

A COMPARATIVE ANALYSIS OF TEACHER EVALUATION POLICY

by

ERROLL QUINN HEADEN

BOB L. JOHNSON, JR., COMMITTEE CHAIR

DAVID DAGLEY

BRENDA MENDIOLA

MARGARET RICE

FOSTER WATKINS

A DISSERTATION

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

TUSCALOOSA, ALABAMA

2014

Copyright Erroll Q. Headen 2014
ALL RIGHTS RESERVED

ABSTRACT

Educational accountability has moved to the forefront of the educational reform movement. There is a call to have every student taught by a highly effective teacher at a highly effective school. How do we get there? As states compete for funding under the Race to the Top initiative (RTTT), new teacher evaluation policies have become one of the leading tools to assess teacher effectiveness.

The purpose of this study is to examine select state teacher evaluation policies and the literature on teacher evaluation for the larger purpose of informing teacher evaluation policies and practices in Alabama. The goal is to make Alabama education policymakers and leaders aware of the current trends in teacher evaluation and an analysis of their successes and failures. The study begins with educational accountability in the United States, starting with the signing of the Elementary and Secondary Education Act (ESEA) and follows the different acts, policies, initiatives, and competitions ending with RTTT. The second stage looks at the literature on educational reforms and the causal assumptions behind these reforms.

The research was based on a comparative policy analysis using qualitative methods. The study includes a comparative data matrix, which includes anchors and subanchors from five states: Georgia, Tennessee, New York, Illinois, and Alabama.

The move toward new teacher evaluation policies and models has been eminently contested and debated by all stakeholders. Several implications for states and teachers emerged from this study, including the following: (a) creating new teacher evaluation policies using multiple measures, (b) using student data as a percentage of a teacher's final evaluation rating,

and (c) the need for teachers to have a voice in the development of the new teacher evaluation policies.

DEDICATION

This dissertation is dedicated to my family. To my wife, Renee', and children, Quinn II and Ashley Rose, thank you for making this journey. To my mother, Emily H. Jenkins and my in-laws Nathaniel and Rosa Ashley, thank you for all the support you gave and keep giving me. I would like to give a special dedication to my late grandmother, Bernice Headen, who always believed in her youngest grandson. To my brother Scott (Stephanie), sister-in-law Tammie (Victor), and a host of uncles, aunts, nieces, nephew, and cousins, thank you for being a loving family.

ACKNOWLEDGMENTS

Praise God, from Whom all blessings flow; Praise Him, all creatures here below; Praise Him above, ye heavenly host; Praise Father, Son, and Holy Ghost. (Thomas Ken, 1674, Awake, my Soul, and with the Sun).

The undertaking of the dissertation process is not completed by one individual. It takes a multitude of support to travel from the beginning to the end of the process. I would like to acknowledge the many educators who were placed in my path from kindergarten through college for my educational base. In particular, I thank Drs. Bob L. Johnson Jr., David Dagley, Brenda Mendiola, Margaret Rice, and Foster Watkins for serving on my dissertation committee. Professors, your guidance, support, and patience were instrumental in the completion of this dissertation. To my wife, Dr. Renee' A. Headen, and my two children, Quinn II and Ashley Rose, your love and support is what made me continue and complete this dissertation. Thanks are also given to my mother, Emily H. Jenkins, who loves me unconditionally and to my father and mother-in-law, Nathaniel and Rosa Ashley, who love me as their son. The Headens, LTC. Scott, his wife Stephanie, Taylor, and Gabrielle thanks for always being there when I needed them. Thanks to my cousin Dr. Bertis D. English for always being encouraging and critical with every question I asked. A special thanks to my editor Ms. Sherri Edwards for her patience and dedication. There are many individuals that I would like to thank who have played a part in who I am today and who I will become in the future.

CONTENTS

ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGMENTS	v
LIST OF TABLES	10
I INTRODUCTION	1
Purpose of the Study	4
Questions Asked by the Study	5
The Rationale and Significance of the Study	6
Assumptions	7
Organization of the Study	7
II LITERATURE REVIEW	8
Background of Educational Reform	11
Elementary and Secondary Education Act 1965	11
A Nation at Risk	11
America 2000: An Education Strategy	12
Goals 2000: Educate America Act	13
No Child Left Behind	14
Race to the Top	14
School Accountability	16
Common Core Standards	18

High Stakes Testing	18
Merit Pay/Performance Pay	19
Charter Schools.....	21
State Government’s Role in Public Education.....	22
Highly Effective Teaching.....	23
History of Evaluation	25
Teacher Evaluation	26
Purpose of Teacher Evaluation	30
Use of Teacher Evaluation.....	31
Methods Used for Teacher Evaluations	31
Evaluation Data from Multiple Sources	35
Student Growth Models	36
Value-Added Models (VAM) of Teacher Evaluation	38
Problems with Teacher Evaluations.....	39
Teacher Evaluation for School Improvement.....	41
Licensure Test for Evaluation.....	42
Legal Issues.....	43
Politics and Teacher Evaluation.....	44
Summary.....	45
III METHODOLOGY	51
Rationale for Research Strategy.....	52
Research Design.....	53
Unit of Analysis	53

Study Data.....	53
Data Collection Methods	54
Sampling Logic and Strategy.....	55
Data Analysis Methods.....	57
Standards of Rigor in Study.....	58
Situating Self as Researcher.....	61
Limitations of Study	63
Summary.....	63
IV NARRATIVE AND SUMMARY OF A COMPARISON DATA MATRIX OF STATE TEACHER EVALUATION MODELS	64
Data Matrix Description Definition and Decision-Rule Logic.....	64
Decision-Rules-Logic	69
Georgia Teacher Evaluation Model: Teacher Keys Effectiveness System	71
Concerns of Teacher Keys Effectiveness System in Georgia.....	74
Georgia: Data Matrix Summary.....	76
Tennessee Teacher Evaluation Model: Tennessee Educator Acceleration Model	82
Concerns with TEAM.....	85
Tennessee: Data Matrix Summary.....	86
New York Teacher Evaluation Model: Annual Professional Performance Review.....	91
Concerns with New York Teacher Evaluation Model.....	92
New York: Data Matrix Summary.....	94
Illinois Teacher Evaluation Model/Chicago Teacher Evaluation Model: Performance Evaluation Reform Act.....	102
Chicago Teacher Evaluation Model.....	104

Concerns with Performance Evaluation Reform Act.....	104
Illinois/Chicago: Data Matrix Summary.....	107
Alabama Teacher Evaluation Model: PEPE to EducateAlabama.....	118
Description of Alabama’s Teacher Evaluation Policies	118
EducateAlabama	119
Alabama: Data Matrix Summary	120
Comparative Analysis of Teacher Evaluatoin Policy Across States	123
Summary	133
V SUMMARY, IMPLICATIONS, DISCUSSION, AND RECOMMENDATIONS	
FOR FUTURE RESEARCH	136
Questions Asked by the Study	137
Implications for States Looking to Design New Teacher Evaluation Policies.....	142
Implications for Teachers as New Teacher Evaluation Policy is Implemented	146
Discussions/Recommendations for Alabama	151
Recommendations for Future Research.....	155
REFERENCES	158

LIST OF TABLES

1	States Data Matrix.....	65
2	Completed State Data Matrix.....	124

CHAPTER I

INTRODUCTION

The need for teacher evaluation policies has increased as school reform and accountability look for ways to improve school effectiveness. These new policies need to be supportable and reliable to help distinguish teacher effectiveness (Glazerman, Leob, Goldhaber, Staiger, Raudenbush, & Whitehurst, 2010). Local, state and federal governments are spending in excess of \$400 billion per year on public education in the United States. As that amount continues to increase, stakeholders demand more accountability for student achievement outcomes (O'Day, 2002). Teachers have the most effect on student achievement outcomes and academic growth, so now the focus of 21st century education has targeted teacher quality as a policy issue (Manginate, 2010). As teacher quality becomes a focal point of policymakers, teacher evaluation has become the tool looked upon to achieve quality teaching in all schools.

According to Green, Baker, and Oluwole (2012), political think tanks are calling for more rigorous teacher evaluation models to judge the effectiveness of teachers. As states compete for additional funding, several have taken the recommendations of the political think tanks and are basing new evaluation models on student performance (Green et al., 2012).

This study focused on teacher evaluation and the role accountability plays in the public school reform movement. Teacher evaluations have been a part of the educational process since the beginning of American formal education (Ellett & Teddle, 2003). This research has led stakeholders to look for improved ways to advance student learning and America's global

assessments ranking. Accountability has become a major topic in the education discussion in America. However, accountability of Kindergarten through high school (K-12) public schools is not a new phenomenon. In an effort to improve K-12 public education, the federal government produced reforms, initiatives, and reform competitions; this study begins with *A Nation at Risk*, *Goals 2000*, *Goals 2000 Educate America Act*, *No Child Left Behind Act*, and *Race to the Top*. Through these reforms, policy initiatives, and educational reform competitions, teacher evaluation plays a major role in the accountability of public K-12 schools. New teacher evaluation systems designed and implemented by states also play a major role in the accountability of public K-12 schools. This study analyzes the newly designed systems of several states in an effort to provide professional insight on what is working and what is not as Alabama looks to implement its new system.

According to Dagley and Veir (2002a) and Kelly and Orris (2011), policymakers are now beginning to hold educators to a professional standard for student educational success or failure. This shift in responsibility by policymakers has increased the pressure on school leaders for school reform and student achievement. This idea has been the cornerstone of the development of new teacher evaluation policies.

As states compete for funds under the *Race to the Top*, which has set aside \$297 million to improve teacher effectiveness, many states will look to find ways to evaluate teacher performance (“Race to the Top,” 2011). Teachers are evaluated on the basis of what has become known as “best practices.” Best practices are methods that gained momentum in the 1990s, which used research and scientific studies to determine the best course of action for educational practices. The use of these practices should support teachers in their daily quest to improve student learning and hence improve student achievement (Iwanicki, 1990). Though best practices

exist, individual states are developing and implementing their own teacher evaluation systems. However, there is not a common method or policy for teacher evaluations.

