was met for perceived influence of the TKES on teaching practices for the transformed data, $F(21,324) = 1.225$, $p = .227$.

The results of the two-factor ANOVA conducted based on the transformed data indicated that there was no significant interaction between grade level and educational level ($F(14,324) = .664$, $p = .809$). These results indicated that there were no statistically significant differences in the mean perceived level of influence of the TKES on teaching practices between any combinations of grade levels and educational levels. In addition, the main effect of educational level on perceived influence of the TKES on teaching practices was not statistically significant ($F(3,324) = 0.143$, $p = .934$). It can be concluded from these results that there was no significant difference in the mean of transformed perceived influence of the TKES on teaching practices scores between those who had a bachelor's degree, master's degree, educational specialist's degree, and doctoral degree. However, the main effect of grade level was found to be statistically significant, $(5,324) = 2.496$, $p = .031$, $\eta^2 = .037$.

Post hoc comparisons were conducted using Turkey's HSD procedure for this ANOVA to determine which grade level means significantly differed. These results indicated that teachers who taught students in grades pre-kindergarten to kindergarten reported significantly higher mean scores on perceived influence of the TKES on teaching practices compared to the teachers who taught High School students in grades 9 – 12 ($p = .043$). In addition, teachers who taught students in other grades ascribed

significantly higher scores on perceived influence of the TKES on teaching practices compared to the teachers who taught middle school students in grades 6-8 and high school students in grades 9-12 (p = .014). No other significant differences were identified. Hence the answer to RQ2a is that there was a statistically significant difference in teachers' perceived influence of the TKES on teaching practices by grade level, while there was no significant difference by educational level, or the interaction between grade level and educational level.

RQ2B: Is there a significant difference by grade level taught and level of education on teachers' perceived accuracy of the TKES assessment?

The two independent variables in this question are grade level taught (pre-K to K, grades 1-3, grades 4-5, grades 6-8, grades 9-12, and other) and level of education (bachelor's degree, master's degree, educational specialist's degree, and doctoral degree). The dependent variable is perceived accuracy of the TKES assessment. Descriptive statistics were calculated on both independent variables and the overall sample.

Regarding grade level, the scores for pre-K to K had a range from 20 to 50 with a mean of 36.46 (SD = 7.690), for grades 1-3 ranged from 10 to 50 with a mean of 33.44 (SD = 8.399), for grades 4-5 ranged from 12 to 50 with a mean of 33.77 (SD = 9.063), for grades 6-8 had a range from 11 to 50 with a mean of 32.03 (SD = 9.458), for grades 9-12 ranged from 10 to 49 with a mean of 30.89 (SD = 7.966), and for other grades ranged from 10 to 50 with a mean of 33.05 (SD = 10.118). Regarding educational level, the scores for bachelor's degree ranged from 10 to 50 with a mean of 34.08 (SD = 9.284), for master's degree ranged from 10 to 50 with a mean of 32.05 (SD = 8.639), for educational

specialist's degree ranged from 10 to 50 with a mean of 31.95 (*SD* = 8.832), and for doctoral degree ranged from 20 to 50 with a mean of 34.86 (*SD* = 8.305).

The scores for pre-K to K and educational specialist's degree ranged from 33 to 47 with a mean of 38.8 (*SD* = 5.404). The scores for grades 1-3 and doctorate degree ranged from 40 to 40 with a mean of 40 (*SD* =0). The scores for grades 4-5 and bachelor's degree ranged from 30 to 41 with a mean of 37.00 (*SD* = 4.472). The scores for grades 6-8 and master's degree ranged from 14 to 50 with a mean of 33.84 (*SD* = 9.56).

Table 11

Descriptive Statistics for Perceived Accuracy of the TKES Assessment by Grade Level and Educational Level

Grade Level	Educational Level	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Elementary: Pre-K to K	Bachelor's degree	9	36.44	9.787	20	50	-0.016	-0.293
	Master's degree	10	35.30	6.961	29	50	1.087	0.577
	Educational Specialist's degree	5	38.80	5.404	33	47	0.800	0.596
	Total	24	36.46	7.690	20	50	0.226	-0.15
Elementary: Gr1-3	Bachelor's degree	17	34.53	8.854	13	50	-0.513	0.89
	Master's degree	22	32.91	7.521	20	48	-0.104	-0.597
	Educational Specialist's degree	22	32.82	9.215	10	50	-0.5	0.54
	Doctorate degree	1	40.00	0	40	40	0	0
	Total	62	33.44	8.399	10	50	-0.403	0.209
Elementary: Gr4-5	Bachelor's degree	6	37.00	4.472	30	41	-0.966	-0.867
	Master's degree	21	32.76	10.281	12	50	-0.225	-0.381
	Educational Specialist's degree	23	33.48	8.398	15	48	-0.594	-0.326
	Doctorate degree	3	36.67	14.048	22	50	-0.423	0
	Total	53	33.77	9.063	12	50	-0.422	-0.26
Middle School: Gr6-8	Bachelor's degree	20	32.40	10.932	11	50	-0.46	-0.109
	Master's degree	25	33.84	9.560	14	50	0.083	-0.3
	Educational Specialist's degree	22	29.82	8.926	10	42	-0.741	-0.021

	Doctorate degree	8	31.50	6.414	20	40	-0.227	0.767
	Total	75	32.03	9.458	11	50	-0.266	0.024
High School: Gr9-12	Bachelor's degree	21	33.48	9.163	10	50	-0.739	1.034
	Master's degree	45	30.04	6.691	17	43	-0.213	-0.448
	Educational Specialist's degree	24	30.38	9.098	10	41	-0.935	0.275
	Doctorate degree	4	30.00	7.439	20	38	-0.787	1.848
	Total	94	30.89	7.966	10	50	-0.508	0.226

Table 11

Descriptive Statistics for Perceived Accuracy of the TKES Assessment by Grade Level and Educational Level

Grade Level	Educational Level	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Other	Bachelor's degree	6	34.00	9.839	21	45	-0.556	-1.735
	Master's degree	12	30.33	12.324	10	50	-0.183	-1.077
	Educational Specialist's degree	15	31.73	8.811	14	50	-0.028	0.706
	Doctorate degree	6	40.83	5.742	33	50	0.471	0.83
	Total	39	33.05	10.118	10	50	-0.401	-0.434
Total	Bachelor's degree	79	34.08	9.284	10	50	-0.56	0.282
	Master's degree	135	32.05	8.639	10	50	-0.043	-0.151
	Educational Specialist's degree	111	31.95	8.832	10	50	-0.596	0.126
	Doctorate degree	22	34.86	8.305	20	50	-0.079	-0.233
	Total	347	32.66	8.854	10	50	-0.339	-0.016

The scores for grades 9- 12 and bachelor's degree ranged from 10 to 50 with a mean of 33.48 (*SD* = 9.163). The scores for other grades and doctorate degree ranged from 33 to 50 with a mean of 40.83 (*SD* = 5.742).

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived accuracy of the TKES assessment by grade level, educational level, and any of the combinations of grade level and educational level. Statistical

assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived accuracy of the TKES assessment for each category of grade level and educational level to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for pre-K to K and bachelor's degree ($SW(9) = .948$, $p = .667$), pre-K to K and educational Specialist's degree ($SW(5) = .946$, $p = .709$), grades 1-3 and bachelor's degree ($SW(17) = .936$, $p = .274$), grades 1-3 and master's degree ($SW(22) = .932$, $p = .133$), grades 1-3 and educational Specialist's degree ($SW(22) = .969$, $p = .68$), grades 4-5 and bachelor's degree ($SW(6) = .849$, $p = .155$), grades 4-5 and master's degree ($SW(21) = .97$, $p = .738$), grades 4-5 and educational Specialist's degree ($SW(23) = .932$, $p = .122$), grades 4-5 and doctorate degree ($SW(3) = .993$, $p = .843$), grades 6-8 and bachelor's degree ($SW(20) = .934$, $p = .182$), grades 6-8 and master's degree ($SW(25) = .955$, $p = .318$), grades 6-8 and educational Specialist's degree ($SW(22) = .926$, $p = .1$), grades 6-8 and doctorate degree ($SW(8) = .887$, $p = .222$), grades 9-12 and bachelor's degree ($SW(21) = .944$, $p = .258$), grades 9-12 and doctorate degree ($SW(4) = .922$, $p = .548$), other grades and bachelor's degree ($SW(6) = .881$, $p = .272$), other grades and master's degree ($SW(12) = .94$, $p = .5$), other grades and educational Specialist's degree ($SW(15) = .97$, $p = .862$), and other grades and doctorate degree ($SW(6) = .965$, $p = .86$). On the contrary, the results of these tests were significant for pre-K to K and master's degree ($SW(10) = .816$, $p = .023$), grades 9-12 and master's

degree ($SW(45) = .939$, $p = .019$), and grades 9-12 and educational Specialist's degree ($SW(24) = .884$, $p = .01$). The results from Levene's test of equality of variances indicated that the homogeneity of variances assumption was met for perceived accuracy of the TKES assessment $F(21,324) = 1.280$, $p = .186$.

Due to the normality assumption being violated for three out of the 22 subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. However, it was found that no improvements could be achieved in the normality of the data using this transformation method. Hence, this question was evaluated using the raw data.

The results of the two-factor ANOVA revealed that there was no significant interaction between educational level and grade level ($F(1,324) = .641$, $p = .830$). These results indicated that there were no statistically significant differences in the mean of perceived accuracy between any combinations of grade levels and educational levels. Additionally, the main effect of educational level ($F(1,324) = 1.410$, $p = .240$) or grade level ($F(1,324) = 2.247$, $p = .050$) on perceived accuracy of the TKES assessment was not statistically significant. These results indicated that there were no statistically significant differences in the mean perceived accuracy of the TKES assessment between different educational levels or between different grade levels. Thus, the answer to RQ2b is that there was no significant difference in teachers' perceived accuracy of the TKES assessment by educational level, grade level, or the interaction between these categorical variables.

RQ2C: Is there a significant difference by grade level taught and level of education on the perceived level of creativity allowed by the TKES assessment?

The two independent variables in this question are grade level taught (pre-K to K, grades 1-3, grades 4-5, grades 6-8, grades 9-12, and other) and level of education (bachelor's degree, master's degree, educational specialist's degree, and doctoral degree). The dependent variable is perceived level of creativity allowed by the TKES assessment. Descriptive statistics were calculated on both independent variables and the overall sample.

Regarding grade level, the scores for pre-K to K had a range from 23 to 50 with a mean of 35.08 (*SD* = 8.439), for grades 1-3 ranged from 10 to 50 with a mean of 29.81 (*SD* = 10.106), for grades 4-5 ranged from 10 to 50 with a mean of 29.75 (*SD* = 9.667), for grades 6-8 had a range from 10 to 50 with a mean of 30.15 (*SD* = 9.456), for grades 9-12 ranged from 10 to 40 with a mean of 28.78 (*SD* = 8.630), and for other grades ranged from 10 to 50 with a mean of 30.15 (*SD* = 10.101). Regarding educational level, the scores for bachelor's degree ranged from 10 to 50 with a mean of 30.51 (*SD* = 9.079), for master's degree ranged from 10 to 50 with a mean of 30.38 (*SD* = 9.197), for educational specialist's degree ranged from 10 to 50 with a mean of 29.00 (*SD* = 9.703), and for doctoral degree ranged from 10 to 50 with a mean of 30.86 (*SD* = 11.192).

Table 12

Descriptive Statistics for Perceived Level of Creativity Allowed by the TKES Assessment by Grade Level and Educational Level

Grade Level	Educational Level	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Elementary: Pre-K to K	Bachelor's degree	9	36.00	9.287	25	50	0.650	-1.016
	Master's degree	10	34.00	8.014	23	50	0.690	0.312
	Educational Specialist's degree	5	35.60	9.343	26	50	0.995	0.636
	Total	24	35.08	8.439	23	50	0.655	-0.626
	Bachelor's degree	17	29.65	9.354	15	48	0.264	-0.802

Elementary: Gr1-3	Master's degree	22	30.18	8.483	10	43	-1.117	1.383
	Educational Specialist's degree	22	28.64	11.705	10	50	0.125	-0.766
	Doctorate degree	1	50.00	.	50	50	.	.
	Total	62	29.81	10.106	10	50	-0.115	-0.39
Elementary: Gr4-5	Bachelor's degree	6	34.17	4.355	30	40	0.245	-2.377
	Master's degree	21	29.86	10.451	10	50	-0.307	-0.354
	Educational Specialist's degree	23	28.48	9.110	10	44	-0.626	-0.013
	Doctorate degree	3	30.00	17.321	20	50	1.732	.
	Total	53	29.75	9.667	10	50	-0.308	-0.164

Table 12

Descriptive Statistics for Perceived Level of Creativity Allowed by the TKES Assessment by Grade Level and Educational Level

Grade Level	Educational Level	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Middle School: Gr6-8	Bachelor's degree	20	29.35	9.799	10	50	-0.047	0.175
	Master's degree	25	33.28	8.354	19	49	0.000	-0.839
	Educational Specialist's degree	22	27.64	10.495	10	50	0.165	-0.378
	Doctorate degree	8	29.25	7.573	20	40	0.224	-1.006
	Total	75	30.15	9.456	10	50	-0.055	-0.373
High School: Gr9-12	Bachelor's degree	21	29.67	8.410	12	40	-0.754	0.086
	Master's degree	45	28.62	8.367	10	40	-0.682	0.127
	Educational Specialist's degree	24	29.00	9.113	10	40	-0.917	0.422
	Doctorate degree	4	24.50	11.958	10	37	-0.356	-1.957
	Total	94	28.78	8.630	10	40	-0.729	-0.044
Other	Bachelor's degree	6	27.83	10.553	20	48	1.840	3.488
	Master's degree	12	29.17	12.791	10	50	-0.006	-0.768

	Educational Specialist's degree	15	30.13	7.530	18	43	0.068	-0.664
	Doctorate degree	6	34.50	10.672	13	41	-2.310	5.456
	Total	39	30.15	10.101	10	50	-0.09	-0.584
Total	Bachelor's degree	79	30.51	9.079	10	50	0.013	-0.14
	Master's degree	135	30.38	9.197	10	50	-0.381	0.041
	Educational Specialist's degree	111	29.00	9.703	10	50	-0.191	-0.273
	Doctorate degree	22	30.86	11.19	10	50	-0.112	-0.7812
	Total	347	30.00	9.453	10	50	-0.212	-0.199

The scores for pre-K to K and bachelor's degree ranged from 25 to 50 with a mean of 36 (*SD* = 9.287). The scores for grades 1-3 and doctorate degree ranged from 50 to 50 with a mean of 50 (*SD* =0). The scores for grades 4-5 and bachelor's degree ranged from 30 to 40 with a mean of 34.17 (*SD* = 4.355). The scores for grades 6-8 and master's degree ranged from 19 to 49 with a mean of 33.28 (*SD* = 8.354). The scores for grades 9-12 and bachelor's degree ranged from 12 to 40 with a mean of 29.67 (*SD* = 8.41). The scores for other grades and doctorate degree ranged from 13 to 41 with a mean of 34.5 (*SD* = 10.672).