While test scores are looking less acceptable and the cost of funding education increases, stakeholders and policymakers are beginning to focus on teachers for more accountability; therefore, teacher evaluations are becoming the tool of choice (Johnson, 1997). The methods used for evaluating teachers should also contain provisions that will support their professional development. Data gathered from evaluations, likewise, should help determine promotions, incentives, and merit pay for teachers (Darling-Hammond, 2004; Stark & Lowther, 1984). It is important to note a teacher evaluation system must address the concerns of all members of a society before it can be considered viable (Johnson, 1997).

The objective in defining and finding the right formula to evaluate teachers should start with what makes a teacher effective. Teacher effectiveness research has shown there is a link between the characteristic of good teaching and student learning and achievement (Ellett & Teddle, 2003). The focus on effective teaching led Secretary of Education Arne Duncan and education leaders to introduce *Shared Vision*, a comprehensive plan to improve teacher quality. *Shared Vision* has three main goals and seven core principles. The goals include preparing students to become college and career ready as well as good citizens, closing the achievement and opportunity gap, and preparing students to be globally competitive (USDOE, 2012). Ellett and Teddle (2003) concluded that new teacher evaluation systems should use both teacher effectiveness research and school effectiveness research. The concept of influencing student outcomes with teacher behavior and by using research-based data to develop new teacher evaluation systems affords practitioners a level of confidence in the system.

Purpose of the Study

The purpose of this study was to review and examine select state teacher evaluation policies and the literature on teacher evaluation for the larger purpose of informing teacher evaluation policies and practices in Alabama. The goal is to make Alabama education policymakers and leaders aware of the current trends in teacher evaluation and provide an analysis of their successes and failures. This study took a qualitative look at teacher evaluation. Research was gained to add to the body of knowledge of teacher evaluation policy as states, and in particular Alabama, begin to develop and pilot new teacher evaluation models. Educate Alabama is Alabama's teacher evaluation model; however, it does not have a component to evaluate teachers based on student data. Strong and Tucker (2003) explained the nature of teacher evaluation as being a process that has been around as long as teaching itself and its purpose is for accountability or improvement, and in some cases, both. As schools have moved into a time of high stakes accountability, the evaluation process has become a valuable part of the process.

This study explored several state models of teacher evaluation in order to give a research based assessment for further development of a new model for Alabama. As new evaluation policies are being designed this study looked at the purpose of evaluations. Marzano (2012) gave two purposes for teacher evaluation: to foster teacher learning and/or to measure teacher competence. He discusses the failure of past evaluation models as they have not done a good job distinguishing between effective and ineffective teachers nor have they helped develop a highly effective teacher workforce (Marzano, 2012). Marzano (2012) goes on to explain that teacher evaluations used for measuring the effectiveness of teachers and one that focuses on the development of teachers will look very different.

In an effort to understand the effect these new teacher evaluation policies and models will have on stakeholders, this study explored the research regarding teacher evaluation and what these new policy focuses will have on student learning, practitioner attitudes towards students and colleagues, the curriculum, and the profession of teaching.

In essence, the need to provide every student with a high quality education is paramount. A highly effective and highly qualified teacher is the only way school reform can be successful, and the only way to ensure our teachers are highly effective is to have a highly effective evaluation system in place. Not only must the evaluation system be highly effective, but its implementation has to be completed successfully (Strong & Tucker, 2003). Also, according to Zepeda and Kruskamp (2012) ensuring all students are provided with an excellent teacher is of the utmost urgency to all stakeholders as school move into the accountability age of common core standards.

Questions Asked by the Study

1. What does the literature have to say about teacher evaluation policy and how evaluations are being used as an accountability tool for states and school district?
2. What is the purpose and nature of teacher evaluation policy in the state of Alabama?
3. What is the purpose and nature of teacher evaluation policy enacted by a sample of 4 state designs other than Alabama?
4. What can the state of Alabama and its school districts learn from: a) the literature on teacher evaluation policy and b) a review of teacher evaluation policy in a select group of states that were examined in this study? What information do these data sources provide that could improve teacher evaluation policy in Alabama?

The Rationale and Significance of the Study

The significance of this study is to research teacher evaluation policies and models in an effort to understand and provide research based assessments of these new policies and models. Research has shown that teachers are the most important factor that affects student achievement in schools (Mangiante, 2010). However, the way in which teachers are evaluated in schools is broken and in need of reform (Papay, 2012). According to Goldrick (2002) evaluations are a means to determine a teacher's skill knowledge. However, Goldrick (2002) concludes evaluation methods could be subjective and not used as a tool for improving teaching. Evaluators must approach evaluations as a method to ensure quality and effective teaching. As Darling-Hammond (2004) explains, highly qualified teachers are needed to ensure students are given the same opportunities to succeed on the same set of learning standards. The children that attend America's public schools depend heavily on the millions of public school teachers that are charged with managing and dispensing knowledge in classrooms daily (Gordon, Kane, & Staiger, 2006). The quest to have every student taught by a highly effective teacher has shifted the reform focus to teacher evaluation and how to determine what is effective. Once it is determined what is effective, states will have to look for guidance from the United States Department of Education. According to Gutierrez (2012) Secretary of Education Arne Duncan was quoted as saying, "evaluation should never be based only on test scores. That would be ridiculous. It should also include factors like principal observation or peer review, student work, parent feedback. It should be designed locally and teachers should be at the table to help design it."

Assumptions

The following assumptions were present in this study:

1. The policies analyzed and compared from the Georgia Department of Education, Tennessee Department of Education, New York Department of Education, Illinois Department of Education, and the Alabama Department of Education websites were accurate.
2. The Race to the Top data analyzed from the United States Department of Education was accurate.

Organization of the Study

The final project is composed of five chapters. The first introduces the study. The second includes a review of the literature for school reform and the most current literature on teacher evaluation, consisting of types of evaluation, methods, issues, and possibilities. Chapter 3 is composed of the qualitative methodology for the study, Chapter 4 gives a narrative and summary of the Comparison Data Matrix of State Teacher Evaluation Models, and Chapter 5 is the summary and findings, implications, conclusion, and recommendations for future research.

CHAPTER II

LITERATURE REVIEW

The literature review of this study brings together relevant works to provide a conceptual and contextual framework for this examination of teacher evaluation policy. According to Fraenkel and Wallen (2003) the literature review should serve two purposes, one is to allow the researcher to locate and identify works of others who have researched the topic of interest and the other is for the researcher to be able to evaluate the existing research on the topic. Both primary and secondary sources will be used to complete the literature review.

Teacher evaluation policy is a tool used to judge or critique teacher practices in an effort to determine what is effective teaching. As America mulls over its education standing in the global society, a tool such as teacher evaluation has to be sharpened to insure that its purpose is sound and its reliability is unquestioned.

Imbedded in the literature review is a collage of information concerning educational reform from 1965 through 2009 and several tools implemented in an effort to improve education in America. The timeframe investigated is not a complete analysis of educational reform, but rather a good foundation for the role teacher evaluation has played and will play in the twenty-first century.

The idea of educational reform can be traced to the beginning of formal education. Teacher evaluation plays a significant role in the present day educational reform movement. However, before exploring teacher evaluation policy the literature on educational reform has to be investigated. Problems within the educational community have to be explored and an attempt

must be made to correct these problems. This literature review will problematize educational reform and examine the attempts to improve the educational system in America.

The literature review begins with a look at educational reform from the implementation of the Elementary and Secondary Education Act 1965 (ESEA), to the investigation of the National Commission on Excellence that released *A Nation at Risk: The Imperative for Educational Reform*. The study continues with explanations of George H. W. Bush's educational proposal and his attempt to reauthorize ESEA as *America 2000*. The Bush's administration's plan did lay the groundwork for the Clinton administration's educational plan entitled *Goals 2000*. The literature review goes on to examine the *No Child Left Behind Act* that reauthorized ESEA in 2002 and then to the educational competition of the Obama administration, *Race To The Top*.

Upon studying the educational reform literature, there have been numerous tools/programs to help with the school accountability/reform movement. Five approaches to school reform will be examined in Chapter two. These approaches are: the common core standards, high-stakes testing, merit pay, charter schools, and teacher evaluation policy.

In an effort to understand a small sample of ways educational organization have tried to improve schools, a cursory review of the literature on school accountability was conducted.

First, the common core state standards, a set of guidelines to ensure all stakeholders know what is expected for each student to learn in order to compete with other students throughout the world ("Common core state, 2010"). Second, high-stakes testing was explored as a tool for students to be evaluated based on a test usually given to a specific group of student to determine proficiency or there rank in that specific group. Third, merit pay was highlighted as a means to reward teachers who were deemed effective by a specific standard. Merit pay can be used in a

variety of ways to include extra pay for student performance, teaching in a high poverty school, or teaching an undermanned subject. Fourth, Charter schools as Peters (1994) proclaimed them “dynamic new public schools” legislation was first introduced in Minnesota. The literature review concentrated on the concepts, management, and spread of charter schools in America. Finally, the literature review focused on teacher evaluation policy. The review clarified the history of teacher evaluation to show its function in the educational process. Teacher evaluation is defined and the purpose of evaluation is explored. Formative and summative types of evaluations are given a close look and past and current methods of teacher evaluations are illustrated. The evaluation process using multiple data sources to include: classroom observations, artifacts of teachers work, surveys, and tools used to measure student progress help guide the literature review to its current place in American schools. The literature on student growth models and value-added models shifted the illustration in the review to focus on how technology is beginning to play a major role in teacher evaluation policy.

The literature review moves from the types and methods of teacher evaluation to common problems with teacher evaluation policy, to how teacher evaluation is being used to improve schools and how teacher licensure is being used to evaluate teachers. The literature review concludes with topics of legal issues and the politics of teacher evaluation policy.

As Johnson (1997) explained, stakeholders are seeing an increase in education expenditures and low returns on educational outputs. New reforms, programs, initiatives and competitions are being counted on to improve outcomes and provide mechanisms for stakeholders to determine the effectiveness of schools.