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived level of creativity allowed by the TKES assessment by grade level, educational level, and any of the combinations of grade level and educational level. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived level of creativity allowed by the

TKES assessment for each category of grade level and educational level to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for pre-K to K and bachelor's degree ($SW(9) = .883$, $p = .17$), pre-K to K and master's degree ($SW(10) = .943$, $p = .583$), pre-K to K and educational Specialist's degree ($SW(5) = .941$, $p = .67$), grades 1-3 and bachelor's degree ($SW(17) = .96$, $p = .64$), grades 1-3 and educational Specialist's degree ($SW(22) = .958$, $p = .448$), grades 4-5 and bachelor's degree ($SW(6) = .84$, $p = .13$), grades 4-5 and master's degree ($SW(21) = .949$, $p = .326$), grades 4-5 and educational Specialist's degree ($SW(23) = .92$, $p = .066$), and grades 6-8 and bachelor's degree ($SW(20) = .969$, $p = .73$). In addition, the normality assumption was deemed valid for grades 6-8 and master's degree ($SW(25) = .941$, $p = .152$), grades 6-8 and educational specialist's degree ($SW(22) = .958$, $p = .45$), grades 6-8 and doctorate degree ($SW(8) = .898$, $p = .276$), grades 9-12 and doctorate degree ($SW(4) = .969$, $p = .834$), other grades and master's degree ($SW(12) = .962$, $p = .81$), and other grades and educational specialist's degree ($SW(15) = .953$, $p = .577$). However, the results of these tests were significant for grades 1-3 and master's degree ($SW(22) = .888$, $p = .017$), grades 4-5 and doctorate degree ($SW(3) = .75$, $p < .001$), grades 9-12 and bachelor's degree ($SW(21) = .886$, $p = .019$), grades 9-12 and master's degree ($SW(45) = .915$, $p = .003$), grades 9-12 and educational Specialist's degree ($SW(24) = .869$, $p = .005$), other grades and bachelor's degree ($SW(6) = .783$, $p = .041$), and other grades and doctorate degree ($SW(6) = .648$, $p = .002$). The results from Levene's test of equality of variances indicated that

the homogeneity of variances assumption was met for perceived level of creativity allowed by the TKES assessment $F(21,324) = 0.940$, $p = .539$.

Due to the normality assumption being violated for seven out of the 22 subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. However, it was found that no improvements could be achieved in the normality of the data using this transformation method. Hence, this question was evaluated using the raw data.

The results of the two-factor ANOVA indicated that there was no significant interaction effect between educational level and grade level ($F(1,324) = .895$, $p = .565$). These results indicated that there were no statistically significant differences in the mean of perceived creativity between any combinations of grade levels and educational levels. Moreover, the main effect of educational level was not statistically significant ($F(1,324) = .992$, $p = .397$). It can be concluded from these results that there was no significant difference in the mean of perceived level of creativity allowed by the TKES between different educational levels. On the other hand, it was found that the main effect of grade level was significant, $F(5,324) = 2.548$, $p < .05$, $\eta^2 = .038$. Post hoc comparisons were conducted using Turkey's HSD procedure to determine which grade level means differed significantly from one another. The results from this analysis revealed that teachers who taught students in grades pre-kindergarten to kindergarten reported significantly lower mean scores on perceived level of creativity allowed by the TKES compared to the teachers who taught High School students in grades 9 – 12 ($p < .05$). No other significant differences were identified. Thus, the answer to RQ2c is that there was a significant difference in teachers' perceived level of creativity allowed by the TKES by grade level,

while there were no significant differences by educational level, or the interaction between grade level and educational level.

Research Question 3

RQ3: Is there a significant difference by years of experience as a certified teacher in Georgia and educational level on selected dependent variables?

RQ3A: Is there a significant difference by years of experience as a certified teacher in Georgia and educational level on teachers' perceived influence of the TKES on teaching practices?

The two independent variables in this question are number of years certified as a teacher (1-5 years, 6-10 years, 11-15 years, 16-20 years, and 21+ years) and level of education (bachelor's degree, master's degree, educational specialist's degree, and doctoral degree). The dependent variable is perceived level of influence of the TKES on teaching practices. Descriptive statistics were calculated on both independent variables and the overall sample.

Table 13

Descriptive Statistics for Perceived Influence of the TKES on Teaching Practices by Number of Years Certified as a Teacher and Educational Level

Educational Level	Number of Years Certified as a Teacher	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Bachelor's degree	1-5 years	25	28.32	8.240	12	49	0.376	0.494
	6-10 years	15	33.47	10.260	10	50	-0.254	1.018
	11-15 years	16	31.63	11.337	10	49	-0.459	-0.125
	16-20 years	9	26.22	8.743	10	34	-1.315	0.346
	21+ years	14	28.29	8.686	11	41	-0.33	-0.35
	Total	79	29.72	9.527	10	50	-0.074	0.037
Master's degree	1-5 years	26	31.38	9.449	11	50	-0.45	0.09

	6-10 years	20	28.80	9.956	10	50	-0.209	0.351
	11-15 years	21	25.67	11.051	10	50	0.219	-0.4
	16-20 years	26	31.19	8.045	10	41	-0.831	0.722
	21+ years	42	28.88	9.605	10	50	-0.188	-0.076
	Total	135	29.30	9.641	10	50	-0.295	-0.230
Educational Specialist's degree	1-5 years	4	19.25	6.652	11	27	-0.21	0.548
	6-10 years	7	29.57	9.778	16	43	-0.147	-1.092
	11-15 years	21	27.67	10.679	10	49	-0.112	-0.488
	16-20 years	32	30.41	10.105	10	50	-0.426	-0.198
	21+ years	47	29.89	9.159	10	50	-0.013	-0.349
	Total	111	29.22	9.783	10	50	-0.150	-0.503
Doctorate degree	6-10 years	1	21.00	.	21	21	.	.
	11-15 years	3	36.00	15.100	20	50	-0.586	.
	16-20 years	7	32.43	7.976	24	46	1.058	-0.16
	21+ years	11	31.09	10.802	14	46	-0.109	-1.371
	Total	22	31.73	10.194	14	50	0.161	-1.065
Total	1-5 years	55	29.11	9.162	11	50	-0.007	-0.261
	6-10 years	43	30.37	10.019	10	50	-0.111	0.053
	11-15 years	61	28.43	11.266	10	50	-0.029	-0.676
	16-20 years	74	30.36	9.050	10	50	-0.519	0.150
	21+ years	114	29.44	9.341	10	50	-0.101	-0.398
	Total	347	29.52	9.672	10	50	-0.163	-0.324

Regarding number of years certified as a teacher, the scores for 1-5 years ranged from 11 to 50 with a mean of 29.11 (*SD* = 9.162), for 6-10 years ranged from 10 to 50 with a mean of 30.37 (*SD* = 10.019), for 11-15 years ranged from 10-50 with a mean of 28.43 (*SD* = 11.266), for 16-20 years ranged from 10 to 50 with a mean of 30.36 (*SD* = 9.050), and for 20+ years ranged from 10 to 50 with a mean of 29.44 (*SD* = 9.341).

The scores for bachelor's degree and 6-10 years ranged from 10 to 50 with a mean of 33.47 (*SD* = 10.26). The scores for master's degree and 1-5 years ranged from 11 to 50 with a mean of 31.38 (*SD* = 9.449). The scores for educational specialist's degree and 16-

20 years ranged from 10 to 50 with a mean of 30.41 (*SD* = 10.105). The scores for doctoral degree and 11-15 years ranged from 20 to 50 with a mean of 36 (*SD* = 15.1).

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived level of influence of the TKES on teaching practices by number of years certified as a teacher, educational level, and any of the combinations of number of years as a certified teacher and educational level. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived level of influence of the TKES on teaching practices for each category of number of years certified as a teacher and educational level to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for1-5 years and bachelor's degree (*SW*(25) = .98, *p* = .887), 1-5 years and master's degree (*SW*(26) = .961, *p* = .415), 1-5 years and educational specialist's degree (*SW*(4) = .995, *p* = .983), 6-10 years and bachelor's degree (*SW*(15) = .933, *p* = .303), 6-10 years and master's degree (*SW*(20) = .941, *p* = .251), 6-10 years and educational specialist's degree (*SW*(7) = .968, *p* = .886), 11-15 years and bachelor's degree (*SW*(16) = .949, *p* = .476), 11-15 years and master's degree (*SW*(21) = .948, *p* = .313), 11-15 years and educational specialist's degree (*SW*(21) = .959, *p* = .503), 11-15 years and doctorate degree (*SW*(3) = .987, *p* = .78), 16-20 years and educational specialist's degree (*SW*(32) = .962, *p* = .306), 16-20 years and doctorate degree (*SW*(7) = .857, *p* = .144), 21+ years and bachelor's degree (*SW*(14) =

.969, p = .856), 21+ years and master's degree (SW(42) = .967, p = .264), 21+ years and educational specialist's degree (SW(47) = .984, p = .771), and 21+ years and doctorate degree (SW(11) = .938, p = .495). On the other hand, the results of these tests were significant for 16-20 years and bachelor's degree (SW(9) = .798, p = .019), 16-20 years and master's degree (SW(26) = .897, p = .014). Based on the results from Levene's test of equality of variances, there was no significant deviation from the homogeneity of variances assumption for perceived influence of the TKES on teaching practices, F(17,328) = 0.517, p = .944.

Due to the normality assumption being violated for two out of the 18 subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. However, it was found that no improvements could be achieved in the normality of the data using this transformation method. Hence, this question was evaluated using the raw data.

The results of the two-factor ANOVA revealed that there was no significant interaction between educational level and number of years certified as a teacher (F(1,328) = 1.444, p = .152). These results indicated that there were no statistically significant differences in the mean perceived level of influence between any combinations of educational levels and number of years certified as a teacher. Additionally, the main effect of educational level (F(1,328) = 0.610, p = .609) or number of years certified as a teacher (F(1,328) = 0.711, p = .585) on perceived level of influence was not statistically significant. These results indicated that there were no statistically significant differences in the mean perceived level of influence of the TKES on teaching practices between different educational levels or between different number of years certified as a teacher.

Hence, the answer to RQ3a is that there was no significant difference in teachers' perceived influence of the TKES on teaching practices by educational level, number of years certified as a teacher, or the interaction between these categorical variables.

RQ3B: Is there a significant difference by years of experience as a certified teacher in Georgia and educational level on teachers' perceived accuracy of the TKES assessment?

The two independent variables in this question are number of years certified as a teacher (1-5 years, 6-10 years, 11-15 years, 16-20 years, and 21+ years) and level of education (bachelor's degree, master's degree, educational specialist's degree, and doctoral degree). The dependent variable is perceived accuracy of the TKES assessment. Descriptive statistics were calculated on both independent variables and the overall sample.

Regarding number of years certified as a teacher, the scores for 1-5 years ranged from 10 to 50 with a mean of 33.13 (*SD* = 9.349), for 6-10 years ranged from 12 to 50 with a mean of 34.00 (*SD* = 8.255), for 11-15 years ranged from 10-50 with a mean of 32.95 (*SD* = 9.601), for 16-20 years ranged from 10 to 50 with a mean of 32.55 (*SD* = 7.995), and for 20+ years ranged from 10 to 50 with a mean of 31.84 (*SD* = 9.010).

The scores bachelor's degree and 6-10 years ranged from 20 to 50 with a mean of 37.73 (*SD* = 7.914). The scores for master's degree and 1-5 years ranged from 19 to 50 with a mean of 34.85 (*SD* = 8.79).

Table 14

Descriptive Statistics for Perceived Accuracy of the TKES Assessment by Number of Years Certified as a Teacher and Educational Level

Educational Level	Number of Years Certified as a Teacher	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Bachelor's degree	1-5 years	25	32.40	9.256	13	50	-0.016	-0.12
	6-10 years	15	37.73	7.914	20	50	-0.442	0.596
	11-15 years	16	36.31	10.480	10	50	-1.135	1.534
	16-20 years	9	29.33	8.201	11	40	-1.359	3.03
	21+ years	14	33.64	8.984	11	45	-1.301	1.832
	Total	79	34.08	9.284	11	50	-0.560	0.282
Master's degree	1-5 years	26	34.85	8.790	19	50	0.097	-0.569
	6-10 years	20	32.40	6.924	20	50	0.557	1.022
	11-15 years	21	29.48	9.564	17	50	0.328	-0.768
	16-20 years	26	32.77	6.993	18	43	-0.577	-0.457
	21+ years	42	31.00	9.502	10	50	-0.093	0.107
	Total	135	32.05	8.639	10	50	-0.043	-0.151
Educational Specialist’s degree	1-5 years	4	26.50	12.477	10	40	-0.683	1.286
	6-10 years	7	30.00	10.693	12	40	-0.835	-0.47
	11-15 years	21	32.57	7.985	14	48	-0.373	0.284
	16-20 years	32	32.41	8.976	10	50	-0.997	1.166
	21+ years	47	32.13	8.714	10	50	-0.301	-0.271
	Total	111	31.95	8.832	10	50	-0.596	0.126
Doctorate degree	6-10 years	1	38.00	.	38	38	.	.
	11-15 years	3	42.00	6.928	38	50	1.732	.
	16-20 years	7	36.57	5.769	29	44	-0.209	-1.969
	21+ years	11	31.55	9.213	20	50	0.554	0.162
	Total	22	34.86	8.305	20	50	-0.079	-0.233
Total	1-5 years	55	33.13	9.349	10	50	-0.132	-0.122
	6-10 years	43	34.00	8.255	12	50	-0.281	0.366
	11-15 years	61	32.95	9.601	10	50	-0.279	-0.465
	16-20 years	74	32.55	7.995	10	50	-0.917	1.02
	21+ years	114	31.84	9.010	10	50	-0.25	-0.138
	Total	347	32.66	8.854	10	50	-0.163	-0.324

The scores for educational specialist’s degree and 11-15 years ranged from 14 to 48 with a mean of 32.57 (*SD* = 7.985). The scores for doctorate degree and 11-15 years ranged from 38 to 50 with a mean of 42 (*SD* = 6.928).

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived accuracy of the TKES assessment by number of years certified as a teacher, educational level, and any of the combinations of number of years as a certified teacher and educational level. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived accuracy of the TKES assessment for each category of number of years certified as a teacher and educational level to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for 1-5 years and bachelor's degree ($SW(25) = .953$, $p = .298$), 1-5 years and master's degree ($SW(26) = .957$, $p = .343$), 1-5 years and educational specialist's degree ($SW(4) = .97$, $p = .841$), 6-10 years and bachelor's degree ($SW(15) = .933$, $p = .304$), 6-10 years and master's degree ($SW(20) = .927$, $p = .137$), 6-10 years and educational specialist's degree ($SW(7) = .873$, $p = .198$), 11-15 years and bachelor's degree ($SW(16) = .904$, $p = .092$), 11-15 years and master's degree ($SW(21) = .93$, $p = .138$), 11-15 years and educational specialist's degree ($SW(21) = .975$, $p = .843$), 16-20 years and bachelor's degree ($SW(9) = .891$, $p = .203$), 16-20 years and doctorate degree ($SW(7) = .873$, $p = .196$), 21+ years and master's degree ($SW(42) = .951$, $p = .069$), 21+ years and educational specialist's degree ($SW(47) = .978$, $p = .503$), and 21+ years and doctorate degree ($SW(11) = .924$, $p = .35$). However, the results of these tests were significant for 11-15 years and doctorate degree ($SW(3) = .75$, $p < .001$), 16-20

years and master's degree ($SW(26) = .913$, $p = .031$), 16-20 years and educational specialist's degree ($SW(32) = .885$, $p = .003$), and 21+ years and bachelor's degree ($SW(14) = .869$, $p = .04$).

Due to the normality assumption being violated for four out of the 18 subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. It was found that the data were rendered closest to a normal distribution, as determined based on the Shapiro-Wilk test, for a λ of 0.7 in the Box-Cox transformation method.

Based on the results of the Shapiro-Wilk test for the transformed data, no significant departures from normality were indicated for bachelor's degree and 1-5 years ($SW(25) = .936$, $p = .12$), bachelor's degree and 6-10 years ($SW(15) = .931$, $p = .282$), bachelor's degree and 11-15 years ($SW(16) = .947$, $p = .445$), bachelor's degree and 16-20 years ($SW(9) = .923$, $p = .419$), bachelor's degree and 21+ years ($SW(14) = .903$, $p = .126$), master's degree and 1-5 years ($SW(26) = .935$, $p = .104$), master's degree and 11-15 years ($SW(21) = .913$, $p = .063$), and master's degree and 16-20 years ($SW(26) = .925$, $p = .058$). Moreover, there were no significant deviations from normality for educational specialist's degree and 1-5 years ($SW(4) = .984$, $p = .924$), educational specialist's degree and 6-10 years ($SW(7) = .883$, $p = .24$), educational specialist's degree and 11-15 years ($SW(21) = .981$, $p = .936$), educational specialist's degree and 21+ years ($SW(47) = .983$, $p = .738$), doctoral degree and 16-20 years ($SW(7) = .879$, $p = .221$), and doctoral degree and 21+ years ($SW(11) = .888$, $p = .132$). However, the results of these tests were significant for master's degree and 6-10 years ($SW(20) = .882$, $p = .019$), master's degree and 21+ years ($SW(42) = .936$, $p = .021$), educational specialist's degree and 16-20 years

($SW(32) = .914$, $p = .015$), and doctoral degree and 11-15 years ($SW(3) = .75$, $p < .001$). Based on the results from Levene's test of equality of variances, there was no significant deviation from the homogeneity of variances assumption for the transformed data, $F(17,328) = 0.430$, $p = .978$.

The results of the two-factor ANOVA conducted based on the transformed data revealed that there was no significant interaction between educational level and number of years certified as a teacher ($F(1,328) = 1.526$, $p = .120$). These results indicated that there were no statistically significant differences in the mean perceived accuracy between any combinations of educational levels and number of years certified as a teacher. Moreover, the main effect of educational level ($F(1,328) = 2.273$, $p = .080$) or number of years certified as a teacher ($F(1,328) = 0.878$, $p = .477$) on perceived accuracy of the TKES assessment was not statistically significant. These results indicated that there were no statistically significant differences in the mean perceived accuracy of the TKES assessment between different educational levels or between different number of years certified as a teacher. Therefore, the answer to RQ3b is that there was no significant difference in teachers' perceived accuracy of the TKES assessment by educational level, number of years certified as a teacher, or the interaction between these categorical variables.

RQ3C: Is there a significant difference by years of experience as a certified teacher in Georgia and educational level on the perceived level of creativity allowed by the TKES assessment?

The two independent variables in this question are number of years certified as a teacher (1-5 years, 6-10 years, 11-15 years, 16-20 years, and 21+ years) and level of

education (bachelor's degree, master's degree, educational specialist's degree, and doctoral degree). The dependent variable is perceived level of creativity allowed by the TKES assessment.

Table 15

Descriptive Statistics for Perceived Level of Creativity Allowed by the TKES Assessment by Number of Years Certified as a Teacher and Educational Level

Educational Level	Number of Years Certified as a Teacher	*N*	Mean	*SD*	Min	Max	Skewness	Kurtosis
Bachelor's degree	1-5 years	25	28.16	9.003	12	50	0.325	0.032
	6-10 years	15	35.40	9.326	20	50	0.265	-0.959
	11-15 years	16	31.19	8.841	12	48	-0.251	0.405
	16-20 years	9	27.44	8.748	10	40	-0.786	1.108
	21+ years	14	30.64	8.289	13	42	-0.639	0.005
	Total	79	30.51	9.079	13	50	0.013	-0.140
Master's degree	1-5 years	26	33.69	8.694	10	50	-0.528	1.076
	6-10 years	20	28.70	7.915	10	40	-0.435	0.081
	11-15 years	21	26.10	12.482	10	49	0.119	-1.328
	16-20 years	26	33.00	5.953	17	41	-0.514	0.331
	21+ years	42	29.64	9.004	10	50	-0.04	0.656
	Total	135	30.38	9.197	10	50	-0.381	0.041
Educational Specialist's degree	1-5 years	4	21.25	8.539	10	30	-0.753	0.343
	6-10 years	7	26.71	11.528	10	40	-0.377	-1.215
	11-15 years	21	29.71	10.479	10	48	-0.577	-0.225
	16-20 years	32	30.00	9.745	10	50	-0.199	0.471
	21+ years	47	29.00	9.184	10	50	0.019	-0.335
	Total	111	29.00	9.703	10	50	-0.191	-0.273
Doctorate degree	6-10 years	1	20.00	.	20	20	.	.
	11-15 years	3	42.00	7.000	37	50	1.574	.
	16-20 years	7	33.86	11.112	20	50	-0.125	-0.983
	21+ years	11	26.91	10.329	10	41	-0.252	-0.948
	Total	22	30.86	11.192	10	50	-0.112	-0.781
Total	1-5 years	55	30.27	9.423	10	50	-0.102	-0.179
	6-10 years	43	30.51	9.563	10	50	-0.033	-0.096
	11-15 years	61	29.46	11.061	10	50	-0.312	-0.74

	16-20 years	74	31.11	8.671	10	50	-0.376	0.563
	21+ years	114	29.24	9.053	10	50	-0.113	-0.13
	Total	347	30.00	9.453	10	50	-0.212	-0.199

Descriptive statistics were calculated on both independent variables and the overall sample. Regarding number of years certified as a teacher, the scores for 1-5 years ranged from 10 to 50 with a mean of 30.27 (*SD* = 9.423), for 6-10 years ranged from 10 to 50 with a mean of 30.51 (*SD* = 9.563), for 11-15 years ranged from 10-50 with a mean of 29.46 (*SD* = 11.061), for 16-20 years ranged from 10 to 50 with a mean of 31.11 (*SD* = 8.671), and for 20+ years ranged from 10 to 50 with a mean of 29.24 (*SD* = 9.053).

The scores for bachelor's degree and 6-10 years ranged from 20 to 50 with a mean of 35.4 (*SD* = 9.326). For master's degree and 1-5 years ranged from 10 to 50 with a mean of 33.69 (*SD* = 8.694). For educational specialist's degree and 16-20 years ranged from 10 to 50 with a mean of 30.00 (*SD* = 9.745). For doctoral degree and 11-15 years ranged from 37 to 50 with a mean of 42 (*SD* = 7).

A factorial ANOVA was conducted to determine whether there was a significant difference in perceived level of creativity allowed by the TKES assessment by number of years certified as a teacher, educational level, and any of the combinations of number of years as a certified teacher and educational level. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived level of creativity allowed by the TKES assessment for each category of number of years certified as a teacher and educational level to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was

used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for 1-5 years and bachelor's degree ($SW(25) = .951$, $p = .26$), 1-5 years and master's degree ($SW(26) = .955$, $p = .304$), 1-5 years and educational specialist's degree ($SW(4) = .971$, $p = .85$), 6-10 years and bachelor's degree ($SW(15) = .918$, $p = .182$), 6-10 years and master's degree ($SW(20) = .951$, $p = .38$), 6-10 years and educational specialist's degree ($SW(7) = .922$, $p = .485$), 11-15 years and bachelor's degree ($SW(16) = .976$, $p = .921$), 11-15 years and master's degree ($SW(21) = .912$, $p = .059$), 11-15 years and educational specialist's degree ($SW(21) = .929$, $p = .129$), 11-15 years and doctorate degree ($SW(3) = .862$, $p = .274$), 16-20 years and bachelor's degree ($SW(9) = .956$, $p = .759$), 16-20 years and doctorate degree ($SW(7) = .914$, $p = .425$), 21+ years and bachelor's degree ($SW(14) = .949$, $p = .547$), 21+ years and master's degree ($SW(42) = .95$, $p = .064$), 21+ years and educational specialist's degree ($SW(47) = .966$, $p = .19$), and 21+ years and doctorate degree ($SW(11) = .951$, $p = .656$). On the contrary, the results of these tests were significant for 16-20 years and master's degree ($SW(26) = .851$, $p = .002$), and 16-20 years and educational specialist's degree ($SW(32) = .928$, $p = .034$). Based on the results from Levene's test of equality of variances, there was no significant deviation from the homogeneity of variances assumption for perceived level of creativity allowed by the TKES assessment, $F(17,328) = 1.184$, $p = .275$.

Due to the normality assumption being violated for two out of the 18 subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. However, it was found that

no improvements could be achieved in the normality of the data using this transformation method. Hence, this question was evaluated using the raw data.

The results of the two-factor ANOVA indicated that there here was a significant interaction effect between educational level and number of years certified as a teacher ($F(1,328) = 2.515$, $p = .005$). A post hoc test was conducted to determine where these differences had occurred. It was found that among the teachers who had a master's degree, those who had 11-15 years of experience reported significantly lower scores on perceived level of creativity allowed by the TKES assessment than those with 1-5 and 16-20 years of teaching experience ($p < .05$). Moreover, among the teachers who had a doctorate degree, those who had 11-15 years of experience as a certified teacher reported significantly higher scores than those with 6-12 and 21+ years of teaching experience ($p < .05$). Furthermore, among teachers with a bachelor's degree, those who had 6-10 years of experience reported significantly higher scores than those with 1-5 and 16-20 years of teaching experience ($p < .05$). On the other hand, the main effect of number of years certified as a teacher ($F(1,328) = 1.334$, $p = .257$) or educational level ($F(1,328) = 1.470$, $p = .223$) was not statistically significant. It can be concluded from these results that there was no significant difference in the mean of perceived level of creativity allowed by the TKES assessment between different educational levels, or between different numbers of years certified as a teacher. Thus, the answer to RQ3c is that there was a significant difference in teachers' perceived creativity allowed by the TKES by the interaction between number of years certified as a teacher and educational level, while no significant differences were identified between different numbers of years certified as a teacher, or between different educational levels.

Research Question 4

RQ4: Is there a significant difference by content area taught on the selected dependent variables?

RQ4A: Is there a significant difference by content area taught on teachers' perceived influence of the TKES on teaching practices?

The independent variable in this question is content area (elementary grades, general, special education, English language arts, mathematics, social studies, vocation, career, or technical education, science, and other). The dependent variable is perceived level of influence of the TKES on teaching practices. Descriptive statistics were calculated on the independent variable and the overall sample.

The scores elementary grades, general ranged from 10 to 50 with a mean of 32.10 (*SD* = 10.29), for special education ranged from 10 to 47 with a mean of 30.48 (*SD* = 9.18), for English language arts ranged from 10 to 49 with a mean of 28.20 (*SD* = 8.94), for mathematics ranged from 10 to 50 with a mean of 29.12 (*SD* = 10.63), for social studies ranged from 10 to 46 with a mean of 27.21 (*SD* = 9.70), for vocation, career, or technical education ranged from 10 to 41 with a mean of 26.90 (*SD* = 9.33), for science ranged from 10 to 49 with a mean of 27.62 (*SD* = 8.49), and for other content areas ranged from 10 to 50 with a mean of 29.60 (*SD* = 9.33).

Table 16

Descriptive Statistics for Perceived Influence of the TKES on Teaching Practices by Content Area

	N	Mean	*SD*	Min	Max	Skewness	Kurtosis
Elementary grades, general	67	32.10	10.29	10	50	-0.356	-0.056
Special education	51	30.78	9.18	10	47	-0.532	-0.221

English Language Arts	54	28.20	8.94	10	49	-0.124	0.393
Mathematics	51	29.12	10.63	10	50	-0.148	-0.786
Social Studies	24	27.21	9.70	10	46	-0.339	-0.586
Vocation, career, or technical education	21	26.90	9.33	10	41	-0.312	-0.631
Science	21	27.62	8.49	10	49	0.003	1.75
Other	58	29.60	9.33	10	50	0.152	-0.405
Total	347	29.52	9.67	10	50	-0.163	-0.324

It was planned to conduct a one-way ANOVA to determine whether there was a significant difference in perceived level of influence of the TKES on teaching practices by content area. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived level of influence of the TKES on teaching practices for each category of content area to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for English language arts ($SW(54) = .96$, $p = .069$), mathematics ($SW(51) = .963$, $p = .111$), social studies ($SW(24) = .946$, $p = .222$), vocation, career, or technical education ($SW(21) = .952$, $p = .373$), science ($SW(21) = .916$, $p = .071$), and other content areas ($SW(58) = .973$, $p = .213$). However, the results of these tests were significant for elementary grades, general ($SW(67) = .95$, $p = .01$) and special education ($SW(51) = .954$, $p = .047$).

Due to the normality assumption being violated for two out of the eight subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. It was found

that the data were rendered closest to a normal distribution, as determined based on the Shapiro-Wilk test, for a λ of 0.9 in the Box-Cox transformation method.

Based on the results of the Shapiro-Wilk test for the transformed data, no significant departures from normality were indicated for special education ($SW(51) = .961$, $p = .094$), English language arts ($SW(54) = .96$, $p = .069$), mathematics ($SW(51) = .966$, $p = .153$), social studies ($SW(24) = .951$, $p = .282$), vocation, career, or technical education ($SW(21) = .955$, $p = .423$), science ($SW(21) = .91$, $p = .056$), and other content areas ($SW(58) = .970$, $p = .156$). However, the results of these tests were significant for elementary grades, general ($SW(67) = .956$, $p = .017$).

Provided that even after applying the Box-Cox transformation method, there were still significant departures from the normality assumption of the one-way ANOVA, this analysis did not seem an appropriate parametric procedure to address RQ4a. Hence, this question was evaluated using a Kruskal-Wallis test, which is a nonparametric alternative to one-way ANOVA. The results of the Kruskal-Wallis test conducted based on the raw data indicated that the effect of content area was not statistically significant ($H(7) = 12.552$, $p = .084$). It can be concluded from these results that perceived level of influence of the TKES on teaching practices did not significantly vary between different types of content area. Hence, the answer to RQ4a is that there was no difference by content area on teachers' perceived influence of the TKES on teaching practices.

RQ4B: Is there a significant difference by content area taught on teachers' perceived accuracy of the TKES assessment?

The independent variable in this question is content area (elementary grades, general, special education, English language arts, mathematics, social studies, vocation,

career, or technical education, science, and other). The dependent variable is perceived accuracy of the TKES assessment. Descriptive statistics were calculated on the independent variable and the overall sample.

The scores elementary grades, general ranged from 10 to 50 with a mean of 33.75 (*SD* = 9.99), for special education ranged from 10 to 45 with a mean of 32.39 (*SD* = 8.14), for English language arts ranged from 10 to 50 with a mean of 32.37 (*SD* = 8.42), for mathematics ranged from 10 to 50 with a mean of 34.08 (*SD* = 9.26), for social studies ranged from 12 to 44 with a mean of 32.08 (*SD* = 7.66), for vocation, career, or technical education ranged from 11 to 40 with a mean of 29.33 (*SD* = 8.03), for science ranged from 19 to 44 with a mean of 32.24 (*SD* = 7.43), and for other content areas ranged from 10 to 50 with a mean of 32.26 (*SD* = 9.36).

Table 17

Descriptive Statistics for Perceived Accuracy of the TKES Assessment by Content Area

	N	Mean	*SD*	Min	Max	Skewness	Kurtosis
Elementary grades, general	67	33.75	9.99	10	50	-0.283	-0.366
Special education	51	32.39	8.14	10	45	-0.747	0.033
English Language Arts	54	32.37	8.42	10	50	-0.029	0.528
Mathematics	51	34.08	9.26	10	50	-0.53	0.551
Social Studies	24	32.08	7.66	12	44	-0.944	0.818
Vocation, career, or technical education	21	29.33	8.03	11	40	-0.636	0.027
Science	21	32.24	7.43	19	44	-0.392	-1.03
Other	58	32.26	9.36	10	50	-0.336	-0.241
Total	347	32.66	8.85	10	50	-0.163	-0.324

It was planned to conduct a one-way ANOVA to determine whether there was a significant difference in perceived accuracy of the TKES assessment by content area. Statistical assumptions were checked before performing this analysis. The statistical

assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived accuracy of the TKES assessment for each category of content area to investigate the presence of outliers in the data and no significant outliers were identified. The next assumption to check was the normality of the data. The results of the Shapiro-Wilk test indicated that there were not significant departures from normality for English Language Arts ($SW(54) = .958$, $p = .054$), social Studies ($SW(24) = .922$, $p = .066$), vocation, career, or technical education ($SW(21) = .936$, $p = .181$), or science ($SW(21) = .918$, $p = .08$). However, the results of these tests were significant for elementary grades, general ($SW(67) = .959$, $p = .028$), special education ($SW(51) = .903$, $p = .001$), mathematics ($SW(51) = .934$, $p = .007$), and other content areas ($SW(58) = .958$, $p = .043$).