Background of Educational Reform

Elementary and Secondary Education Act 1965

May 1964, President Lyndon B. Johnson outlined a new vision and principles for the United States of America in a commencement speech at the University of Michigan. The *Great Society* was his goals for equality of opportunity, enhancement of urban life, restoration of natural beauty, improvement of education, ending poverty, and implementing racial justice (Boyer, 2001). President Johnson's focus was on civil rights and poverty. After work to get the Civil Rights Act of 1964 passed, the U.S. Office of Education issued a major report titled *Compensatory Education for Cultural Deprivation* (NYSED, 2009). The crown jewel of President Johnson's *Great Society* was his *War on Poverty*, in which the President wanted to get more federal aid for education to the states' poorest students. On April 11, 1965, President Johnson signed the Elementary and Secondary Schools Act into law. The Act had six original titles:

- Title I--Financial Assistance To Local Educational Agencies For The Education Of Children Of Low-Income Families
- Title II--School Library Resources, Textbooks, and other Instructional Materials
- Title III--Supplementary Educational Centers and Services
- Title IV--Educational Research And Training
- Title V--Grants To Strengthen State Departments Of Education
- Title VI--General Provisions

A Nation at Risk

In 1983, the Reagan Administration released *A Nation at Risk: The Imperative for Educational Reform*. A group was formed and headed by Education Secretary T. H. Bell consisting of education and business leaders. This group became known as the National

Commission on Excellence in Education. *A Nation at Risk* studied the educational landscape of America and compiled a list of findings and recommendations for the American public.

According to *A Nation at Risk*, six major findings were offered:

- The top students graduating from high school and college were not entering into the teaching profession
- More time should have been spent on teaching the content instead of method classes in the teacher education programs
- Salaries for teachers were too low
- Teacher shortages of certain areas
- Shortages in teachers for math and science were at an alarming rate
- Teachers that were being employed were not qualified to teach Math, Science and English

Based on the offerings regarding teaching in public education, The National Commission on Excellence in Education issued seven recommendations:

- Teachers should meet high academic standards and be competent in an academic area of concentration.
- Teachers' salaries should increase to be competitive in the American job market, with an effective teacher evaluation system to determine promotion, pay increases, retention and tenure.
- Teachers should work longer for professional growth and curriculum evaluation.
- Stakeholders should work together to develop steps for teachers, from beginners to master teachers.
- Allow qualified individuals to become certified in areas of math and science where there may be a shortage of teacher candidates.
- Incentive pay for teachers.
- Mentor programs by master teachers for novice teachers.

The findings and recommendations for teaching found in *A Nation at Risk* was to be a blueprint for schools as America looked to improve student achievement. The findings and recommendations were a new outlook on American education.

America 2000: An Education Strategy

President George H. W. Bush convened a National Education summit with the governors of each state. The term “systemic reform” became a key talking point in education circles.

Systemic reform referred to aligning curriculum, standards, assessments, teacher training, and resources (NYSED.gov 2012). *America 2000* was introduced to America following the National Education Summit; it was tasked with creating educational standards (Kasper, 2005). The Goals' standards were:

- All children in America will start school ready to learn.
- The high school graduation rate will increase to at least 90 percent.
- American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.
- U.S. students will be first in the world in science and mathematics achievement.
- Every adult American will be literate and possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.
- Every school in America will be free of drugs and violence and will offer a safe, disciplined environment conducive to learning.

America 2000 was a follow up to the findings in *A Nation at Risk*, which stated that there was “a rising tide of mediocrity” in America’s educational system. Goals 2000 passed the House of Representatives in the fall of 1992, but it was allowed to die in the United States Senate.

Goals 2000: Educate America Act

President Bush’s plan however laid the groundwork for the Clinton Administration’s educational reform plan, *Goals 2000: Educate America Act* which was the reauthorization of the Elementary and Secondary Education Act. Congress reauthorized the ESEA with a focus on school readiness, school completion, student academic achievement, leadership in math and science, adult literacy, and safe and drug-free schools, which were hold-over ideas from *America 2000* (Paris, 1994). However, *Goals 2000: Educate America Act* added two additional standards to help improve K-12 public education:

- The nation's teaching force will have access to programs for the continued improvement of their professional skills and the opportunity to acquire the knowledge and skills needed to instruct and prepare all American students for the next century.
- Every school will promote partnerships that will increase parental involvement and participation in promoting the social, emotional, and academic growth of children.

No Child Left Behind

Elementary and Secondary Education Act became known as the *No Child Left Behind Act* (NCLB Act) in 2002 under the Bush Administration. NCLB's major cornerstone was increased accountability for all students. Public schools that received federal funds in each state were required to make Adequate Yearly Progress (AYP) with all students. The year 2014 was the goal set for all students to reach proficiency regardless of race, socioeconomic status, language barriers, or special education determination. NCLB also would allow Parent Choice in their child's education for all students who are zoned to "failing schools" to be able to transfer to schools that had met the NCLB AYP requirement. President Bush wanted every child reading by the end of the third grade, so NCLB implemented the Reading First Initiative. NCLB is in the process of being reauthorized by congress and states have been given some flexibility for meeting the 2014 deadline for 100% proficiency.

Race to the Top

President Barack Obama and Education Secretary, Arne Duncan, set out to change the course of education reform with the Administration's ambitious plan called *A Race to the Top* fund. President Obama stated, "It's time to stop talking about education reform and start actually doing it. It's time to make education America's national mission." The federal government provided \$4.35 billion to fund this initiative, and it was to be divided among the states that met

certain criteria when they applied for funds (*Race to the Top*, 2009). The *Race to the Top Fund* has a set of six selection criteria:

- State Success Factors (125 points)
- Standards and Assessments (70 points)
- Data Systems to Supports Instruction (47 points)
- Great Teachers and Leaders (138 points)
- Turning Around the Lowest-Achieving Schools (50 points)
- General Selection Criteria (55 points)

Great Teachers and Leaders will be the focus of this study. There are five components that make up the selection criteria for Great Teachers and Leaders.

- Providing high-quality pathways for aspiring teachers and principals (21 points)
- Improving teacher and principal effectiveness based on performance (58 points)
- Ensuring equitable distribution of effective teachers and principals (25 points)
- Improving the effectiveness of teacher and principal preparation programs (14 points)
- Providing effective support to teachers and principals (20 points)

The Department of Education released 10 Flexibility Options for schools under the *Race* initiative: Flexibility regarding the 2013-2014 Timeline for determining AYP, Flexibility in Implementation of School Improvement Requirements, Flexibility in Implementation of LEA Improvement requirements, Flexibility for Rural LEAs, Flexibility for School wide programs, Flexibility to Support School Improvement, Flexibility for Reward Schools, Flexibility regarding Highly Qualified Teacher improvement plans, Flexibility of transfer certain funds and Flexibility to use School Improvement Grant Funds to support priority schools (USDOE, 2012).

According to the Flexibility Option, State Education Agency (AEA) and the Local Education Agency (LEA) are developing and adopting new evaluations for teachers and principals that will meet the following criteria: show continual improvement of instruction, meaningful differentiated performance using at least three performance levels, use multiple valid measures in determining performance levels, regular evaluations for principals and teachers, clear and timely feedback, and used in personnel decisions (USDOE, 2012). The move from

“Highly Qualified” teachers under NCLB is being replaced by “Highly Effective” teachers under the flexibility option.

The Flexibility Option also afford states an opportunity to gain waivers from parts of the NCLB requirements if the states implement new and innovative ways of improving student academics and school reforms (Klein, 2012). There have been 32 states and the District of Columbia to receive waivers under the Flexibility Option.

There are certain items specific to obtaining a waiver that focus on overall school improvement that follow four principles. Those principles are as follows Adopting college- and career-ready standards; creating state-defined accountability systems that reward success and promote improvement; supporting effective instruction and leadership; and reducing duplication and administrative burden placed on districts and schools (Ayers, Owens, et al., 2012). The spotlight of these four principles will focus on the effective instruction and leadership as it relates to teacher evaluation.

Ayers, Owen et al. (2012) discussed the different of the waiver principles as they relate to NCLB. NCLB required teachers to be “highly qualified”, but did little to ensure that the “highly qualified” teachers were effective and those students who needed the more effective teachers were provided those teachers. The waiver process was asking schools to be more accountable to all stakeholders.

School Accountability

Educational policy has shifted from what goes into a school to the quality of the student that leaves a school. Since the publication of *A Nation At Risk* in 1983, schools have faced increased scrutiny over student outcomes (Lee & Wong, 2004). Public schools and its certified

employees are entering into an era of increased accountability for more effective schools.

According to Darling-Hammond (2004) accountability systems were first designed on a standards-based system with testing as the key assessment tool which featured rewards and sanctions.

Newmann, King, and Rigdon (1997) surmise that educational accountability systems should have four parts:

- Information about the school's performance
- Standards for judging the quality or degree of success of organizational performance
- Significant consequences to the organization (good or bad) for its success or failure in meeting specific standards
- An agent or constituency that receives information on organizational performance, judges the extent to which standards have been met, and distributes rewards and sanctions

Darling-Hammond (2004) explained that several concepts of educational accountability exist. They include, but are not limited to, political accountability, legal accountability, bureaucratic accountability, professional accountability, and market accountability. The National Governors Association has produced guidelines for teacher effectiveness and a majority of states have begun to accept parts of these recommendations (Hazi & Arredondo-Rucinski, 2009b). As states look for new ways to finance public education, legislatures are asking questions about the outcomes of public school students and looking to evaluate the effectiveness of certified employees that are tasked to teach students. The end goal is to produce students that are "College and Career Ready" (USDOE, 2012b).

Cumulatively, these Federal reform movements have focused on different fixes for the education problem in the United States. Adopting common core standards, increased testing, establishing charter schools, merit pay, and improved teacher evaluation systems are some of the

most common ideas to improve public education. Teachers have the greatest impact on student achievement than any other variable associated with school organizations (White, 2004).

Common Core Standards

The Common Core Standards movement has gained momentum, with 44 states adopting the standards. According to the National Governors Association, the standards will give American children the opportunity to be “College and Career Ready” (2011). The Common Core Standards Initiative states these standards are going to give every student, teacher, and parent a concrete blueprint of what all students should be taught and what each student should learn in order to become college and career ready (Initiative, 2011).

As state after state begin to implement common core standards, ways of assessing the students’ knowledge of the standards have become more prominent. High stakes testing is used to determine if elementary students are meeting benchmarks, middle school students are progressing as they should, and high school graduation tests have become the norm. William (2010) stated, there is a place for high stake tests in today’s public education organization, but work needs to be done to ensure the tests are giving the accurate results.

High Stakes Testing

High stakes testing gets a different perspective among teachers. According to Anderson (2009), teachers feel testing is harmful if teachers are trying to cram all the standards in before the test and students are not given the time needed to digest the material. However, school administrators view testing as a positive influence on student achievement (Jones et al., 2006). Testing for accountability will be a part of the American public education system for the

foreseeable future (Jacob, 2005), and with the passage of NCLB test-based accountability has become the norm for schools (“New assessments for,” 2011).

Test makers are trying to create an assessment that will provide the best feedback for accountability. However, in doing so, assessments are created with duplicate questions on assessments from year to year, hence allowing teachers to focus on certain questions that are going to appear on the assessments from year to year. This type of assessment for accountability has drawn criticism from many academics, because the fear is students are not being developed socially, emotionally, and academically (“New assessments for,” 2011).

As accountability becomes the norm for public education, states have begun to search for innovative ways to attract and retain “highly effective” teachers and explore new ways of disseminating education to get students college and career ready. Merit pay and charter schools are two hot button topics that states are researching in an effort to achieve the accountability standards they are looking to reach.

Merit Pay/Performance Pay

Teacher quality has become one of the most valued commodities in today’s education landscape. Teachers are the most vital contributors when it comes to student efficiency. It begs to reason that the traditional single salary schedule will be debated as being obsolete in favor of merit based pay or performance pay (Adams, Heywood, & Rothstein, 2009).

Merit/performance pay for teachers has to be attached to student performance and this performance has to be measured. The idea is to create a link between the accountability goals of a school, teacher performance, and student performance (Adams et al., 2009).

Merit pay for teachers is determined using several different methods, with some being based on student performance, extra pay for an understaffed discipline such as math and science, or pay for working in a high poverty area. According to McFadyen (2001), teachers have long viewed merit pay as a way to reward teachers in “better” schools with more advanced students if the pay is based on student achievement. Dee (2004) stated that with a good teacher evaluation system the more effective teachers can be rewarded with merit pay.

According to Gordon et al. (2006), merit pay should be given to high performing teachers that serve in the lowest performing schools. The authors contend that the traditional pay system of a uniform pay scale does not help in rewarding the most effective teachers. Current research suggests teachers should receive between 10 and 20% merit raise from their base salaries (Gordon et al., 2006)

The question needing answered is does merit/performance pay work. The University of Vanderbilt’s Peabody School of Education conducted a 3-year study on the effects of incentive pay on middle school math teachers in a metropolitan school system. The study tested the notion that incentive pay would help increase student performance on standardized tests. The study found that there was no increase in test scores based on the group that was eligible for the incentive pay.

As the merit/performance pay debate continues, Marietta City Schools, Georgia, and Fulton County Schools, Georgia, are planning to implement performance pay for teachers beginning during the 2015-2016 school year, using teacher evaluation policy (Field, 2013).

Charter Schools

According to Peters (1994), the early 1990s brought about a push for “dynamic new public schools,” or what became known as Charter Schools or the Accelerated Schools model. The first charter school legislation passed in Minnesota in 1991. These “dynamic new public schools” would fit the needs of the most underprivileged students and give the same opportunity as other students and they would focus on accelerating the learning, not focusing on remediation for America’s most at-risk students (Peters, 1994). Charter schools are the new and innovative way to get the most out of students in public education. The most common charter schools are those run by State Boards of Education, but they do not have all the limitations and policies akin to traditional K-12 schools. During the school year 2009-2010 there were 4,952 charter schools operating in the United States (SER, 2011). As more states adopt charter schools legislation, the number of charter schools will increase. Parents are choosing charter schools based on academic success and academic programs that traditional public schools are not providing (Vanderhoff, 2008).

President Obama and his Education Secretary are proponents of charter school legislation and it is a key component of their Race to the Top federal education competition. The more charter schools increase in the United States, the more the impact on traditional public schools will be debated. Bohte (2004) concluded that traditional public schools will see an increase in success because of the competition from charter schools. Not only will traditional public schools improve from competition, but some will improve as the student population changes, as some low performing students transfer to charter schools. The success of charter schools will be determined by the type of legislation governing the schools and the mission of the schools themselves (Lawton, 2009).

State Government's Role in Public Education

Education is not mentioned in the United States Constitution; hence as stated by the Tenth Amendment, "The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people." This gives states the power to regulate education as long as this power does not violate any provision of the United States Constitution (Dennis, 2000).

States have addressed education in their state constitutions. The state legislative branch exercises the power that is granted to it to provide for the education of its citizens. Some but not all of the duties that are as follows:

1. Providing funding for public education
2. Licensing or chartering private schools
3. Providing oversight and guidance to local school boards
4. Setting broad policies for school-level curricula, texts, standards, and assessments
5. Licensing teachers
6. Overseeing the provision of educational services, and other special needs populations
7. Electing or appointing some or all of the members of the governing state boards of education

The role of state government in education can be mapped out in several stages. The first stage is known as the permissive era (1642-1821), in which states began enacting education laws to require citizens to be able to read and understand religious teachings. In the encouraging era (1826-1851), local governments began to encourage the creation of school districts and looked for ways to raise revenue to support these districts; however, states did not require the establishment of schools and students were not yet required to attend school. The next stage is identified as the compulsory era (1855-1980). This stage of government requires the establishment of school districts, the creation of school curriculums, and the enactment of compulsory education for all students. Finally, during the freedom or schools choice era (1980-present), the options for students began to expand. States began to accept homeschooling as a

form of appropriate education, vouchers became a new option for moving students from failing school, tuition tax credits were given, and charter schools became sanctioned by several states.

The role of states in public education has evolved over time and is continuing to have an impact on teachers, students, and communities.

Highly Effective Teaching

According to Tucker and Stronge (2005), the quality of a child's teacher, whether good or bad, has an educational effect on the child for years. The goal should now be to assure every student has a quality or effective teacher. Teachers still have the greatest influence on student learning in schools. The influence can be positive or negative, large or small, so policymakers and school leaders have to provide key tools for improving teacher performance (Zepeda & Kruskamp, 2012).

The No Child Left Behind Act required every teacher to be highly qualified, but this does not mean highly effective. NCLB and its high quality teacher qualification was a good starting point in the effort to ensure all teachers are teaching a high level (Tucker & Stronge, 2005). Researchers have agreed overwhelmingly that effective teaching correlates to student success (Stronge, Ward, Tucker, & Hindman, 2008)

Tucker and Stronge (2005) view teacher evaluation as a process to improve performance and for documenting accountability. Performance improvement is seen as formative in nature, while the accountability of evaluation is summative. As Tucker and Stronge (2005) pointed out, classroom observations have been the primary tool used for evaluating teachers. Observations however are lacking the needed components to be an effective stand-alone tool for an evaluation system.

As student progress has become one of the focal points of teacher evaluation systems, teachers are concerned with how student progress will be measured, by whom it will be measured, and with what resources. Accountability is a collective responsibility that should be shared by all stakeholders and not left to the teacher and student. This list of stakeholders includes parents and all school personnel from the teacher to the superintendent and board members, as they all have a function in student learning. Although other stakeholders play a role, it is ultimately the teacher and student who are interacting daily to achieve goals and standards that have been put in place (Tucker & Stronge, 2005).

According to Gordon et al. (2006), when teachers are classified as being highly effective, the conversation turns to compensation. Compensation for excellent teaching may lead to a better distribution of teachers across districts. The distribution of effective teachers to the neediest students should be a priority in American schools; however, this is not the case. Adjusting salaries for highly effective teachers that teach in low performing schools would be a great starting point in the school accountability movement. The standard uniform salary scales used by most school districts do not award highly effective teachers or penalize incompetent teachers (Gordon et al., 2006). Gordon et al. (2006) also suggested that effective teachers be granted a 10% to 20% pay increase when teaching in high poverty schools. For salary increases, a school with 75% free and reduced lunch students is determined to be a high poverty school.

The effort to ensure quality teaching has to focus on improving the teacher, and with an improved teacher comes school improvement (Strong & Tucker, 2003). These ideals have to work in concert with one another to get the accountability needed to maximize student achievement (Strong & Tucker, 2003). Peterson (1987) stated that the key to understanding effective teaching is through teacher evaluations. A key component to understanding what

teachers need professionally to improve instruction as well as evaluations can provide evidence for rewarding effective performance in the classroom.

History of Evaluation

According to Darling-Hammond (1990), teacher evaluation has not always been a high-stakes activity for schools. These early evaluations were no more than routine rituals performed by administrators with few resources and little organization. Darling-Hammond (1990) predicted that teacher evaluation would eventually have a greater role in education decision making.

Strong and Tucker (2003) later noted that evaluations have been a part of education from the beginning, usually for improvement or accountability and in some cases both. The process has now become more high stakes as schools have moved to the era of high standards with district and school accountability.

Teacher appraisal systems and other educational accountability measures can be traced back to the turn of the 20th century as teacher salaries began to increase and the profession gained status. This led to an increased interest in the profession by its stakeholders. Teacher effectiveness has been a concern of policymakers and other stakeholders since the 1900s with the implementation of evaluation tools, and student observations of teachers being one of the first methods used to gage the value of teachers (Good & Mulryan, 2006). The Committee on Child Development of the National Research Council was one of the first to study teacher effectiveness and created the first observational development research tools. The Ohio Teaching Record was created to encourage cooperation between teachers and supervisors to gain the most effective practice for teaching (Good & Mulryan, 2006).