Due to the normality assumption being violated for four out of the eight subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. However, it was found that no improvements could be achieved in the normality of the data using this transformation method. Hence, this question was evaluated using the raw data. A Kruskal-Wallis test was employed as the nonparametric alternative to the one-way ANOVA to evaluate whether there was a significant difference in perceived accuracy of the TKES assessment by content area.

The results of the Kruskal-Wallis test conducted based on the raw data indicated that the effect of content area was not statistically significant ($H(7) = 5.855$, $p = .557$). It can be concluded from these results that perceived accuracy of the TKES assessment did

not significantly differ between different types of content area. Hence, the answer to RQ4b is that there was no difference by content area on teachers' perceived accuracy of the TKES assessment.

RQ4C: Is there a significant difference by content area taught on teachers' perceived level of creativity allowed by the TKES assessment?

The independent variable in this question is content area (elementary grades, general, special education, English language arts, mathematics, social studies, vocation, career, or technical education, science, and other). The dependent variable is perceived level of creativity allowed by the TKES assessment. Descriptive statistics were calculated on the independent variable and the overall sample.

The scores elementary grades, general ranged from 10 to 50 with a mean of 30.12 (*SD* = 10.98), for special education ranged from 10 to 50 with a mean of 31.96 (*SD* = 9.38), for English language arts ranged from 10 to 50 with a mean of 29.30 (*SD* = 9.79), for mathematics ranged from 10 to 50 with a mean of 31.00 (*SD* = 9.22), for social studies ranged from 12 to 50 with a mean of 29.21 (*SD* = 8.27), for vocation, career, or technical education ranged from 10 to 40 with a mean of 28.00 (*SD* = 8.17), for science ranged from 10 to 42 with a mean of 29.05 (*SD* = 8.05), and for other content areas ranged from 10 to 50 with a mean of 29.29 (*SD* = 8.99).

Table 18

Descriptive Statistics for Perceived Level of Creativity Allowed by the TKES Assessment by Content Area

	N	Mean	*SD*	Min	Max	Skewness	Kurtosis
Elementary grades, general	67	30.12	10.98	10	50	-0.053	-0.385
Special education	51	31.96	9.38	10	50	-0.712	0.34

English Language Arts	54	29.30	9.79	10	50	-0.262	-0.366
Mathematics	51	31.00	9.22	10	50	-0.331	-0.132
Social Studies	24	29.21	8.27	12	50	0.147	1.125
Vocation, career, or technical education	21	28.00	8.17	10	40	-0.675	0.434
Science	21	29.05	8.05	10	42	-0.51	0.186
Other	58	29.29	8.99	10	50	-0.003	-0.323
Total	347	30.00	9.45	10	50	-0.212	-0.199

A one-way ANOVA was conducted to determine whether there was a significant difference in perceived level of creativity allowed by the TKES assessment by content area. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for perceived level of creativity allowed by the TKES assessment for each category of content area to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. These results did not indicate significant violations of the normality assumption for social studies ($SW(24) = .94$, $p = .16$), vocation, career, or technical education ($SW(21) = .934$, $p = .165$), science ($SW(21) = .954$, $p = .406$), or other content areas ($SW(58) = .966$, $p = .109$). However, the results of these tests were significant elementary grades, general ($SW(67) = .955$, $p = .017$), special education ($SW(51) = .93$, $p = .005$), English language arts ($SW(54) = .957$, $p = .049$), and mathematics ($SW(51) = .944$, $p = .018$).

Due to the normality assumption being violated for four out of the eight subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. It was found

that the data were rendered closest to a normal distribution, as determined based on the Shapiro-Wilk test, for a λ of 0.9 in the Box-Cox transformation method.

Based on the results of the Shapiro-Wilk test for the transformed data, no significant departures from normality were indicated for English language arts ($SW(54) = .962$, $p = .081$), social studies ($SW(24) = .933$, $p = .116$), vocation, career, or technical education ($SW(21) = .94$, $p = .216$), science ($SW(21) = .958$, $p = .481$), and other content areas ($SW(58) = .966$, $p = .106$). However, the results of these tests were significant for elementary grades, general ($SW(67) = .954$, $p = .015$), special education ($SW(51) = .94$, $p = .012$), mathematics ($SW(51) = .948$, $p = .026$).

Given that even after applying the Box-Cox transformation method, there were still significant deviations from the normality assumption of the one-way ANOVA, this analysis did not seem an appropriate parametric procedure to address RQ4c. Thus, this question was evaluated using a Kruskal-Wallis test. The results of the Kruskal-Wallis test conducted based on the raw data indicated that the effect of content area was not statistically significant ($H(7) = 6.121$, $p = .526$). It can be concluded from these results that perceived level of creativity allowed by the TKES did not significantly vary between different types of content area. Therefore, the answer to RQ4c is that there was no difference by content area on teachers' perceived creativity allowed by the TKES.

Research Question 5

RQ5: Is there a significant difference between level of effective training for the TKES on TKES teaching practices?

It was planned to conduct a one-way ANOVA to address this question. The dependent variable in this analysis was level of effective training for the TKES on

teaching practices and the independent variable was the ratings for the TKES training (small, moderate, large, very large). Descriptive statistics were calculated on the independent variable and the overall sample.

Table 19

Descriptive Statistics for Level of Effective Training for the TKES on Teaching Practices by the Ratings for the TKES Training

	N	Mean	*SD*	Min	Max	Skewness	Kurtosis
Small	61	20.11	7.497	10	37	0.246	-0.997
Moderate	131	28.08	7.901	10	50	-0.366	0.692
Large	123	32.55	7.793	10	48	-0.756	0.281
Very Large	32	41.69	7.908	26	50	-0.6	-0.935
Total	347	29.52	9.672	10	50	-0.163	-0.324

The scores for those who reported ‘small' ranged from 10 to 37 with a mean of 20.11 (*SD* = 7.497), for those who reported 'moderate’ ranged from 10 to 50 with a mean of 28.08 (*SD* = 7.901), for those who reported 'large' ranged from 10 to 48 with a mean of 32.55 (*SD* = 7.793), and for those who reported 'very large’ ranged from 26 to 50 with a mean of 41.69 (*SD* = 7.908).

It was planned to perform a one-way ANOVA to determine whether there was a significant difference in level of effective training for the TKES on teaching practices by the ratings for the TKES training. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for level of effective training for the TKES on teaching practices for each level of the ratings for

the TKES training to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, significant deviations from normality were observed for those who responded 'small' ($SW(61) = .931$, $p = .002$), 'moderate' ($SW(131) = .947$, $p < .001$), 'large' ($SW(123) = .940$, $p < .001$), and 'very large' ($SW(32) = .881$, $p = .002$).

Due to the normality assumption being violated for all four subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. However, it was found that no improvements could be achieved in the normality of the data using this transformation method. Hence, this model was evaluated using the raw data. A Kruskal-Wallis test was employed as the nonparametric alternative to the one-way ANOVA test to evaluate whether there was a significant difference in level of effective training for the TKES on teaching practices by the ratings for the TKES training.

The results of the Kruskal-Wallis test showed that the effect of the ratings for the TKES training was statistically significant ($H(3) = 121.435$, $p < .001$), indicating that was a statistically significant difference in level of effective training for the TKES on teaching practices between different ratings for the TKES training. To determine where these differences had occurred, a series of post hoc Mann-Whitney U tests were conducted. Using the Bonferroni correction method, the significance level was adjusted by dividing the original significance level of.05 by the number of pairs being assessed. Using this procedure, the significance level was set at $\alpha = .05/6 = .008$.

The results of post hoc Mann-Whitney U tests revealed that respondents who rated ‘small’ reported significantly higher scores compared to those who rated ‘moderate’ ($p < .001$), ‘large’ ($p < .001$), and ‘very large’ ($p < .001$). Similarly, respondents who rated ‘moderate’ reported significantly higher scores compared to those who rated ‘large’ ($p < .001$), and ‘very large’ ($p < .001$). In addition, respondents who rated ‘large’ reported significantly higher scores than those who rated ‘very large’ ($p < .001$).

Research Question 6

RQ6: Is there a significant difference between level of effective training for the TAPS on TKES teaching practices?

It was planned to conduct a one-way ANOVA to address this question. The dependent variable in this analysis was level of effective training for the TKES on teaching practices and the independent variable was the ratings for the TAPS training (not at all, small, moderate, large, very large). Descriptive statistics were calculated on the independent variable and the overall sample.

Table 20 shows the descriptive statistics for level of effective training for the TKES on teaching practices by the ratings for the TAPS training. The mean of this variable for those who reported 'not at all' ranged from 10 to 41 with a mean of 21.38 (*SD* = 10.591), for those who reported 'small’ ranged from 11 to 42 with a mean of 23.65 (*SD* = 7.726), for those who reported 'moderate' ranged from 10 to 50 with a mean of 28.74 (*SD* = 7.595), for those who reported 'large' ranged from 19 to 48 with a mean of 34.54 (*SD* = 6.226), and for those who reported 'very large' ranged from 31 to 50 with a mean of 43.74 (*SD* = 6.51).

Table 20

Descriptive Statistics for Level of Effective Training for the TKES on Teaching Practices by the Ratings for the TAPS Training

	N	Mean	*SD*	Min	Max	Skewness	Kurtosis
Not at all	50	21.38	10.591	10	41	0.412	-1.248
Small	46	23.65	7.726	11	42	0.444	-0.184
Moderate	137	28.74	7.595	10	50	-0.030	1.002
Large	91	34.54	6.226	19	48	-0.499	0.100
Very Large	23	43.74	6.510	31	50	-0.721	-0.711
Total	347	29.52	9.672	10	50	-0.163	-0.324

It was planned to perform a one-way ANOVA to determine whether there was a significant difference in level of effective training for the TKES on teaching practices by the ratings for the TAPS training. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for level of effective training for the TKES on teaching practices for each level of the ratings for the TAPS training to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for those who indicated 'small' ($SW(46) = .959$, $p = .109$). On the other hand, the results of these tests were significant for those who stated 'not at all' ($SW(50) = .875$, $p < .001$), 'moderate' ($SW(137) = .956$, $p < .001$), 'large' ($SW(91) = .951$, $p = .002$), and 'very large' ($SW(23) = .858$, $p = .004$).

Due to the normality assumption being violated for four out of the five subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of the data can be made more similar to the normal distribution. However, it was found that no improvements could be achieved in the normality of the data using this transformation method. Hence, this model was evaluated using the raw data. A Kruskal-Wallis test was employed as the nonparametric alternative to the one-way ANOVA test to evaluate whether there was a significant difference in level of effective training for the TKES on teaching practices by the ratings for the TAPS training.

The results of the Kruskal-Wallis test indicated that the effect of the ratings for the TAPS training was statistically significant ($H(4) = 121.204$, $p < .001$), indicating that was a statistically significant difference in level of effective training for the TKES on teaching practices between different ratings for the TAPS training. To determine where these differences had occurred, a series of post hoc Mann-Whitney U tests were conducted. Using the Bonferroni correction method, the significance level was set at $\alpha = .05/10 = .005$.

The results of post hoc Mann-Whitney U tests revealed that respondents who rated "not at all" reported significantly higher scores compared to those who rated 'moderate' ($p < .001$), 'large' ($p < .001$), and 'very large' ($p < .001$). Similarly, respondents who rated t 'small' reported significantly higher scores compared to those who rated 'moderate' ($p < .001$), 'large' ($p < .001$), and 'very large' ($p < .001$). Moreover, respondents who rated 'moderate' reported significantly higher scores than those who rated 'large' ($p < .001$), and 'very large' ($p < .001$). In addition, respondents

who rated ‘large’ reported significantly higher scores than those who rated ‘very large’ ($p < .001$).

Research Question 7

RQ7: Is there a significant difference between level of effective training for the SGP on TKES teaching practices?

It was planned to conduct a one-way ANOVA to address this question. The dependent variable in this analysis was level of effective training for the TKES on teaching practices and the independent variable was the ratings for the SGP training (not at all, small, moderate, large, very large). Descriptive statistics were calculated on the independent variable and the overall sample.

Table 21 shows the descriptive statistics for level of effective training for the TKES on teaching practices by the ratings for the SGP training. The scores for those who reported 'not at all' ranged from 10 to 30 with a mean of 16.74 (SD = 7.335), for those who reported 'small' ranged from 10 to 42 with a mean of 22.45 (SD = 7.295), for those who reported 'moderate’ ranged from 10 to 50 with a mean of 28.52 (SD = 7.283), for those who reported 'large' ranged from 28 to 50 with a mean of 35.78 (SD = 6.154), and for those who reported 'very large' ranged from 24 to 50 with a mean of 41.96 (SD = 8.062).

Table 21

Descriptive Statistics for Level of Effective Training for the TKES on Teaching Practices by the Ratings for the SGP Training

	N	Mean	*SD*	Min	Max	Skewness	Kurtosis
Not at all	31	16.74	7.335	10	30	0.733	-0.892
Small	53	22.45	7.295	10	42	0.158	-0.211
Moderate	141	28.52	7.283	10	50	-0.198	1.014
Large	98	35.78	6.154	18	50	-0.536	0.767
Very Large	24	41.96	8.062	24	50	-0.731	-0.681
Total	347	29.52	9.672	10	50	-0.163	-0.324

It was planned to perform a one-way ANOVA to determine whether there was a significant difference in level of effective training for the TKES on teaching practices by the ratings for the SGP training. Statistical assumptions were checked before performing this analysis. The statistical assumptions including no missing data, no outliers, homogeneity, normality, independence of observations, and interval or ratio level of measurement were considered for this analysis. *Z*-scores were calculated for level of effective training for the TKES on teaching practices for each level of the ratings for the SGP training to investigate the presence of outliers in the data and no significant outliers were identified. The Shapiro-Wilk test was used to examine the normality assumption of the data. Based on the results of this test, no significant departures from normality were observed for those who responded 'small' ($SW(53) = .959$, $p = .066$). On the contrary, the results of these tests were significant for those who indicated 'not at all' ($SW(31) = .826$, $p < .001$), 'moderate' ($SW(141) = .957$, $p < .001$), 'large' ($SW(98) = .946$, $p = .001$), and 'very large' ($SW(24) = .871$, $p = .006$).

Due to the normality assumption being violated for four out of the five subgroups, the researcher examined the Box-Cox transformation method to see if the distribution of

the data can be made more similar to the normal distribution. However, it was found that no improvements could be achieved in the normality of the data using this transformation method. Hence, this model was evaluated using the raw data. A Kruskal-Wallis test was employed as the nonparametric alternative to the one-way ANOVA test to evaluate whether there was a significant difference in level of effective training for the TKES on teaching practices by the ratings for the SGP training.