Teacher Evaluation

As policymakers search for new and innovative ways to improve teaching, they have taken a turn from the traditional ways that were led by trying to raise hurdles to get into the teaching profession. A system of measurement needs to be developed that will measure a teacher's effectiveness based on what is being done on the job (Gordon et al., 2006). Teacher evaluation is a vital component to the overall environment of a school organization. As more conditions change in schools such as student populations, proration, and the conditions of the school and its students it becomes imperative that schools develop consistent policies to evaluate teachers (Sullivan & Zirkel, 1998). Strong and Tucker (2003) explained the value in designing a good teacher evaluation system that has the key components for success. Communication, commitment, and collaboration are the backbone to any good teacher evaluation system.

According to Strong and Tucker (2003), communication has to begin with the design of the evaluation system by allowing teachers to be a part of the building process, distributing handbooks to all evaluators and evaluatees, and having a system with continued training after the system has been put into practice. Also, Strong and Tucker (2003) explained the need for ongoing communication between the evaluator and teacher. Ongoing communication is paramount not only to the building of trust but to preventing any problems that may arise.

Commitment to excellence is a phrase made famous by former Oakland Raiders owner Al Davis. This concept should be the building block of any teacher evaluation system. Strong and Tucker (2003) illustrated the need for commitment in order to design a teacher evaluation system that is going to be a quality system. Commitment by policymakers, state- and district-level leaders, administrators, and teachers is essential if the process is to improve teacher quality. There has to be an organizational commitment to improving teaching that will foster a culture of

teaching excellence. Administrators and teachers must be given the time to learn the system and for implementation of the process. Finally, school systems must allocate time and money for professional growth activities (Strong & Tucker, 2003).

Ding (2009) understands, too, that a key component in defining and improving accountability will be teacher evaluation policy. Goldrick (2002) defined teacher evaluation as professionally assessing a teacher's work performance. Formal teacher evaluation has been defined using a moralistic and ethical approach, followed by the influence of personal traits of good teaching; classroom research that linked teacher behavior and student outcomes; teaching practices and student outcomes; linking teacher license to evaluation; and using teacher evaluation for accountability, professional development, and school improvement (Ellett & Teddle, 2003). In the past, teacher evaluations have focused on the curriculum, day-to-day management of schools, and new programs, but the new focus of teacher evaluations is to improve teacher quality (Millman, Darling-Hammond et al., 1990). The goal of the teacher evaluation policy is to help teachers grow professionally and to improve accountability. Although these two goals are most often separate, a good teacher evaluation system must combine these goals to maximize the system and the teachers who are being evaluated (Strong & Tucker, 2003). However, Millman, Darling-Hammond et al. (1990) goes further to define the purpose for evaluating teachers, to include the following areas: certification, formative and summative assessment, professional development, incentive pay and promotion, and retention and/or dismissal.

Teacher evaluation should try and use several approaches to determine if a teacher is promoting teaching and learning as well as to show the effectiveness of the teacher (Kyriakides & Demetriou, 2007). According to Kyriadies and Demetriou, teachers maintain the notion that

evaluation systems must be valid and reliable, and are meeting the educational function of improving teaching and learning. Teachers are comfortable with evaluations and prefer to be evaluated as opposed to not being evaluated. Teachers view evaluations by the building administrator as the most effective and rate peer evaluations as a good tool for schools also (Bouchamma, 2007).

According to Logue-Belden (2007), teachers have a favorable view of the evaluator, and that evaluators have good intent in administering the evaluation. However, perceptions of teacher evaluations do have a different take when using schools that have met Annual Yearly Progress (AYP) and those that did not meet AYP. However, according to Ovando and Ramirez (2007), administrators prefer a multi-year evaluation system and also one that allows experienced teachers more elasticity in their evaluation process.

Teachers are more accepting of evaluations if they have an input in the final product. Evaluations will be more effective if they are not used as punishment, but rather as a tool to help enhance the teaching process. Ovando and Ramirez (2007) believed that a teacher's belief in the process of teacher evaluations should fall under the umbrella of instructional leadership. Students also have an effect on teacher evaluations. Students who are highly motivated and are going to do well on tests tend to lead teachers in the direction of not changing for evaluations (Hall, 2008).

Chow, Wong, Yeung, and Mo (2002), found novice teachers perceived veteran teachers more competent to evaluate them than the building principal, while veteran teachers were comfortable with building principals as their evaluator. Both groups of teachers were concerned with the process of the evaluation being both formative and summative.

Since the beginning of the education reform movement in the 1980s, the dominant theme of nearly every professional journal and research report on school reform has been the topic of accountability in education. In the early years of the reform movement, the states wanted accountability-based teacher evaluations. Toward the end of the reform movement, however, many legislatures gave control over accountability and evaluations to the discretion of policymakers (Dagley & Veir, 2002a).

Teacher evaluation is one of the primary means of improving educational instruction, enhancing educational services, and justifying the removal of substandard teachers. The responsibility of implementing teacher evaluations is ultimately the responsibility of the administrator at the local school site. The accountability movement of the 1980s viewed teacher evaluation as one of the problems in the school system. During the 1980s, the inadequacy of teacher evaluations stifled efforts to improve teaching and learning in the school system (Dagley & Veir, 2002b). Before the reform movement gained momentum in the 1980s, relatively little legislation focused directly on teacher evaluation. From 1983 to 1985, there were only 20 states with legislation directly addressing teacher evaluation.

The inconsistent observations with little to no feedback serve no purpose in improving teaching and learning (Papay, 2012). As policymakers look to improve teacher performance, a serious conversation concerning the purpose of teacher evaluation has to be undertaken. Now, since the publication of *A Nation at Risk*, teaching has become one of the most scrutinized professions in the world. Researchers have studied ways to improve student and teacher performance; with the passage of NCLB, more data have become available to weigh student achievement to teacher performance (Staiger & Rockoff, 2010).

Purpose of Teacher Evaluation

As the American educational system continues to take on more criticism about performance on tests that compare American students to international students, the focus has shifted to teacher effectiveness. This issue of teacher effectiveness leads to the need for teacher evaluation models. Teacher evaluations are performed daily on an informal basis by students, parents, other teachers, and stakeholders (Natriello, 2006).

Natriello (2006) discussed three purposes of formal teacher evaluation processes. First, the evaluation may be used to control the performance of a teacher. Next, the evaluation could be used to control the movement in or out of a position. Lastly, the evaluation should be used to legitimize the control of the organization.

Teacher evaluation as a performance control works to influence teacher behavior that should lead to expectable practices within set criteria. This type of control can be used in determining merit pay or teacher assignments. The control of movement illustrates a desire to award teacher certification, hiring or firing, and the decision to move teachers up the career ladder. The third purpose involves using the process as a means of legitimizing control of the organization and wrestles with the notion of assigning equality and justice to the organization. Evaluation can be the tool to settle conflicts between the administration and teachers as long as the method of evaluation is seen as democratic (Natriello, 2006).

Papay (2012) added to the research on the purpose for teacher evaluation and stated teacher evaluations can be used in two ways. One is to assess the teacher's effectiveness and the other is to help the teacher grow professionally. These are formative approaches to teacher evaluations, and, according to Papay (2012), most policy debates on teacher evaluation have

focused on the formative component of evaluations, but there are three criteria on which teacher evaluations should be based; the evaluation should be unbiased, reliable, and valid.

Use of Teacher Evaluation

Teacher evaluation can be formative, summative, or both. As a formative tool, evaluations focused more on the professional growth of the teacher, and as a summative tool, evaluation focus more on educational decisions for teachers (Strong et al., 2008).

Formative evaluation, as used in the arena of teacher evaluation, is the process of analyzing the strengths and weaknesses of the educator. These evaluations provide opportunities for both the teacher and the administrator to reflect on the educator's performance, to obtain feedback, and to provide for the professional development of the educator (Dagley & Veir, 2002b).

The summative phase of the cycle shows whether the data, the documentation, and the observations demonstrate the improvements and changes sought. During this summative phase, the administrators make decisions based on the data regarding the teacher's employment status. The dominant statutory use of the evaluation system is for dismissal of problem teachers. Other uses of these evaluation systems delineated by statutes include preparation for hearings, production of evidence, discovery, demotion, immediate discharge, production of exhibits, and, in one state's legislation, tenure.

Methods Used for Teacher Evaluations

Beyond the formative and summative evaluation purposes, Millman, Darling-Hammond, et al. (1990) examined several methods used for teacher evaluation systems. They include

classroom observations, self-assessment, student data, and multiple measurable evaluations. Also there are other methods such as student assessment, teacher portfolios, and performance tests and/or simulations used in evaluating teachers (Millman, Darling-Hammond et al., 1990).

The use of data to determine if a school is meeting AYP is the cornerstone of NCLB, and now policymakers and education leaders are moving toward using data to determine the effectiveness of teachers. As Hill et. al. (2010) contended, over the past decade of American evaluation policy, value-added models have become increasingly popular for both state and federal policymakers. President Barack Obama's education initiative Race to the Top, has pressed for value-added models to be included in state evaluation policy (Hill et. al., 2010).

One of the most common methods of teacher evaluation is classroom observation. Stodolsky (1990) explained how classroom observations are ideal in determining how a teacher interacts with a group of students or the teacher's overall behavior in a classroom. However, a classroom observation will not show how well students digested the lesson or if the students understood the lesson. Observations lead to the belief that what is being evaluated can be judged by sight. Observations do not account for a teacher's response to intervention, planning, assessments, and assessment practices and relationships with other stakeholders (Stodolsky, 1990). Classroom observation is a tool for evaluation but should not be used alone to determine a teacher's effectiveness.

Stodolsky (1990) explained the early nature of classroom observations as a system of impromptu visits by the school principal or designee with the observer focusing on what was seen during the visit. As evaluations have become a part of the accountability system of schools and school systems, different instruments are starting to be used during classroom observations.

Stodolsky (1990) listed these as follows: behavior checklist, narrative records, summaries, category systems, and rating systems.

Teachers should be given timely feedback and given a critique of their strengths and weakness witnessed during a classroom observation. Classroom observations are formative in nature and can provide keys to help teachers gain an understanding of the profession (Zepeda & Kruskamp, 2012).