The results of the Kruskal-Wallis test indicated that the effect of the ratings for the SGP training was statistically significant ($H(4) = 167.811$, $p < .001$), indicating that was a statistically significant difference in level of effective training for the TKES on teaching practices between different ratings for the SGP training. To determine where these differences had occurred, a series of post hoc Mann-Whitney U tests were conducted. Using the Bonferroni correction method, the significance level was set at $\alpha = .05/10 = .005$.

The results of post hoc Mann-Whitney U tests revealed that respondents who rated "not at all" reported significantly higher scores compared to those who rated 'small' ($p = .001$), 'moderate' ($p < .001$), 'large' ($p < .001$), and 'very large' ($p < .001$). Similarly, respondents who rated 'small' reported significantly higher scores compared to those who rated 'moderate' ($p < .001$), 'large' ($p < .001$), and 'very large' ($p < .001$). In addition, respondents who rated 'moderate' reported significantly higher scores than those who rated 'large' ($p < .001$), and 'very large' ($p < .001$). Moreover, respondents who rated 'large' reported significantly higher scores than those who rated 'very large' ($p = .001$).

Summary

This study is designed to identify whether teachers perceive the Georgia TKES to improve their instructional practice, including any differences in perceptions by teacher characteristics of experience, grade level taught, and level of education. For this purpose, seven research questions and their corresponding hypotheses were formulated. Multiple two-way ANOVAs were conducted to evaluate Research Questions 1-3. In addition, a series of Kruskal Wallis tests were performed to evaluate Research Questions 4-7. The dependent variables were perceived influence of the TKES on teaching practices, perceived accuracy of the TKES assessment, and perceived level of the TKES. The independent variables were gender, race/ethnicity, grade level taught, educational level, content area taught, number of years certified as a teacher, the ratings for the TKES training, the ratings for the TAPS training, and the ratings for the SGP training. It was found that the effect of gender on both perceived accuracy of the TKES assessment and perceived level of creativity allowed by the TKES assessment was significant. Males reported significantly higher scores on perceived accuracy of the TKES assessment on teaching practices and significantly lower scores on teachers' perceived level of creativity allowed by the TKES assessment compared to females. These results indicated that the effect of grade level on perceived influence of the TKES on teaching practices and perceived level of creativity allowed by the TKES assessment was significant. These results indicated that teachers who taught students in grades pre-kindergarten to kindergarten reported significantly higher mean scores on perceived influence of the TKES on teaching practices compared to the teachers who taught High School students in grades 9 – 12. In addition, teachers who taught students in other grades ascribed

significantly higher scores on perceived influence of the TKES on teaching practices compared to the teachers who taught middle school students in grades 6-8 and high school students in grades 9-12. It was also found that teachers who taught students in grades pre-kindergarten to kindergarten reported significantly lower mean scores on perceived level of creativity allowed by the TKES compared to the teachers who taught High School students in grades 9 – 12.

The results also indicated that the interaction between number of years certified as a teacher and educational level had a significant effect on perceived level of creativity allowed by the TKES assessment. The results from the Kruskal-Wallis tests revealed that there were significant differences by the ratings for the TKES, TAPS, and SGP training on all three dependent variables. A series of post hoc Mann-Whitney U tests were employed to determine where these differences had occurred. It was found that among the teachers who had a master's degree, those who had 11-15 years of experience reported significantly lower scores on perceived level of creativity allowed by the TKES assessment than those with 1-5 and 16-20 years of teaching experience. Moreover, among the teachers who had a doctorate degree, those who had 11-15 years of experience as a certified teacher reported significantly higher scores than those with 6-12 and 21+ years of teaching experience. Furthermore, among teachers with a bachelor's degree, those who had 6-10 years of experience reported significantly higher scores than those with 1-5 and 16-20 years of teaching experience.

Moreover, it was found that the effects of the ratings for the TKES, TAPS, and SGP training were statistically significant on level of effective training for the TKES on teaching practices. It was found that, regarding the ratings for the TKES training, those

who rated ‘small’ reported significantly higher scores compared to those who rated ‘moderate’, ‘large’, and ‘very large’. Respondents who rated ‘moderate’ reported significantly higher scores compared to those who rated ‘large’, and ‘very large’. In addition, respondents who rated ‘large’ reported significantly higher scores than those who rated ‘very large’. Moreover, regarding the ratings for the TAPS training, those who rated “not at all” reported significantly higher scores compared to those who rated ‘moderate’, ‘large’, and ‘very large’. Similarly, respondents who rated t ‘small’ reported significantly higher scores compared to those who rated ‘moderate’ , ‘large’ , and ‘very large’. Moreover, respondents who rated ‘moderate’ reported significantly higher scores than those who rated ‘large’, and ‘very large’. In addition, respondents who rated ‘large’ reported significantly higher scores than those who rated ‘very large’. Lastly, regarding the ratings for the SGP training, those who rated “not at all” reported significantly higher scores compared to those who rated ‘small’, ‘moderate’, ‘large’, and ‘very large’. Similarly, respondents who rated ‘small’ reported significantly higher scores compared to those who rated ‘moderate’, ‘large’, and ‘very large’. In addition, respondents who rated ‘moderate’ reported significantly higher scores than those who rated ‘large’ , and ‘very large’ . Moreover, respondents who rated ‘large’ reported significantly higher scores than those who rated ‘very large’.

Chapter V

DISCUSSION

The important role of evaluations as a strategy for assessing teacher quality has led to a growing consensus on the need for evaluation systems that can yield higher quality information that can be used to improve teacher performance (Taylor & Tyler, 2012). The research problem addressed in the proposed study was based on the fact that while stakeholders in the educational system have worked diligently to reform teacher evaluation to impact teacher performance, it is not known how teachers perceive the way specific evaluation instruments improve instructional practice in order to increase student learning outcomes. Considering the resources invested in the evaluation process both in terms of finance and time, it was found important to examine the perceptions and views of teachers involved in the evaluation process to help improve the educational reform movement (Shakman et al., 2012). Consequently, the purpose of this quantitative nonexperimental study was to identify teacher perceptions of the Georgia TKES evaluation; in particular, whether teachers perceived the Georgia TKES to improve their instructional practice, including any differences in perceptions by teacher characteristics of experience, grade level taught, and level of education.

Data were collected using through the Teacher Survey of the Georgia TKES, and the survey used was adapted from a survey developed in previous research by Battle-Edwards (2017) with a focus on collecting data on teachers' perceptions of the TKES. The dependent variables in this study were (1) teachers' perceived influence of the TKES on teaching practices, (2) teachers' perceived accuracy of the TKES assessment, and (3) the perceived level of creativity allowed by the TKES. The independent variables were a)

gender, (b) race or ethnicity, (c) grade level taught, (d) level of education, (e) years of education as a certified teacher, (f) content area taught, (g) level of effective training for the TKES, (h) level of effective training for the TAPS, and (i) level of effective training for the SGP. A total of 347 participants took part in the study, of whom 280 were female, 305 were reported as White or Caucasian. Of the total, 94 participants formed the largest group of high school teachers in terms of grade level taught, followed by middle school teachers at 75. With regards other characteristics, a majority of the participants held post-graduate degree and taught general subjects in elementary grades.

The findings of the study suggested that there was a significant impact of gender on both perceived accuracy and perceived level of creativity, while perceived level of creativity was also significantly affected by the interaction between number of years certified as a teacher and educational level. Further, findings showed significant differences based on the ratings for the TKES, TAPS, and SGP training on all three dependent variables. The present chapter contains a summary of the results, interpretations in relation to existing literature, specification of limitations of the study, and discussion of recommendations and implications.

Literature Review

The review of the literature focused on teacher evaluations and how teachers perceive these systems. Conducting the literature review was organized by five different areas. These areas included: the history of teacher evaluations, effective teaching, reforming teacher evaluation, perceptions of teacher evaluation systems, and the Georgia Department of Education Teacher Keys Effectiveness system.

Jewel (2015) noted that prior to 1965, the responsibility for assessing teachers' performance rested primarily with community leaders who oversaw local schools. From 1930 onwards, the evaluation of teachers was influenced by the Hawthorne studies, which suggested that observing and providing feedback on employees' performance improves workplace efficiency (Taylor & Tyler, 2012). The Hawthorne-based model for evaluating teachers involved conducting evaluations one to three times a year, based mainly on a checklist of observable classroom behaviors. The results of these evaluations were recorded in a file in the human resources department and were used for documenting years of teaching experience, as well as for hiring and firing teachers. However, this evaluation system did not place emphasis on how the results could be used to enhance teachers' effectiveness (Taylor & Tyler, 2012).

According to Jewel (2015), the Hawthorne model was utilized until the 1960s, and evaluations of teachers persisted to identify the most effective ways to enhance instruction. Jewel (2015) also noted that starting in the 1970s, standardized test scores were used to assess instructional quality. Donaldson and Papay (2014) stated that teacher evaluations were integrated as part of educational reform during this period, holding teachers accountable for student performance. This approach was called clinical supervision, where the school principal took charge of conducting pre-observations, observations, and post-observations of teachers in the classroom. Furthermore, principals met with teachers to discuss their professional practice, with the goal of enhancing overall teacher quality (Donaldson & Papay, 2014). Compared to previous models, the principal's role was more prominent in clinical supervision.

As of late 2015, 42 states had been granted waivers from NCLB requirements; however, many districts still employed value-added assessment models as part of their teacher evaluations (Akers, 2016). These 43 states made the switch to new student testing systems that aligned with college and career readiness standards, in conjunction with the implementation of teacher effectiveness policy (Doherty & Jacobs, 2015). The states that have adopted the new evaluation systems are utilizing a value-added model that estimates the impact of individual teachers on their students' learning outcomes (Koedel, Mihaly, & Rockoff, 2015).

According to Doherty and Jacobs (2015), Georgia's teacher evaluation system is comparable to those implemented in 43 other states across the US. Despite much educational research and widespread confusion among researchers, the public, and politicians regarding the advantages and disadvantages of the value-added model and its impact on student achievement, there is a lack of qualitative research on the pedagogical influences of the value-added model of evaluation from the perspective of teachers who have undergone three years of evaluation implementation (Bogart, 2013; Taylor & Tyler, 2012).

Reforming the teaching profession now centers on teacher effectiveness, with policymakers implementing initiatives that ensure students receive high-quality instruction leading to improved learning outcomes (Welsh, 2011). Recognizing the significance of an effective teacher evaluation system, school district leaders and stakeholders have prioritized its design to produce reliable results (Darling-Hammond, 2014). The validity of an evaluation instrument is measured by its ability to accurately

reflect a teacher's practices that contribute to student learning outcomes (Goe et al., 2008).

According to recent research by Chetty et al. (2014) and Jackson (2012), teacher effectiveness has a lasting impact on student learning outcomes. Despite the overwhelming evidence in the literature regarding the crucial role of teacher effectiveness, reformed teacher evaluations have failed to accurately distinguish between effective and ineffective teachers (Marzano, 2012). The evaluation of effective teaching is challenging, as identifying the characteristics of an effective teacher is difficult (Rockoff, 2004), and there is no agreement among educators on how to measure it. Stigler and Hiebert (1999) noted that there is no consensus on whether to evaluate teacher effectiveness based on teaching outcomes, teacher perceptions, teaching methods, or a combination of these factors. Furthermore, they emphasized that teaching itself, rather than teachers, is a crucial factor in student academic achievement. This suggests that instructional systems used in teaching have a more significant impact on student learning outcomes than teaching strategies, professional development, assessments, curriculum, and other factors related to the teaching profession. Given the complexity of education and the diverse working conditions of educators, measuring effectiveness remains a challenging task.

The implementation of teacher and principal evaluation systems across the United States has been revised to include student performance data, in response to states receiving federal grants from Race to the Top. The need for these educational reforms was initially highlighted by the National Commission on Excellence in Education in 1983, with the publication of A Nation at Risk, which identified shortcomings in the

American public education system. NCLB (2002) also put pressure on states to revise their evaluations. The concern driving these reforms is that a decline in educational quality may have negative implications for economic growth in America, as argued by Hanushek (2009).

According to Taylor and Tyler (2012), the reformation of teacher evaluation systems is a common component of public education reform. These evaluation tools are typically designed to measure a teacher's professional growth and competency (Weems & Rogers, 2010). A comprehensive teacher evaluation system includes both formative and summative evaluations, which Shinkfield and Stufflebeam (2012) distinguish as follows: formative evaluations aim to provide feedback that could enhance teacher effectiveness, while summative evaluations assess a teacher's overall value to the school system. In the period from 2009 to 2012, nearly two-thirds of U.S. school districts revised their teacher evaluation systems to include specific performance standards for measuring teacher effectiveness and practices (Jerald, 2012).

Researchers and stakeholders seeking to understand how teacher evaluations can affect student learning outcomes need to gain a better understanding of teacher perceptions of the implementation of performance-based evaluation instruments, according to Darling-Hammond et al. (2012). Despite the importance of teacher perspectives, limited research has focused on their perceptions of new evaluation systems and how they measure teacher effectiveness. In a recent study by Finster and Milanowki (2018), the researchers examined teacher perceptions of the impact of a new performance evaluation system on their practice. They found that teachers need to see how the components of an evaluation system are linked to student learning outcomes. Another

study by Donahue and Vogel (2018) investigated teacher perceptions of an evaluation system on classroom practice in a single school district across subject areas. The researchers identified five themes from the data that highlighted components of the evaluation system that were beneficial to teachers, including feedback, quality of relationships, the evaluation rubric, modeling, personal integrity, and self-reflection.

According to Rogers (2019), teacher evaluation systems in Georgia are similar to those used across the United States, relying on student achievement data to determine teacher effectiveness. The Georgia Department of Education (2013) reports that the Georgia Teacher Keys Effectiveness System (TKES) was fully implemented as the official evaluation tool for Georgia public schools during the 2012-2013 academic year. The TKES aims to improve both student achievement and teacher effectiveness (Georgia Department of Education, 2013, 2018) and includes a rubric used for both summative and formative evaluations. The TKES measures teacher effectiveness based on three components, as outlined by the Georgia Department of Education (2018): Teacher Assessment on Performance Standards (50% of overall score), Professional Growth (20%), and Student Growth (30% of overall score). The Assessment on Performance Standards is a qualitative evaluation method that uses rubrics to measure teacher quality on ten performance standards. Professional Growth is assessed through teachers' achievement of Professional Growth Goals, while Student Growth is measured through Student Growth Percentile Measures based on state assessment data.

In the existing literature, researchers had examined the variable of gender in the context of evaluation. For instance, in a quantitative descriptive study, Torbert (2014) studied 126 teachers and administrators from an urban school district in Georgia

regarding the perceived effectiveness of the TKES. Regarding gender, Torbert (2014) did not find statistically significant difference in participants' perceptions of the TKES by years of teaching experience, gender, or school level for teachers and for administrators. In this regard, the findings of the current study, which showed that there was a significant effect of gender on both perceived accuracy and perceived level of creativity, did not confirm the findings of Torbert (2014). Further, the current study also showed that the number of years certified as a teacher and educational level did have a significant effect on perceived level of creativity. Teachers' grade level also affected perceived influence and perceived level of creativity. However, to some extent, the findings did align with Torbert (2014). For instance, the findings showed there was no significant difference in perceived influence by gender. Likewise, the findings showed there was no significant difference in perceived influence and perceived accuracy by number of years certified as a teacher and in perceived accuracy by grade level. These differences may be explained by the fact that the current study undertook a more granular inquiry, dividing perceptions of the TKES into multiple elements, such as perceived influence, perceived accuracy, and perceived level of creativity. Further, while Torbert (2014) focused on both teachers and administrators, this was not true of the current study, which was based on teachers' perceptions.

Another researcher, Griffith (2017), had also focused on teachers' perceptions of the TKES. In that study, the focus was on high school teachers' experiences with the Georgia TKES, where the findings showed that feedback from the TKES has the potential to support a positive change in the classroom, provided appropriate time and resources are dedicated to implementing the evaluative process with fidelity. In the

current study, the influence of grade level on perceived influence was found significant. Here, again, the granular focus of the current study considered multiple variables, thus providing a richer understanding compared to past research.

In a study, Redman (2018) focused on teachers' perceptions of a teacher evaluation system, Kentucky's Professional Growth and Effectiveness System (PGES), with attention to the impact of years of experience. In that study, it was found that there was no statistically significant difference between teachers' perceptions of the evaluation system and the years of experience. In this regard, the findings of Redman's (2018) study align with those of the current research. In the current study, the variable of years of experience was studied in relation to teachers' perceived influence of the TKES on teaching practices, perceived accuracy of the TKES assessment, and perceived level of creativity allowed by the TKES. Findings showed there was no significant difference in perceived influence by number of years certified as a teacher, no significant difference in perceived accuracy by number of years certified as a teacher and no significant difference in perceived level of creativity by experience. Thus, in alignment with the findings of Redman (2018), it appears that teachers' years of experience does not have an effect on their perceptions of the evaluation system even across different evaluation systems, such as the TKES studied in the current study and the PGES studied by Redman (2018).
The findings of the study expand and complicate the current understanding on the conceptual framework of this study, which was based on the view that teacher evaluation systems can help improve a teacher's professionalism through the provision of useful feedback and support that address specific teachers' needs at their current level of practice and relevant professional growth opportunities. The findings of this study

showed that, while there was a significant effect of the interaction between number of years certified as a teacher and educational level on perceived level of creativity with respect to TKES, overall it was found that there was no significant difference in perceived influence by number of years certified as a teacher, no significant difference in perceived accuracy by number of years certified as a teacher and no significant difference in perceived level of creativity by experience. Thus, the variable of experience, which was assumed in the conceptual framework as relevant in the way the evaluation systems can help teachers, may not be as effective in their perceptions of the evaluation system. However, in general, teachers' did find grade level did affect their perceived influence and perceived level of creativity with respect to the evaluation system. Consequently, the findings broadened the conceptual framework by highlighting the various ways in which teacher-related variable may affect their perception of the evaluation system, and thus the effect it has.

Methodology

This study utilized a survey design with group comparisons, which is a nonexperimental approach that does not involve the manipulation of an independent variable or random assignment of participants to conditions (Babones, 2016; Della Porta & Keating, 2008). Instead, the study described the characteristics of a group of individuals through the use of self-developed or previously validated surveys (Johnson & Christensen, 2012; Kelley-Quon, 2018). Due to the impossibility of including the entire population of interest, a sample was recruited through stratified random sampling to ensure representativeness of each group. Participants were asked to complete web-based surveys, although those who prefer hard copy surveys were provided one with postage

paid by the researcher. To ensure the validity of the survey results, the data was subjected to statistical considerations and analyses. Skewness and kurtosis values were also calculated to assess the normality of the data.

The study investigated nine independent variables, which included gender (male or female), race or ethnicity (Asian, Black or African American, Latino/Hispanic, White, Native Hawaiian or other Pacific Islander, and American Indian or Alaska Native), grade level (ranging from prekindergarten to 12th grade), educational attainment level (bachelor's degree, master's degree, Educational Specialist's degree, and doctorate), years of teaching experience (1–2 years, 3–5 years, 6–10 years, and more than 10 years), subject taught (mathematics, science, language arts, and electives), and the extent of influence of training on TKES, TAPS, and SGP. Nominal scales were used to measure gender, race or ethnicity, grade level, educational attainment, and content area taught. Years of experience were measured on an interval scale, while training was measured on a nominal scale.

The study measured three dependent variables: influence on teaching practices, agreement with assessment score, and allowance of teacher creativity. These variables were assessed using a 10-item scale in each section, with a 5-point rating scale on an interval level. The total score for each variable ranged from 10 to 50, with each item rated on a scale of 1 to 5. Section B asked about the influence of the Georgia TKES on teaching practices, Section C asked about agreement with assessment scores, and Section D asked about the allowance of teacher creativity. The total scores for each section were used to answer research questions 1 to 4. Multiple two-way ANOVAs were conducted to

evaluate Research Questions 1 through 3. In addition, a series of one-way ANOVAs were performed to evaluate Research Questions 4 through 7.

Results

Research Question 1

With the first research question, the goal was to determine if there was a significant difference by gender and race or ethnicity on teachers' perceived influence of the TKES on teaching practices, perceived accuracy of the TKES assessment, and perceived level of creativity allowed by the TKES. The null hypothesis for the first element of this research question, which stated that there would be no significant difference by gender and race or ethnicity on teachers' perceived influence of the TKES on teaching practices, was not rejected by the findings, as the results showed that there was no significant interaction between gender and race ($F(1,343) = 0.204$, $p = .652$), and the main effect of race ($F(1,343) = 0.188$, $p = .665$) or gender ($F(1,343) = 1.437$, $p = .231$) on perceived influence of the TKES on teaching practices were not statistically significant. The null hypothesis for the second element of this research question, which stated that there would be no significant difference by gender and race or ethnicity on teachers' perceived accuracy of the TKES assessment, had mixed results. Findings showed that the effect of gender on perceived accuracy of the TKES assessment was found to be statistically significant, $F(1,343) = 3.928$, $p = .048$, $\eta 2 = .011$; however, there was no significant interaction between gender and race ($F(1,343) = 0.012$, $p = .913$), and the main effect of race on perceived accuracy of the TKES assessment was not statistically significant ($F(1,343) = 0.171$, $p = .679$).). The null hypothesis for the third element of this research question, which stated that there would be no significant

difference by gender and race or ethnicity on teachers' perceived level of creativity allowed by the TKES, also had mixed results. Findings showed that the effect of gender on perceived creativity was significant, $F(1,343) = 4.934$, $p < .05$, $\eta 2 = .014$, as male participants reported significantly lower scores on perceived creativity; however, there was no significant difference in perceived accuracy by ethnicity ($F(1,343) = 0.131$, $p = .718$) or the interaction between gender and ethnicity ($F(1,343) = 1.178$, $p = .279$).

Research Question 2

With the second research question, the goal was to determine if there was a significant difference by grade level taught and level of education on teachers' perceived influence of the TKES on teaching practices, perceived accuracy of the TKES assessment, and perceived level of creativity allowed by the TKES. The null hypothesis for the first element of this research question, which stated that there would be no significant difference by grade level taught and level of education on teachers' perceived influence of the TKES on teaching practices, had mixed results. Findings showed that the main effect of grade level was found to be statistically significant, $(5,324) = 2.496$, $p = .031$, $\eta 2 = .037$; however, no significant differences were identified by educational level ($F(3,324) = 0.143$, $p = .934$) or the interaction between grade level and educational level ($F(14,324) = .664$, $p = .809$). The null hypothesis for the second element of this research question, which stated that there would be no significant difference by grade level taught and level of education on teachers' perceived accuracy of the TKES assessment, was not rejected, as the findings showed there was no significant difference in perceived accuracy by educational level ($F(1,324) = 1.410$, $p = .240$), grade level ($F(1,324) = 2.247$, $p = .050$), or the interaction between these factors ($F(1,324) = .641$, $p = .830$). The null

hypothesis for the third element of this research question, which stated that there would be no significant difference by grade level taught and level of education on teachers' perceived level of creativity allowed by the TKES, had mixed results. Findings showed that the effect of grade level on perceived level of creativity was significant, $F(5,324) = 2.548$, $p < .05$, $\eta 2 = .038$; however, there was no significant difference by educational level ($F(1,324) = .992$, $p = .397$) and the interaction between grade level and educational level for perceived level of creativity ($F(1,324) = .895$, $p = .565$).

Research Question 3

With the third research question, the goal was to determine if there was a significant difference by years of experience as a certified teacher in Georgia and educational level on teachers' perceived influence of the TKES on teaching practices, perceived accuracy of the TKES assessment, and perceived level of creativity allowed by the TKES. The null hypothesis for the first element of this research question, which stated that there would be no significant difference by years of experience as a certified teacher in Georgia and educational level on teachers' perceived influence of the TKES on teaching practices, was not rejected, as the findings showed there was no significant difference in perceived influence by number of years certified as a teacher ($F(1,328) = 0.711$, $p = .585$), educational level ($F(1,328) = 0.610$, $p = .609$), or the interaction between these categorical variables ($F(1,328) = 1.444$, $p = .152$). The null hypothesis for the second element of this research question, which stated that there would be no significant difference by years of experience as a certified teacher in Georgia and educational level on teachers' perceived accuracy of the TKES assessment, was not rejected, as the findings showed there was no significant difference in perceived accuracy

by number of years certified as a teacher ($F(1,328) = 0.878$, $p = .477$), educational level ($F(1,328) = 2.273$, $p = .080$), or the interaction between these variables ($F(1,328) = 1.526$, $p = .120$). The null hypothesis for the third element of this research question, which stated that there would be no significant difference by years of experience as a certified teacher in Georgia and educational level on teachers' perceived level of creativity allowed by the TKES, had mixed results. Findings showed there was no significant difference in perceived level of creativity by experience ($F(1,328) = 1.334$, $p = .257$) or educational level ($F(1,328) = 1.470$, $p = .223$); however, the interaction between experience and educational level had a significant effect on perceived level of creativity ($F(1,328) = 2.515$, $p = .005$).

Research Question 4

With the fourth research question, the goal was to determine if there was a significant difference by content area taught on teachers' perceived influence of the TKES on teaching practices, perceived accuracy of the TKES assessment, and perceived level of creativity allowed by the TKES. The null hypothesis for the first element of this research question, which stated that there would be no significant difference by content area taught on teachers' perceived influence of the TKES on teaching practices, was not rejected, as the findings showed there was no significant difference by content area for square root transformation of perceived influence, $H(7) = 12.552$, $p = .084$. The null hypothesis for the second element of this research question, which stated that there would be no significant difference by content area taught on teachers' perceived accuracy of the TKES assessment, was not rejected, as the findings showed there was no significant difference by content area for natural log transformation of perceived accuracy, ($H(7) =$

5.855, p = .557). The null hypothesis for the third element of this research question, which stated that there would be no significant difference by content area taught on teachers' perceived level of creativity allowed by the TKES, was also not rejected, as the findings showed there was no significant difference by content area for square root transformation of perceived level of creativity, (H(7) = 6.121, p = .526).

Research Question 5

With the fifth research question, the goal was to determine if there was a significant difference between levels taught of the effectiveness of training on the TKES on teaching practices. The null hypothesis for this research question, that there would be no significant difference between levels taught of the effectiveness of training on the TKES on teaching practices, was not supported, as the findings showed that the effect of the ratings for the TKES training was statistically significant (H(3) = 121.435, p < .001), indicating there was a statistically significant difference in level of effective training for the TKES on teaching practices between different ratings for the TKES training.

Research Question 6

With the sixth research question, the goal was to determine if there was a significant difference between levels taught of the effectiveness of training on the TAPS on teaching practices. The null hypothesis for this research question, that there would be no significant difference between levels taught of the effectiveness of training on the TAPS on teaching practices, was not supported, as the findings showed the effect of the ratings for the TAPS training was statistically significant (H(4) = 121.204, p < .001), indicating that was a statistically significant difference in level of effective training for the TKES on teaching practices between different ratings for the TAPS training.

Research Question 7

With the seventh research question, the goal was to determine if there was a significant difference between levels taught of the effectiveness of training on the SGP on teaching practices. The null hypothesis for this research question, that there would be no significant difference between levels taught of the effectiveness of training on the SGP on teaching practices, was not supported, as the findings showed the effect of the ratings for the SGP training was statistically significant ($H(4) = 167.811$, $p < .001$), indicating there was a statistically significant difference in level of effective training for the TKES on teaching practices between different ratings for the SGP training.

Limitations and Assumptions

Prior to conducting the study, a number of limitations were identified, and their influence was notable throughout the process of conducting the study. However, the impact these limitations had on the findings of the study as well as their applicability was limited through measures undertaken for validity and reliability. Thus, potential limitations with the survey instrument were reduced by adapting a survey that had been previously utilized. In this study, I used a survey developed by Battle-Edwards (2017), with survey items altered based on my research questions and variables examined in the study. In an effort to provide content validity, a formal expert panel reviewed the survey, providing clarity, technical quality of items, and content validity (Brockmeier et al., 2009, 2014; McNeill & Brockmeier, 2005). The expert panel included a university professor in education who helped design the Georgia TKES.

Further, with regards reliability, I tested for the internal consistency of the self-developed survey using Cronbach's alpha. Cronbach's alpha testing procedures were

used to conduct a reliability analysis for Perceived Influence, Perceived Accuracy, Perceived Level of Creativity, Total Scale (i.e., a summed total for perceived influence, perceived accuracy, and perceived level of creativity), and Total Scale plus TKES, TAPS, and SGP (33 items total). The results revealed that each section of the survey (α= .97, .98, and .98), the whole survey (α = .97), and the survey plus TKES, TAPS, and SGP (α = .97) were reliable as the reliability coefficients for all these scales were well above the acceptable level of .70.

A limitation that was identified prior to conducting the study was the fact that the research was based on self-report. It was expected that the anonymity of responses would encourage teachers to respond honestly. In this regard, due to the nature of the research, which focused on teacher perceptions, the use of self-report was found to be appropriate. However, certain limitations inherent in the use of self-report, including the possibility of obfuscation or dishonesty, cannot be ruled out. While the anonymity and privacy provided to the participants was aimed at encouraging honesty, and the use of online survey provided participants security and comfort in providing thoughtful responses, the possibility of obfuscation or deception cannot be avoided, and this remained a limitation throughout the study.

Another limitation mentioned prior to conducting the study was the possibility of low rate of participation. However, this limitation was overcome, as the research received participation from a total of 347 respondents. However, there was limited variation in demographics, with 280 out of 347 being female, while the rest male, 305 out of 347 being White, while the rest non-White. On the contrary, there was greater variation with respect to grade level taught, where the largest group of participants were high school

teachers (94 out of 347, 27.1%), followed by middle school teachers (75 out of 347, 21.6%). There was also diversity with regards education, with 38.9% (135 out of 347) reported to have a master's degree, and 32% (111 out of 347) reported to have an Educational Specialist degree. Likewise, with regard to content area taught, respondents included those who taught general subjects in elementary grades (67 out of 347, 19.3%), followed by respondents who taught Mathematics (51 out 347, 14.7%). Finally, there was variation in the number of years participants had been certified as a teacher in Georgia, where the largest group of respondents had been certified for more than 21 years (114 out of 347, 32.9%).

Another limitation identified prior to conducting the study was the fact that, due to the non-experimental nature of the research study, no attempt was made for determining association and causation. Since the focus of this study was on perceptions of the teachers, this limitation remained as part of the nature of the study and the methodology chosen. Finally, it was noted that the research would compare survey results among a group of teachers and thus will be limited by population and sample size. While the focus remained on a single evaluation system and only teachers participated in this study, it can be noted that some of the findings were in alignment with research conducted in different state with different evaluation system, such as Kentucky's PGES studied by Redman (2018). However, the nature of the research was fixed, and thus the generalizability of the study remain limited by the methodological choices made. Before utilizing one-way ANOVA and two-way factorial ANOVAs as parametric tests, statistical considerations and assumptions had to be met for this study. There are four primary assumptions that must be satisfied for parametric tests, which include: (a)

normality, (b) homogeneity of variance, (c) independence, and (d) the dependent variable being in interval/ratio form (Sedgwick, 2015). To evaluate the normality assumption, a Shapiro-Wilks test was conducted on the dependent variables (Siddiqi, 2014). Skewness and kurtosis values were also calculated for each variable to evaluate the normality of the data and measure the shape of the distribution of each variable.