According to Good and Mulryan (1990), rating scales for teacher evaluations were first introduced in the late 1800s. Rating scales became the primary way of judging a teacher's effectiveness in schools. The first rating scales were designed by administrators, students, and self-rating models by teachers.

There were several rating scales introduced to educators at the turn of the 20th century. They were called the Elliot scale, quality scales, Rugg's rating scale, and Boyce card. These rating scales judged a teacher's performance based on what was considered to be best practices of the time using a numerical system (Good & Mulryan, 1990).

Building administrators and supervisors were the first to use rating scales. The rating scales were used for promotion, transfer, dismissal, judgments, and accountability. At one point, rating scales were used for salary increases, but this was as unpopular with teachers as merit pay is today. As rating scales were increasingly used for teacher evaluations, several complaints were lodged against this method. The following list states the complaints: rating scales were unprofessional, rating scales were not consistent from teacher to teacher or year to year, rating scales were unfair and discriminatory, and rating scales did not lead to good teacher practices or student learning (Good & Mulryan, 1990).

According to Bird (1990), the idea of teacher portfolios for evaluation purposes was borrowed from other professions. Portfolios are a collection of works gathered and presented to demonstrate a person's workmanship. Bird (1990) goes on to say that portfolios should improve the quality of teaching and help with personnel decisions. Portfolios give visual evidence that can help with bringing new ideas into the evaluation process. However, portfolios may not give an accurate illustration of what is being taught. Additionally, portfolios used alone could lead teachers' attention away from the more pressing issue of teaching because they would be focused on developing the portfolio.

Bird (1990) gave examples of how portfolios can be used as a valuable tool for evaluation. School teacher entries, jointly-produced entries, and entries by others are considered the three main participation modes of completing a portfolio. Entries provided solely by a schoolteacher would consist of elective entries, guided entries, and required entries. Each entry by the schoolteacher has its purpose and can be readily distinguished in a portfolio. As school teachers collaborate with others, their portfolio entries produce a collegial product for the evaluatee. Finally, entries produced by others consist of those entries that may be recommendations, awards, or observations. The options provide an array of items to help evaluators see a glimpse of classroom practices.

According to Haertel (1990), new methods of obtaining a teacher's knowledge and skill level are being introduced. These new methods promise to give a clearer picture of the effectiveness of teachers. Performance test and simulations have given evaluators more tools for assessing teacher performance. These teacher assessments are administered outside of the actual classroom, but are conducted using tasks and assignments that simulate real-life teacher duties.

These exercises are based on what has been determined to be best practices for effective teaching but cannot be assessed by multiple choice tests or a routine observation.

Haertel (1990) explained the reasons for new ways of assessing teacher behavior as the expanding of teacher testing. Teacher testing is exerting an influence on teacher education curricula and having a valid and reliable teacher licensure process.

Evaluation Data from Multiple Sources

According to Mehrens (2006), using multiple data sources for teacher evaluation should use one of two approaches. The use of multiple data for professional growth may not be the best option. Skills for professional growth need to be analyzed on an individual basis. On the other hand, when basing a single decision on the evaluation process, multiple data is the better way to make an informed decision.

Utilizing multiple measures is the best way to determine the most effective forms of teaching. The most common multiple measures are classroom observations, artifacts of teachers' work, student surveys, and tools to measure student progress (Stronge, Xianxuan, & Leeper, 2012).

Classroom observations have been the foundation of teacher evaluation models. These observations can serve to fulfill the obligation of legal requirements and can help determine what is needed for professional growth (Zepeda & Kruskamp, 2012). According to Zepeda and Kruskamp, the most common procedure for evaluating teachers, classroom observations, served the purpose of student and teacher improvement.

Teacher artifacts can be used by teachers as a complement to show what may not have been observed during a classroom visit and serve as a good precursor for professional development (Stronge et al., 2012).

Student Growth Models

Gordon et al. (2006) explained the need to use student achievement as part of a teacher evaluation system. Student achievement as a performance measure should not be used alone but as tangible and objective evidence that can be used by school administrators. Ehler, Koedel, Parsons, and Podgursky (2013) also explained that the growth model approach to teacher evaluation has gained momentum among researchers and policymakers. Growth modeling gives clearer pictures of what is happening in a certain school regardless if it is a high achieving school or a low achieving school.

The passage of NCLB required students to meet certain academic standards that lead to 100% proficiency by 2014; this was called Adequate Yearly progress and it was based on the status model approach (Hull, 2007). According to Hull (2007) and O'Malley et al. (2011), growth models are designed to measure a student's progress over two different points in time. Hull (2007) discussed five types of growth models. The improvement model or status model measures scores of students in one grade with different students in the same grade. Status models tend to show whether students can reach a benchmark on a certain test, but the information taken from a status model does not show if underachieving students are making growth or if high achieving students are losing growth (Stronge et al., 2012).

Under NCLB this model was called Safe Harbor. The performance index model measures the degree students who are performing at the lowest academic level have progressed. The model

does not measure students' individual academic growth, but the amount of change in a school's performance from one school year to the next. Simple growth models record change in a student's test scores from one grade to the next. The growth to proficiency model shows stakeholders if students are on track to reach a designated proficiency level or if they are scoring above proficiency.

Other research has found growth-based measures are becoming increasingly common in the newly designed teacher evaluation systems that have high stakes attached; however, research has shown that the effects of test-score growth models are viewed differently by teachers and schools (Ehler et al., 2013). Ehler et al. (2013) examined three approaches to the use of test-growth models for the evaluation of teachers. The approaches are student growth percentiles (SGPs), one-step value-added model (VAM), and VAM that compares observationally similar students.

The SGPs model does not control for other factors that influence a student's performance on a test. This model compares students to their peer group and is descriptive in nature and is usually used to generate further discussion, and they do not recognize causal effects (Ehler et al., 2013).

Finally, the value-added model is a model used to measure change in a student's academic performance over time, but it also incorporates outside influences such as the school, teacher, education program, and previous teachers. Value-added models are a form of a growth model, but not all growth models are value-added models (Hull, 2007).

Value-Added Models (VAM) of Teacher Evaluation

Over the past decade, value-added models have become increasingly popular in trying to determine school and teacher effectiveness as well as merit pay systems (Hill, Kapitula & Umland, 2011). Strong et al. (2008) also identified the contemporary research as having seen a shift in focus of teacher evaluations, with more focus on what value a teacher brings to students. As value-added models for teacher evaluation continues to be studied it is evident that teachers are most essential for students' success or failure in schools.

Papay (2012) stated there are two systems of teacher evaluation tools: value added models and standards-based models. Policymakers have become more enamored with the value-added models because they are quantitative, which allows policymakers to assess student test growth; it is fairly inexpensive to gather the data and the data are believed to be both valid and reliable. Papay (2012) explained that as technology has continued to improve so have new ways of evaluating teachers. Researchers now have databases with collected data that help determine how much effect teachers have on a student's academic achievement. This new technology has given rise to the value-added models for teacher evaluation. He suggested that whether implementing either a value-added model or a standards-based model for teacher evaluation, there must be a commitment to the process and careful implementation of the design.

The Race to the Top fund places a high emphasis on states designing and developing teacher evaluation systems that are rigorous, transparent, and fair, and that will take into account student growth (*Race to the Top*, 2009). The value-added model of teacher evaluation uses student data to determine student growth. Value-added models (VAM), as a teacher evaluation system, use statistics to measure a student's progress from year to year by using test scores and other variables that may influence student achievement (Darling-Hammond et al., 2012).

Researchers believe a teacher's effectiveness can be determined by using VAM (Darling-Hammond et al., 2012). When using value-added models for teacher evaluation, one must control outside factors that may influence the baseline test scores of the students. This is because teachers should be held accountable for student growth, not what students have or have not learned in their past educational experiences (Gordon et al., 2006).

Ehler et al.'s (2013) approach to value-added models is the one-step VAM, which sets up controls for a student's outside influences as well as those influences that may arise in their school environment, while the VAM design, which compares observationally similar students, works to reflect the teacher as he/she compares to other teachers teaching similar content. The one-step VAM is the model most used in research studies that are trying to determine the effectiveness of teachers and schools.

Scherrer (2012) explained the arguments of using VAM for evaluating teachers as researchers debate over the reliability and the margin of error of VAM, while others rave about the possibility of having valuable feedback for the use of VAM. Policymakers must understand the difference in the student population of schools. Research shows schools in poor areas are at a disadvantage from those in more affluent areas thus their VAM scores should be higher. The thing researchers are in agreement about is that the VAM is a tool that is here to stay.

Problems with Teacher Evaluations

With any system of evaluation there will be problems associated with its procedures, implementation, or validity and reliability. Good and Mulryan (1990) discussed the historical problems with teacher evaluation as it related to the gaining of knowledge for improving teacher and student behaviors. The four problems examined were external evaluation, complexity of

teaching, lack of attention to context or theory, and minimum of observations. Teacher evaluation as an external evaluation is the practice of evaluating for outside accountability and not for teacher professional development. Complexity of teaching wanted to place numerous behaviors and personality traits on each teacher as if to use the cookie cutter mold to evaluate teaching. The lack of attention to context or theory eliminated the idea of different methods for different grade levels or the content that was being taught. Finally, classroom observations were few and far between as evaluators thought they could evaluate personalities and behaviors without being in the classrooms.

As time moved into the 1980s and 1990s, the same problems that existed at the beginning of the century were still happening. Six reasons for the continuous problems were given: public ambivalence about the professionalism of teaching, no investment in teacher evaluation, no variation in procedures to evaluate different goals, not enough research on the different variables affecting teacher evaluation and student outcomes, schools tend to either rely not enough or too much on evaluation research, and evaluation has become a rote ritual (Strong & Tucker, 2003).

Strong and Tucker (2003) also discussed several reasons why many teacher evaluation systems have failed. The lack of screening personnel when awarding certification, individual educators are not receiving beneficial feedback, there is no acknowledgment for outstanding services teachers provide, staff development is not ideal, the evaluations are not cost affective, the reliability and validity of the evaluation will not hold up to judicial analysis and/or professional inspection, the ability for teacher evaluation models to help in the termination of ineffective teacher, and the models do not help in creating an inclusive environment among teachers and administrators.