Suggestions for Future Research

Several recommendations for future research can be derived from this study. Of the recommendations derive, some are concerning future research, relevant to other researchers with interest in evaluation systems, teacher performance, and associated variables that were studied in the current research, such as the role of teachers' gender and years of experience on perceptions concerning evaluation systems.

With regards recommendations concerning future research, it can be noted that the self-reported nature of the data collected may have led to bias, as participants may not have felt comfortable sharing some details. Further, the online survey design of this study may have prevented possibilities that could be exploited during face-to-face interviews. Thus, it is recommended that future researchers consider conducting similar research in which qualitative method is used with face-to-face, in-depth interviews for data collection. It may be that lack of face-to-face contact while sharing personal experiences may be a barrier.

Additionally, future researchers may consider using multiple sources of data, including reports and focus groups, in order to ensure there is triangulation of data sources. It is also recommended that future researchers consider attaining greater diversity with respect to gender and race, since the participants in the current study were

primarily female and White. Researchers using quantitative method may consider using an experimental research design to determine the factors responsible for certain perceptions' occurring with certain variables. Finally, future researchers may also consider expanding the population and evaluation systems used to include principles as well as other states' evaluation systems in order to determine the extent to which the findings obtained in this study are replicated.

Conclusion

Despite the fact that stakeholders in the educational system have worked diligently to reform teacher evaluation to impact teacher performance, it was not known how teachers perceive the way specific evaluation instruments improve instructional practice in order to increase student learning outcomes. The purpose of this quantitative nonexperimental study was to identify teacher perceptions of the Georgia TKES evaluation; in particular, whether teachers perceived the Georgia TKES to improve their instructional practice, including any differences in perceptions by teacher characteristics of experience, grade level taught, and level of education.

The dependent variables in this study were teachers' perceptions of the state-level evaluation instrument impacting teaching practices, agreement with assessment score, and allowed teacher creativity. The independent variables were gender, race/ethnicity, grade level taught, educational level, content area taught, number of years certified as a teacher, the ratings for the TKES training, the ratings for the TAPS training, and the ratings for the SGP training. Data were collected through the Teacher Survey of the Georgia TKES, and the survey used was adapted from a survey developed in previous research by Battle-Edwards (2017) with a focus on collecting data on teachers'

perceptions of the TKES. A total of 347 participants took part in the study, of whom 280 were female, 305 were reported as White or Caucasian.

The findings of the study suggested that there was a significant impact of gender on both perceived accuracy and perceived level of creativity, while perceived level of creativity was also significantly affected by the interaction between number of years certified as a teacher and educational level. Further, findings showed significant differences based on the ratings for the TKES, TAPS, and SGP training on all three dependent variables. The findings aligned, to some extent, with previous researchers such as Torbert (2014), Griffith (2017), and Redman (2018). However, the finding that there was a significant effect of gender on both perceived accuracy and perceived level of creativity did not confirm the findings of Torbert (2014).

Some limitations of the research included validity and reliability, use of self-report, possible low rate of participation, non-experimental nature of the study, and limited generalizability. Of these, validity and reliability as well as possibility of low rate of participation were successfully overcome. Remaining limitations provided opportunities for future research, such as the use of multiple sources of data, experimental design, and more diverse participants. Implications of the study, for research, social change, and practice were also noted. Regarding social change, the findings obtained their relevance from the important role of teachers in influencing student outcomes and behavior. With regard practice, some pointers for improving existing evaluation system based on the findings of the study were discussed. As the findings of the study provided teacher perceptions of the Georgia TKES evaluation, showing there was a significant impact of gender on both perceived accuracy and

perceived level of creativity, while perceived level of creativity was also significantly affected by the interaction between number of years certified as a teacher and educational level, in addition to significant differences based on the ratings for the TKES, TAPS, and SGP training on all three dependent variables, this chapter concludes the present study.

REFERENCES

Adkins, R. (2017). *Teacher perceptions of an evaluation system in relation to teacher pedagogy and student achievement* (Order No. 10637969). (1990647886). https://search.proquest.com/docview/1990647886?accountid=6988

Aguilar, C. E., & Richerme, L. K. (2014). What is everyone saying about teacher evaluation? Framing the intended and inadvertent causes and consequences of Race to the Top. *Arts Education Policy Review, 115*(4), 110–120. https://doi.org/10.1080/10632913.2014.947908

Angelle, P. S., & DeHart, C. A. (2011). Teacher perceptions of teacher leadership: Examining differences by experience, degree, and position. *NASSP Bulletin, 95*(2), 141–160. https://doi.org/10.1177/0192636511415397

Akers, R. (2016). *The utility of a standards-based teacher evaluation as a measure of effectiveness* [Doctoral dissertation, Frostburg State University]. Maryland Shared Open Access Repository. https://mdsoar.org/bitstream/handle/11603/3825/Akers_Dissertation.pdf

Babones, S. (2016). Interpretive quantitative methods for the social sciences. *Sociology, 50*(3), 453–469. https://doi.org/10.1177/0038038515583637

Battle Edwards, A. B. (2017). *An examination of the Georgia Teacher Key Effectiveness System* [Doctoral dissertation, Clark Atlanta University]. Atlanta University Center. https://radar.auctr.edu

Bergsmann, E., Schultes, M. T., Winter, P., Schober, B., & Spiel, C. (2015). Evaluation of competence-based teaching in higher education: From theory to practice. *Evaluation and Program Planning*, *52*, 1–9. https://doi.org/10.1016/j.evalprogplan.2015.03.001

Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* ERIC. https://eric.ed.gov/?id=ED540960

Bogart, C. D. (2013). *Teacher evaluation and classroom practice: Teacher perceptions in Northeast Tennessee* (Paper 1177) [Doctoral dissertation, East Tennessee State University]. http://dc.etsu.edu/etd/1177

Bradley-Levine, J., Romano, G., & Reichart, M. (2017). Teacher leaders' influence on teachers' perceptions of the teacher evaluation process. *International Studies In Educational Administration, 45*(3), 65-84.

Brockmeier, L. L., Green, R. B., Archibald, J. G., Pate, J. L., & Leech, D. W. (2014). Validation of the Teacher's High Stakes Testing Survey. *International Journal of Humanities and Social Science, 4*(3), 31–40. http://www.ijhssnet.com/journals/Vol_4_No_3_February_2014/4.pdf

Brockmeier, L. L., Pate, J. L., & Leech, D. W. (2009). Validation of the Principal's High Stakes Testing Survey. *Georgia Educational Researcher, 7*(2), Article 4. https://doi.org/10.20429/ger.2009.070204

Callahan, K., & Sadeghi, L. (2015). Teacher perceptions of the value of teacher evaluations: New Jersey's ACHIEVE NJ. *International Journal of Educational Leadership Preparation*, *10*(1), 46–59.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*, 2593–2632. https://doi.org/10.1257/aer.104.9.2593

Cochran-Smith, M., Piazza, P., & Power, C. (2013). The politics of accountability: Assessing teacher education in the United States. *Educational Forum, 77*(1), 6–27. https://doi.org/10.1080/00131725.2013.739015

Coe, R., Alosi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research.* Sutton Trust. https://www.suttontrust.com/research-paper/great-teaching/

Conceição, S. O. (2016). Competing in the world's global education and technology arenas. *New Directions for Adult & Continuing Education*, *2016*(149), 53–61. https://doi.org/10.1002/ace.20176

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.

Croft, S. J., Roberts, M. A., & Stenhouse, V. L. (2015). The perfect storm of education reform: High-stakes testing and teacher evaluation. *Social Justice, 42*(1)*,* 70–92. http://www.socialjusticejournal.org

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching* (2nd ed.). Association for Supervision and Curriculum Development.

Danielson, C. (2016, April 18). Charlotte Danielson on rethinking teacher evaluation. *Education Week, 35*(28), 20–24. https://www.edweek.org/ew/articles/2016/04/20/charlotte-danielson-on-rethinking-teacher-evaluation.html

Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator*, *38*(1), 4–13.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8–15. https://doi.org/10.1177/003172171209300603

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*, 418–446. https://doi.org/10.1002/pam.20586

Della Porta, D., & Keating, M. J. (Eds.). (2008). *Approaches and methodologies in the social sciences: A pluralist perspective.* Cambridge University Press.

Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, postsecondary attainment and earnings. *Review of Economics and Statistics, 98*(5), 848–862. https://doi.org/10.1162/REST_a_00598

DeWitt, D. (2018). *Teachers perception of the evaluation process on classroom instruction* (Order No. 10816409). (2042371821). Retrieved from https://search.proquest.com/docview/2042371821?accountid=6988

Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading and learning.* National Council on Teacher Quality. https://www.nctq.org/dmsView/StateofStates2015

Domina, T. (2014). An education in politics: The origins and evolution of No Child Left Behind [Review of the book *An education in politics: The origins and evolution of No Child Left Behind,* by J. H. Rhodes]. *Contemporary Sociology, 43,* 253–255. https://doi.org/10.1177/00943061145224151l

Donahue, E., & Vogel, L. (2018). Teacher perceptions of the impact of an evaluation system on classroom instructional practices. *Journal of School Leadership, 28*(1), 31–55. https://doi.org/10.1177/105268461802800102

Donaldson, M. L., & Papay, J. P. (2014). Teacher evaluation for accountability and development. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 190–209). Routledge.

Duran, M. (2018). *Circulating power through teacher evaluation system: A Foucauldian perspective* (Publication No. 10814671) [Doctoral dissertation, University of Texas at San Antonio]. ProQuest Dissertations and Theses Global.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2013). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org10.3758/BRM.41.4.1149

Finster, M., & Milanowski, A. (2018). Teacher perceptions of a new performance evaluation system and their influence on practice: A within- and between-school level analysis. *Education Policy Analysis Archives, 26*(41/42), 1–44. https://doi.org/10.14507/epaa.26.3500

Frasier, A. M. S. (2017). *Examining teacher perceptions of the relationship between evaluation policy and teacher practice in a North Carolina school system* [Doctoral dissertation, Michigan State University]. Michigan State University Digital Repository. https://d.lib.msu.edu/etd/6886

Frey, B. B. (Ed.). (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation*. Sage. https://doi.org/10.4135/9781506326139.n671

Gagnon, D. J., Hall, E. L., & Marion, S. (2017). Teacher evaluation and local control in the US: An investigation into the degree of local control afforded to districts in defining evaluation procedures for teachers in non-tested subjects and grades. *Assessment in Education: Principals, Policy & Practice, 24,* 489–505. https://doi.org/10.1080/0969594X.2016.1167669

Gamson, D. A., McDermott, K. A., & Reed, D. S. (2015). The Elementary and Secondary Education Act at fifty: Aspirations, effects, and limitations. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 1*(3), 1–29. https://doi.org/10.7758/RSF.2015.1.3.01

Georgia Department of Education. (2013). *Teacher Keys Effectiveness System.* https://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/TKES%20Handbook%20FINAL%207-18-2013.pdf

Georgia Department of Education. (2018). *Georgia's Teacher Keys Effectiveness System: Implementation handbook.* https://www.gadoe.org/School-Improvement/Teacher-and-Leader-

Effectiveness/Documents/Finalized%20TKES%20Handbook%20with%20district%20feedback%20%202016-2017.pdf

Goe, L., Bell, C. A., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* National Comprehensive Center for Teacher Quality. https://gtlcenter.org/sites/default/files/docs/EvaluatingTeachEffectiveness.pdf

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher*, *44*(2), 87–95. https://doi.org/10.3102/0013189X15574905

Good, T. L., & Lavigne, A. L. (2017). *Looking in classrooms*. Routledge.

Griffith, D. L. (2017). *Noncore secondary teachers' lived experience of TKES evaluative feedback* (Publication No. 10746749) [Doctoral dissertation, Piedmont College]. ProQuest Dissertations and Theses Global.

Hanlin, D. C. (2014). *The relationship between research-based leadership practices and emotional intelligence of high school principals* [Doctoral dissertation, University of Maryland]. Digital Repository at the University of Maryland. https://drum.lib.umd.edu/handle/1903/15677

Hanushek, E. (2009). School policy: Implications of recent research for human capital investments in South Asia and other developing countries. *Education Economics*, *17*, 291–313. https://doi.org/10.1080/09645290903142585

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal, 51,* 73–112. https://doi.org/10.3102/0002831213517130

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence Based Nursing, 18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Huber, M., & Melly, B. (2015). A test of the conditional independence assumption in sample selection models. *Journal of Applied Econometrics, 30*(7), 1144–1168. https://doi.org/10.1002/jae.2431

Jackson, C. K. (2012). *Non-cognitive ability, test scores, and teacher quality: Evidence from ninth grade teachers in North Carolina* (NBER Working Paper No. 18624). National Bureau for Economic Research.

Jaffurs, A. C. (2017). *Teacher perceptions of teacher evaluation using the teacher performance assessment system and factors that contribute to teacher quality, professional growth, and instructional improvement over time* [Doctoral dissertation, University of Maryland]. Digital Repository at the University of Maryland. https://drum.lib.umd.edu/handle/1903/20342

Jerald, C. (2012). *Movin' it and improvin' it—Using both education strategies to increase teaching effectiveness.* Center for American Progress. http://www.americanprogress.org/wp-content/uploads/issues/2012/01/pdf/movin_it_improvin_it.pdf

Jewel, L. A. (2015). Oil and water: How legal education's doctrine and skills divide reproduce toxic hierarchies. *Colum. J. Gender & L.*, *31*, 111. https://heinonline.org/HOL/LandingPage?handle=hein.journals/coljgl31&div=12&id=&page=

Johnson, R. B., & Christensen, L. B. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). SAGE.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2012). *Have we identified effective teachers? Validating Measures of Effective Teaching using random assignment.* Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://k12education.gatesfoundation.org/resource/have-we-identified-effective-teachers-validating-measures-of-effective-teaching-using-random-assignment/

Kappler, L. A. (2017). *Teacher perceptions of the Mississippi Teacher Evaluation System: Insight for school administrators* (Publication No. 10263490) [Doctoral dissertation, Mississippi State University]. ProQuest Dissertations and Theses Global.

Kelley-Quon, L. I. (2018). Surveys: Merging qualitative and quantitative research methods. *Seminars in Pediatric Surgery, 27*(6), 361–366. https://doi.org/10.1053/j.sempedsurg.2018.10.007

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review, 47,* 180–195. https://doi.org/10.1016/j.econedurev.2015.01.006

Koran, J. (2016). Preliminary proactive sample size determination for confirmatory factor analysis models. *Measurement and Evaluation in Counseling and Development, 49*(4), 296–308. https://doi.org/10.1177/0748175616664012

Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher, 46*, 234–249. https://doi.org/10.3102/0013189X17718797

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: Management of missing values and outliers. *Korean Journal of Anesthesiology, 70*(4), 407–411. https://doi.org/10.4097/kjae.2017.70.4.407

Ladd, S. (2016). *An examination of teachers' perceptions of teacher evaluation in New Jersey school systems* (Publication No. 1776478591) [Doctoral dissertation, Saint Peters University]. ProQuest Dissertations and Theses Global.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods.* Sage.

Marzano, R. (2012). Two purposes of teacher evaluation. *Educational Leadership, 70*(3), 14–19. http://www.ascd.org/publications/educational-leadership/nov12/vol70/num03/The-Two-Purposes-of-Teacher-Evaluation.aspx

Mascha, E. J., & Vetter, T. R. (2017). The statistical checklist and statistical review: Two essential yet challenging deliverables. Anesthesia & Analgesia, 124(3), 719–721.