Strong and Tucker (2003) highlighted many of the concerns with the way teacher evaluation systems have been designed and implemented in schools. The prevailing literature shows a distant belief that teacher evaluations do very little to improve teaching and learning. Also there is the concern of evaluators giving a true evaluation score for fear of legal ramifications associated with an unsatisfactory performance rating. One of the biggest concerns in recent years is the cavalier approach to teacher evaluations as they were not seen as a tool for improving the quality of education. Finally, the lack of trust that has become imbedded in our schools led to a dysfunctional evaluation approach (Strong & Tucker (2003).

The problems associated with teacher evaluation models continue as Varlas (2012) noted, many of today's teacher evaluation models are not built to encourage teachers to perform at a highly effective level and thus do not create an atmosphere for improving student learning. Finally, Papay (2012) believed a teacher evaluation model that promotes continued teacher development will be most valuable in improving student learning.

Teacher Evaluation for School Improvement

According to Iwanicki (1990), there are three parts to an effective teacher evaluation system. They are philosophy and purpose, performance criteria and standards, and evaluation procedures. Developers have placed more emphasis on the accountability and professional growth models for teacher evaluation, but teacher evaluation can play a very vital role in school improvement. As target objectives are devised for school improvement, then those objectives should be integrated into the teacher evaluation process (Iwanicki, 1990). Iwanicki (1990) explained that by using teacher evaluation as a tool for school improvement, three distinct advantages to the process are created. Trust and collegiality is built within the staff, the

evaluation becomes a part of the school and not just a routine for the administration, it becomes more visible to the school and has a greater impact on the school improvement process.

Licensure Test for Evaluation

One level of evaluation is requiring prospective teachers to take a test before a certificate or license to teach is awarded. The licensure procedure is an attempt to ensure all teachers have the minimum competence before being allowed to teach and keeping incompetent people out of the profession. The licensure is not about determining if a teacher is going to be effective or not just that they have the minimum skills to instruct students (Madaus & Mehrens, 1990). The most commonly used type of licensure test has been a multiple choice paper-and-pencil test. Several states including Alabama have started using the Praxis as its entry test into the classroom.

The State of Alabama requires all prospective teachers to take *Praxis II* tests for each area of certification they seek and pass the basic skills assessments required in the Alabama Prospective Teacher Testing Program (ALSED, 2012). Mississippi requires that all educators pass a rigorous test to comply with licensure guidelines and with No Child Left Behind standards. The Praxis series are the tests approved by the Mississippi State Board of Education (MSSED, 2012). Tennessee requires the Praxis I Pre-Professional Skills Tests for admission into an accredited teacher education program and students must pass the Praxis II Subject Assessments, which include the Principles of Learning and Teaching portion and separate Subject Assessment tests for each area of endorsement sought (TNSDE, 2012). The state of Georgia has a basic skills test to become a certified teacher and the subject area competence of the Georgia Assessments for the Certification of Educators (GADE, 2012).

According to Gordon et al. (2006), the practices used today for hiring and retaining teachers are flawed in two ways. Their first concern deals with how teachers are awarded their credentials versus their effectiveness over time. The second flaw discussed is how systems track teacher effectiveness after they are hired. Researchers explain that very little tracking is done. Evidence examined shows no relationship between teacher certification and teacher effectiveness (Gordon et al., 2006).

Legal Issues

Rebell (1990) discussed the movement of the legal system as teacher evaluations have gone from being periodic and having no discernible standards to a system of multiple measures that look to improve the quality of teaching. Rebell examined the shift in the courts' attitude from allowing school evaluators to judge the process to taking a critical look at the processes. However, according to Gordon et al. (2006), very few teacher contracts are non-renewed from year to year leaving teachers who are effective and ineffective in schools that may or may not be performing at an acceptable standard.

Rebell (1990) chronicled the change in judicial opinions on teacher evaluation starting with the 1973 *Scheelhasse v. Woodbury* decision by the United States Court of Appeals for the Eighth Circuit. In the *Shcheelhasse* case, the judge dismissed the teacher's claim for being terminated using student test data versus expert witnesses who stated this was not a valid way to determine termination as not being a matter of constitutional dimension. Fifteen years later, in the *St. Louis Teacher's Union v. Board of Education*, the District Court would not dismiss the teacher's claim that using student achievement gain for the majority of the evaluation process was not designed for this purpose, it was not valid, it was capricious, unconstitutional, and it was

very arbitrary. Rebell (1990) gives four factors that influenced the court's thinking on teacher evaluation:

- Improved evaluation techniques
- Disproportionate number of teachers or would be teachers are members of minority groups whose grievances are given special scrutiny by the courts.
- Legal developments have broadened the jurisdiction, especially of federal courts to consider social reform issues, including testing matters
- Judge's own experience in assessing psychometric techniques in employment discrimination cases and their familiarity with educational reform issues from desegregation, special education and other school-based litigations (pp. 338-339)

Courts have given a positive sign to evaluation systems that have in place objectivity and procedural regularity. However, Sullivan and Zirkel (1998) explained there is still a disconnect between professional wisdom and the law of teacher evaluation.

Politics and Teacher Evaluation

The power to create policy and enforce policy in educational settings is vested with school boards, rendering teachers and administrators as non-political agents. However, teachers and administrators must participate in the process as new educational policies are being designed and, in particular, new teacher evaluation systems (Stronge & Tucker, 1999). The design of a valued and valid teacher evaluation system needs to understand and incorporate the political temperament of the policymakers and have a constructive dialogue between all stakeholders. The design has to encompass the ideas of communication, commitment, and collaboration. Working together is paramount to the success or failure of a complete teacher evaluation design (Stronge & Tucker, 1999).

Policymakers have embraced the idea of using the value-added model approach for teacher evaluation as shown in the Race to the Top competition (Hill et al., 2010). However, as

Johnson (1999) explained, a lack of consensus among stakeholders can cause a new evaluation policy to fail like the policy initiative in Louisiana.

Summary

The intent of this literature review was to gain an understanding of teacher evaluation policy. The review has drawn on the education reform movement beginning in the 1960s with President Lyndon Johnson's War on Poverty and the implementation of the Elementary and Secondary Education Act, President Reagan's *A Nation at Risk*, President George H. W. Bush's America 2000, President Bill Clinton's Goal 2000, President George W. Bush's No Child Left Behind Act, and President Barak Obama's Race to the Top competition.

The review of the literature also examined different types of education reform and school accountability tools beginning with an overview of school accountability and then moving its focus to common core standards, high stakes testing, merit pay, and charter schools. All of these were viewed as ways to improve education in America. However, the bulk of the literature review centered on teacher evaluation policies.

Teacher evaluation has become a policy target for policymakers. Policymakers are looking for a tool to judge the effectiveness of schools in America. This effort is to ensure all students are taught by an effective teacher. However, the literature on teacher evaluation policy gives mixed signals about the most effective procedures and designs of teacher evaluation; as Papay (2012) stated, many stakeholders view teacher evaluation in its present state as broken and in need of immediate repair.

Education literature reveals that there have been numerous reforms, initiatives, acts, and competitions released to improve the education of America's students. A cursory look at the

teacher evaluation literature reveals a disconnect among stakeholders on the purpose of the new teacher evaluation policy designs and with the purpose of each new design in question; the actual procedures are also being debated.

The history of teacher evaluation leads us to understand that evaluations have been a part of the educational system since the inception of formal education. The literature illustrated for us the need for teacher effectiveness to be judged and gave us one of the first observational development research tools from the committee on child development of the National Research Council.

Teacher evaluation as a tool to judge school effectiveness was born out of the need for a need to evaluate teachers in an ever changing school setting. However, as Strong and Tucker (2003) explained, any good teacher evaluation tool must have three functions, communication, commitment, and collaboration. Having these three functions one can define teacher evaluation, as Goldrick (2002) did, as professionally assessing a teacher's work performance. Nevertheless, many of the evaluation policies being used are viewed as being ineffective. A consensus in the literature is to allow teachers to sit at the table when new evaluations designs are discussed. Teachers must be a part of the development process and have equal input. Teachers must be active participants, not just allowed to sit at the table.

A commitment to new teacher evaluation designs need to be explicit and stated by all stakeholders, policymakers, educators, and civic leaders. The commitment should include the time and money needed to produce an effective policy for teacher evaluation. These new teacher evaluation designs must aid in improving teacher quality and growing teachers professionally.

Teacher evaluations must be valid and reliable with an understanding that the evaluations are not used for punitive purposes. The research shows that novice teachers feel that veteran

teachers are better suited to objectively evaluate them, while veteran teachers are comfortable with the building administrators as evaluators. However, concern comes with the formative or summative nature of the evaluation tool. How do schools remove ineffective teachers? One of the primary means is an effective teacher evaluation model. The inadequacy of teacher evaluation methods in the 1980s led to very little legislation being passed, thus leading to a void on the topic.

After the publication of *A Nation at Risk*, policymakers began to dissect teaching in America's public schools. Data are being collected to analyze the value teachers are or are not adding to student performance. The search began to find the purpose and method to evaluate teachers.

Educational stakeholders informally evaluate teachers daily; however, there is a need for defined purposes or a more formal means of evaluation. Researchers have found several purposes for evaluations, as a needed control for teacher performance, control movement in an organization, legitimized the control of the organization, and help teachers become better teachers. These purposes cover both the formative and summative evaluation processes.

Gaining meaningful data is imperative in any evaluation process. Classroom observations, self-assessment, student data, portfolios, high stakes tests, simulations, and the use of multiple measures serve as data collectors. Classroom observations are one of the most common forms of collecting data for teacher evaluations. Though good for collecting data on teacher behavior, it does not give real data on whether the students are retaining what is being taught. Portfolios are another form of data collecting for teachers, but again it does not show if the students are retaining the information. The use of multiple measures is gaining momentum as the clear, reliable, and fair way to assess teacher influence on student achievement.