Master, B. (2013). *Using subjective teacher evaluations to examine principals' personnel management priorities.* Society for Research on Educational Effectiveness. ERIC. https://eric.ed.gov/?id=ED564090

Mathison, S., & Ross, E. W. (2013). Hegemony and "accountability" in schools and universities. *Workplace*, *9*, 88–102. https://doi.org/10.14288/workplace.v0i9.184050

McGuinn, P. (2014, April). Presidential policymaking: Race to the Top, executive power, and the Obama education agenda. *The Forum, 12*(1), 61–79. https://doi.org/10.1515/for-2014-0017

McGuinn, P. (2016). From No Child Left Behind to the Every Student Succeeds Act: Federalism and the education legacy of the Obama administration. *Publius: The Journal of Federalism*, *46*(3), 392–415. https://doi.org/10.1093/publius/pjw014

McNeill, M. H., & Brockmeier, L. L. (2005). Relationships between academic program variables and success on the Registered Health Information Administrator certification examination. *Perspectives in Health Information Management, 2*, Article 4. https://perspectives.ahima.org/relationships-between-academic-program-variables-and-success-on-the-registered-health-information-administrator-certification-examination/

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Government Printing Office.

Nickolas, S. (2020). *How stratified random sampling works.* https://www.investopedia.com/ask/answers/032615/what-are-some-examples-stratified-random-sampling.asp

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).

Norris, J., van der Mars, H., Kulinna, P., Amrein-Beardsley, A., Kwon, J., & Hodges, M. (2017). Physical education teacher perceptions of teacher evaluation. *Physical Educator, 74*(1), 41–62. https://doi.org/10.18666/TPE-2017-V74-I1-6882

Pogrow, S. (2018). *Authentic quantitative analysis for education leadership decision-making and EdD dissertations: A practical, intuitive, and intelligible approach* (2nd ed.). ICPEL.

Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis*, *36*, 399–416. https://doi.org/10.3102/0162373714531851

Ramnarain, U. (2011). Teachers' use of questioning in supporting learners doing science investigations. *South African Journal of Education*, *31*(1), 91–101. https://www.ajol.info/index.php/saje/article/view/63493

Ravitch, D. (2016). *The death and life of the great American school system: How testing and choice are undermining education*. Basic Books.

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal, 48,* 462–501. https://doi.org/10.3102/0002831210382891

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*, 247–252. https://www.jstor.org/stable/3592891

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? (NBER Working Paper No. 14485). *Education Finance and Policy, 6*(1), 43–74. https://doi.org/10.3386/w14485

Rogers, L. M. (2019). *The influence of teacher effectiveness on student achievement: A case study* [Doctoral dissertation, Liberty University]. https://digitalcommons.liberty.edu/doctoral/1964/

Saleh, A., & Bista, K. (2017). Examining factors impacting online survey response rates in educational research: Perceptions of graduate students. Online Submission, 13(2), 63–74.

Salkind, N. J. (Ed.). (2010). Encyclopedia of research design (Vol. 1). Sage.

Sedgwick, P. (2015). A comparison of parametric and non-parametric statistical tests. *BMJ, 350*, Article h2053. https://doi.org/10.1136/bmj.h2053

Shakman, K., Riordan, J., Sanchez, M. T., Cook, K. D., Fournier, R., & Brett, J. (2012). *An examination of performance-based teacher evaluation systems in five states.* (REL 2012-129). Regional Educational Laboratory Northeast and Islands.

Shaw, R. D. (2016). Arts teacher evaluation: How did we get here? *Arts Education Policy Review, 117*(1), 1–12. https://doi.org/10.1080/10632913.2014.992083

Shinkfield, A. J., & Stufflebeam, D. L. (2012). *Teacher evaluation: Guide to effective practice*. Kluwer.

Siddiqi, A. F. (2014). An observatory note on tests for normality assumptions. *Journal of Modelling in Management, 19*(3), 290–305. https://doi.org/10.1108/JM2-04-2014-0032

Spring, J. (2014). *Globalization of education: An introduction*. Routledge.

Stecher, B. M., & Garet, M. (2014). *Introduction to the evaluation of the Intensive Partnerships for Effective Teaching (IP)* (Working Paper WR-1034-BMGF). RAND Corporation.

Stiggins, R. J., & Duke, D. L. (1988). *The case for commitment to teacher growth: Research for teacher evaluation.* State University of New York Press.

Stigler, J., & Hiebert, J. (1999). *The teaching gap.* The Free Press.

Stronge, J. H. (Ed.). (2006). *Evaluating teaching: A guide to current thinking and best practice* (2nd ed.). Corwin Press.

Stronge, J. H. (2018). *Qualities of effective teachers*. Association for Supervision and Curriculum Development.

Strunk, K. O., Weinsten, T. L., & Makkonen, R. (2014). Sorting out the signal: Do multiple measures of teachers' effectiveness provide consistent information to teachers and principals? *Education Policy Analysis Archives, 22*. https://doi.org/10.14507/epaa.v22.1590

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? Evidence of systematic growth in mid-career teachers. *Education Next, 12*(4), 78–84. http://www.educationnext.org/can-teacher-evaluation-improve-teaching/

Tofade, T., Elsner, J., & Haines, S. T. (2013). Best practice strategies for effective use of questions as a teaching tool. *American Journal of Pharmaceutical Education, 77*(7), Article 155. https://doi.org/10.5688/ajpe777155

Torbert, G. J. (2014). *Exploring the perceptions of teachers and principals regarding the Teacher Keys Effectiveness System in Georgia* (Publication No. 1642406799) [Doctoral dissertation]. ProQuest Dissertations and Theses Global.

U.S. Department of Education. (2009, July 24). *President Obama, U.S. Secretary of Education Duncan announce national competition to advance school reform.* [Press release] https://www.ed.gov/news/press-releases/president-obama-us-secretary-education-duncan-announce-national-competition-adva

Weems, D. M. & Rogers, C. B. H. (2010). Are US teachers making the grade? A proposed framework for teacher evaluation and professional growth. *Management in Education, 24*(1), 19–24. https://doi.org/10.1177/0892020609354959

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness* (2nd ed.). New Teacher Project. https://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf

Welsh, M. E. (2011). Measuring teacher effectiveness in gifted education: Some challenges and suggestions. *Journal of Advanced Academics*, *22*, 750–770. https://doi.org/10.1177/1932202X11424882

Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work?. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Routledge.

Wong, H., Wong, R., Rogers, K., & Brooks, A. (2012). Managing your classroom for success. *Science and Children, 49*(9), 60–64.

Appendix A:

Teacher Survey of the Georgia Teacher Keys Effectiveness System (TKES)

Purpose: The purpose of this survey is to identify teacher perceptions of the Georgia Teacher Keys Effectiveness System (TKES) evaluation. The study is designed to identify whether teachers perceive the Georgia TKES to improve their instructional practice and planning. The Georgia TKES was fully implemented during the 2012-2013 school year. As a relatively new evaluation instrument, research is needed on perceived effectiveness.

Consent: Submission of this survey indicates your consent for participation. All responses will be kept strictly confidential, and only group-level results will be reported.

Directions: Please rate the following statements as they apply to you. Select the number from the column that best reflects your opinion by darkening the number. Use the following scale:

1 = Not at all 2 = To a small extent 3 = To some extent 4 = To a large extent 5 = To a very large extent

A.	**The following questions are related to the extent to which the training you received for the TKES components influenced your teaching practice.** **Training**	Not at all	To a small extent	To some extent	To a large extent	To a very large extent
1.	Teacher Keys Effectiveness System (TKES) Rubric	(1)	(2)	(3)	(4)	(5)
2.	Teacher Assessment Performance Standards (TAPS)	(1)	(2)	(3)	(4)	(5)
3.	Student Growth Performance (SGP)	(1)	(2)	(3)	(4)	(5)

Directions: Please rate the following statements as they apply to you. Select the number from the column that best reflects your opinion by darkening the number. Use the following scale:

1 = Not at all 2 = To a small extent 3 = To some extent 4 = To a large extent

5 = To a very large extent

B.	**The following questions are related to the extent you use the TKES Rubric to plan for your teaching practices.** **Influence on Teaching Practices**	Not at all	To a small extent	To some extent	To a large extent	To a very large extent
1.	Professional Knowledge	(1)	(2)	(3)	(4)	(5)
2.	Instructional Planning	(1)	(2)	(3)	(4)	(5)
3.	Instructional Strategies	(1)	(2)	(3)	(4)	(5)
4.	Differentiated Strategies	(1)	(2)	(3)	(4)	(5)
5.	Assessment Strategies	(1)	(2)	(3)	(4)	(5)
6.	Assessment Use	(1)	(2)	(3)	(4)	(5)
7.	Positive Learning Environment	(1)	(2)	(3)	(4)	(5)
8.	Academically Challenging Environment	(1)	(2)	(3)	(4)	(5)
9.	Professionalism	(1)	(2)	(3)	(4)	(5)
10.	Communication	(1)	(2)	(3)	(4)	(5)

Directions: Please rate the following statements as they apply to you. Darken the number from the column that best reflects your opinion. Use the following scale:

1 = Not at all 2 = To a small extent 3 = To some extent 4 = To a large extent

5 = To a very large extent

C.	**The following questions are based upon the extent to which you agree with your evaluator's accuracy of using TKES to assess what you do as a teacher.** **Accuracy**	Not at all	To a small extent	To some extent	To a large extent	To a very large extent
1.	Professional knowledge	(1)	(2)	(3)	(4)	(5)
2.	Instructional planning	(1)	(2)	(3)	(4)	(5)
3.	Instructional strategies	(1)	(2)	(3)	(4)	(5)
4.	Differentiated instruction	(1)	(2)	(3)	(4)	(5)
5.	Assessment strategies	(1)	(2)	(3)	(4)	(5)
6.	Assessment usage	(1)	(2)	(3)	(4)	(5)
7.	Positive learning environment	(1)	(2)	(3)	(4)	(5)
8.	Academically challenging environment	(1)	(2)	(3)	(4)	(5)
9.	Professionalism	(1)	(2)	(3)	(4)	(5)
10.	Communication	(1)	(2)	(3)	(4)	(5)

Directions: Please rate the following statements as they apply to you. Darken the number from the column that best reflects your opinion. Use the following scale:

1 = Not at all 2 = To a small extent 3 = To some extent 4 = To a large extent

5 = To a very large extent

D.	**The following questions are based upon your experience with the extent to which TKES allows you to use creativity in teaching practice.** **Creativity**	Not at all	To a small extent	To some extent	To a large extent	To a very large extent
1.	Professional knowledge	(1)	(2)	(3)	(4)	(5)
2.	Instructional planning	(1)	(2)	(3)	(4)	(5)
3.	Instructional strategies	(1)	(2)	(3)	(4)	(5)
4.	Differentiated instruction	(1)	(2)	(3)	(4)	(5)
5.	Assessment strategies	(1)	(2)	(3)	(4)	(5)
6.	Assessment usage	(1)	(2)	(3)	(4)	(5)
7.	Positive learning environment	(1)	(2)	(3)	(4)	(5)
8.	Academically challenging environment	(1)	(2)	(3)	(4)	(5)
9.	Professionalism	(1)	(2)	(3)	(4)	(5)
10.	Communication	(1)	(2)	(3)	(4)	(5)

Directions: Please check the option that best describes you.

E.	**Demographic Information**
1.	**How long have you been a certified teacher in Georgia?** ( ) Less than 1 year ( ) 1–5 years ( ) 6–10 years ( ) 11–15 years ( ) 16–20 years ( ) 21+ years

2.	**What is the highest level of education you have completed?** ( ) Bachelor's degree ( ) Master's degree ( ) Educational Specialist degree ( ) Doctorate
3.	**Are you of Hispanic, Latino/a, or Spanish origin?** ( ) Yes ( ) No
4.	**What is your race?** ( ) American Indian or Alaska Native ( ) Asian ( ) Black or African American ( ) Native Hawaiian or other Pacific Islander ( ) White or Caucasian ( ) Other, please specify: _____
5.	**What is your gender?** ( ) Female ( ) Male
6.	**What grade level do you currently teach?** ( ) Elementary: pre-kindergarten to kindergarten ( ) Elementary: Grades 1–3 ( ) Elementary: Grades 4–5 ( ) Middle School: Grades 6–8 ( ) High School: Grades 9–12 ( ) If other, please specify: _____

7.	**What content area do you primarily teach?** ( ) Early childhood or pre-kindergarten ( ) Elementary grades, general ( ) Special education ( ) Fine Arts and Music ( ) English Language Arts ( ) English as a Second Language (ESOL) and Title III ( ) World Languages and Global Initiatives ( ) Health and Physical Education ( ) Mathematics ( ) Natural sciences ( ) Social Studies ( ) Vocational, career, or technical education ( ) Gifted ( ) Computer Science ( ) Literacy/Reading ( ) Science ( ) STEAM/STEM ( ) Other, please specify: _____

Appendix B:

Institutional Review Board Protocol Exemption Report

Institutional Review Board (IRB)
For the Protection of Human Research Participants

PROTOCOL EXEMPTION REPORT

Protocol Number: 04098-2020 **Responsible Researcher:** Tonya LeSure

Supervising Faculty: Dr. James L. Pate

Project Title: *Teacher Perception of the Georgia Teacher Evaluation Instrument and its Impact on Teacher Professional Growth.*

INSTITUTIONAL REVIEW BOARD DETERMINATION:

This research protocol is **Exempt** from Institutional Review Board (IRB) oversight under Exemption Category 2. Your research study may begin immediately. If the nature of the research project changes such that exemption criteria may no longer apply, please consult with the IRB Administrator (irb@valdosta.edu) before continuing your research.

ADDITIONAL COMMENTS:

- *Upon completion of this research study all collected data must be securely maintained (locked file cabinet, password protected computer, etc.) and accessible only by the researcher for a minimum of 3 years.*
- *Your research study may begin immediately at the following sites: Atkinson, Bacon, Brantley, Charlton, Clinch, Coffee, Pierce, and Ware County School Districts.*
- *The PDF (paper/pencil) version, if used, must include the Survey Research Statement. If a paper/pencil survey is used, participation is no longer be considered anonymous – but instead confidential.*
  - *The research statement must be modified to reflect that participation is confidential.*
- *All completed paper/pencil surveys must be kept in a separate file from name and email lists.*

☒ ***If this box is checked, please submit any documents you revise to the IRB Administrator at irb@valdosta.edu to ensure an updated record of your exemption.***

Elizabeth Ann Olphie *12.08.2020*
Elizabeth Ann Olphie, IRB Administrator

Thank you for submitting an IRB application.
Please direct questions to irb@valdosta.edu or 229-253-2947.

Revised: 06.02.16

Appendix C:

Letter to Superintendent requesting Permission to Conduct Research

December 02, 2020

Superintendent A
County School District
99 Ave.
Schooltown, GA 31642

RE: Permission to Conduct Research Study

Superintendent A:

I am writing to request permission to conduct a research study in your school district. I am currently enrolled in the Educational Leadership Doctoral program at Valdosta State University in Valdosta, GA. I am conducting a study to examine teacher perceptions of the Georgia Teacher Evaluation Instrument and Its Impact on Teacher Professional Growth on (a) teaching practices, (b) agreement with assessment score, and (c) allows for teacher creativity.

Elementary, middle, and high school certified teachers were randomly selected from a southeast RESA district. If approval is granted, teachers will complete a survey using a link through Qualtrics. Completion of the survey should require no more than 10 minutes, and no personal identifiable information will be in the database. Teacher responses will be kept confidential and only group summaries will be reported. Additionally, the study will adhere to Valdosta State University Institutional Review Board's policy for projects involving human subjects.

Your approval to conduct this study will be important to the success of this study. Your time and contribution to increasing the knowledge of teacher perceptions of the Georgia Teacher Keys Evaluation System and its impact on teacher professional growth is greatly appreciated. Please contact me at 912-381-7709 or email at talesure@valdosta.edu if you have questions about this study.

Sincerely,