Multiple measures should be used as a data collector when making an informed decision about a teacher's impact on student outcomes. Multiple measures include all of the teacher data collectors plus the use of student growth models or value-added models.

Growth models look to explain what is taking place academically in a school regardless of the type of school being observed. The growth model approach is designed to measure a student's progress over two points in time. There are several different growth models being used from the status model, performance index model, simple growth models, and growth to proficiency model to name a few. The No Child Left Behind Act used both the status model and the performance index model to determine a school's accountability status.

Teacher evaluation models have recently used three types of test-growth models in their designs: the student growth percentiles (SGP), the one-step value-added model (VAM), and the value-added model that compares observationally similar students. The SGPs do not control for outside influences on student performance, it compares students in a particular peer group. The VAMs are used to measure a student's growth over time and is seen as more reliable than the SGPs.

VAMs are a form of growth model, but not all growth models are VAMs. VAMs have become the tool of choice in helping determine a school/teacher's effectiveness and with governing how merit pay is distributed. As technology continues to improve, policymakers are becoming increasingly infatuated with VAMs. These models are quantitative and seen as being easy to gather and interpret the data. VAMs use statistical measures to judge a student's progress from year to year, but also to control for outside influences on the student's progress. This control for outside influences allows teachers to be judged on the growth a student has made during the year he/she was taught not on past knowledge or lack of knowledge. VAMs are not

without their critics, some researchers are skeptical of some of the statistical measures being used.

There are problems associated with most educational tools, but historically there are several associated with teacher evaluation models. They are external evaluation, complexity of teaching, lack of attention to context or theory, and the lack of a set number of observations. Teacher evaluations have become associated with a routine ritual that is done every year by the principal with no intrinsic value. The prevailing literature shows that teacher evaluation in its present form does little to improve teacher performance.

There is, however, a movement that believes teacher evaluation can create positive change for schools. When the evaluation becomes a part of the school climate, more teachers will buy into the concept and thus it will begin to have an impact on school improvement. The use of teacher licensure as an evaluation tool for prospective teachers has gained momentum, also. States are requiring teachers to pass a high-stakes test in order to receive their state teacher's license.

Finally Chapter II examined the literature on legal issues and the politics of teacher evaluations. Courts are going to continue to review how evaluations are conducted when school organizations begin to use them in a summative way.

As policymakers influence educational policies more and more, teacher evaluations become political in nature. Researchers are encouraging educators to join the political process during the debate to create new evaluation systems because policymakers are becoming increasingly interested in the VAM approach to teacher evaluation and the percentage it will count toward a teacher's evaluation.

Teacher evaluation as a policy target is becoming a progressively hot topic among all education stakeholders. Chapter II has tried to give the reader an understanding of the history of teacher evaluation and the approaches that have contributed and will continue to contribute to the design of new teacher evaluation policies.

CHAPTER III

METHODOLOGY

The purpose of this study was to take a comparative look at five states' teacher evaluation policies and analyze these policies using qualitative methods. The study used these comparisons in hopes of sharing constructive feedback to stakeholders and policymakers in Alabama as they design and implement a new teacher evaluation policy. The researcher investigated educational documents and educational programs and took a comparative look at the documents and programs to provide insight on teacher evaluation for the state of Alabama. In this process of inquiry, the study methodology was deemed suitable for a qualitative approach.

Creswell (1998) defined qualitative research as “an inquiry process of understanding based on distinct methodological traditions of inquiry that explore a social or human problem” (p. 15). Under the qualitative methodology the study used a multiple case or collective study approach.

Specifically, this study focused on new teacher evaluation policies and new teacher evaluation designs and methods in four states within the United States of America. Teacher evaluation designs and methods were used from states in the South, Midwest, and Northeast.

This study added to the body of knowledge in the field of Educational Leadership by analyzing what was working and what was not working with new teacher evaluation designs and methods. The hope was to gain a clear understanding of teacher evaluation policies and to be able to recommend practices and procedures that have worked in the states that were analyzed by the Alabama Department of Education and the research community.

Rationale for Research Strategy

Strauss and Corbin (1998) explained qualitative research as a method used that will generate answers that are not derived from statistical methods. This study undertook the task of researching state level teacher evaluation policy and how new designs will effect Alabama's teacher evaluation policy in the future.

One of the first questions to ask when deciding on undertaking a qualitative study is how the research questions is asked: usually they began with a how or what (Creswell, 1998). This study had four research questions all asking what as pertaining to teacher evaluation policy. As teacher evaluation policy gains more and more momentum, a study is needed to explore the contributions or lack of contributions it adds to effective teaching. The second reason to use qualitative research is when a topic needs to be explored. Creswell (1998) further explained that qualitative research is the desirable design when a topic is in need of a detailed explanation. As policymakers in the states researched are pushing for more and rigorous teacher evaluations, more research needs to be added to the body of knowledge. Although there are several approaches that can be explored, I will conclude with the final reason Creswell gives. This approach is that of being a storyteller from a participant's view of active learning. Teacher evaluation policies are changing in an effort to improve student learning. This project will compare and contrast the essence/nature, methods, and practices of four state evaluation systems and compare these with the literature on teacher evaluation as well as report the successes and failures from the policies.

Research Design

This study is a comparative policy analysis using qualitative methods. A comparative policy analysis by definition is an item-by-item comparison of two or more comparable alternatives, processes, products, qualifications, sets of data, systems, or the like. According to Marshall (1998), the method of qualitative comparison was coined by Charles Ragin in 1987. Ragin based his technique on the need to compare causal inferences on a small number of cases (Marshall, 1998).

Unit of Analysis

The unit of analysis for this study was the state level. Specifically, this study examined teacher evaluation policies as enacted by five states and compared/contrasted these policies so as to inform educational policymakers in Alabama.

Study Data

Much research is needed to gain a clear understanding of the changes taking place in states to improve teaching. The researcher first drew on the professional literature of teacher evaluation policy. In gathering the professional literature the researcher looked to some of the leaders in the field of educational reform, evaluation, and organizational change. To get a true picture of teacher evaluations the researcher will pull information from Tennessee, Georgia, New York, Illinois, and Alabama.

Data Collection Methods

The data collected for this study were gathered from evaluation policies, procedures, and methods of the newly designed teacher evaluation policies of five states throughout the United States. The states were chosen from the different geographical regions: the South, Midwest, and Northeast. The states were chosen to give a clear understanding of newly developed teacher evaluation policies from across the United States. Each state's teacher evaluation policy was analyzed in an effort to determine its purpose, how it was implemented, the design, the strengths, and the weaknesses. Also, information was collected from outside agencies such as teachers' unions and policymakers to gauge the attitudes of all stakeholders.

The data for this study were found on the state department of education websites for each of the states researched. Also, the pertinent literature on teacher evaluation and educational reform was taken from scholarly research and relevant newspaper articles that reflected the pulse of the nation on the development of new teacher evaluation policies and their use for determining teacher effectiveness. Newspaper articles were also used to determine the successes and failures in the states studied. The data were in document form.

Document analysis as defined by Heffernan (2001) is a social research method and is an important research tool. Document analysis involves reading written material and can be inductive, interpretative, and social constructionist. The idea of document analysis is to inductively generate novel theoretical ideas or hypotheses from the data. This study critically studied the documents of teacher evaluation policy and attempted to give a critical analysis of the new designs and not just a simple description.

The documents that were used for this study were obtained from the state department education websites for each state. The documents were readily available for all stakeholders to

examine. To gather additional information concerning each of the states' new policies, documents were surveyed from other sources. Newspaper articles from some of the leading media outlets, such as *The New York Times*, *Atlanta Journal Constitution*, and *Chicago Sun-Times*, provided a good look into the feelings and beliefs of teacher groups, policymakers, and academics as these policies were being debated.

Sampling Logic and Strategy

The sample for this study came from evaluation policies, procedures, and methods of the teacher evaluation policies of four states throughout the United States. The states were chosen from the different geographical regions; the South, Midwest, and Northeast. The researcher wanted to measure the pulse of the nation and its attitude toward teacher effectiveness and teacher evaluation policies.

The sampling strategy that was used was a purposeful sampling. A purposeful sampling is when cases for study are selected because they are information rich and illuminative, that is they offer useful manifestations of the phenomenon of interest; sampling, then, is aimed at insight about the phenomenon, not empirical generalization from a sample to a population.

The researcher collected a number of reports documenting pros and cons of evaluation policies taking place in the states mentioned above. The documents provided data to describe and assess teacher evaluation policy in each of the respective states.

New York was chosen to evaluate because it had the largest public school system in the United States. New York City has approximately 1.1 million students in 1,700 schools. The teacher workforce is estimated at 75,000 and New York has an annual budget of \$24 billion.

New York also was implementing a new teacher policy that may be the test case for the rest of the United States' public schools.

Tennessee was implementing a new value-added teacher evaluation system and its proximity to Alabama made it an ideal state to add to the study. Tennessee also had awarded over \$501 million dollars under President Barak Obama's Race to the Top fund, which lends itself for the study of new and innovative techniques to improve education. Tennessee aligned its teacher effectiveness measures with Race to the Top guidelines and referred to its programs as being First to the Top. Tennessee had over 137 school districts with more than 1,700 schools.

Georgia's department of education operated over 2,200 schools with an approximate student enrollment of over 1.6 million students. Georgia employed over 100,000 certified teachers. Georgia's proximity to Alabama also loaned itself for this study. However, Georgia had implemented a new teacher evaluation design under the Race to the Top initiative called Teacher Keys Evaluation System. Not only did the creation of Georgia's new teacher evaluation system intrigue the researcher, but also the guidance given to Georgia by one of the preeminent scholars on teacher effectiveness and evaluation, Professor James K. Stronge.

Finally, the state of Illinois and the city of Chicago, in particular, had implemented a new teacher evaluation design that gained global notoriety as teachers in Chicago went on strike to protest the weight that the student data would count toward a teacher's evaluation score. The intrigue behind the strike led the researcher to search out what the new design entailed and what frustrated teachers about this design that forced them to strike. The state of Illinois had 4,493 public schools that educate over 2.1 million students. The city of Chicago operated 681 public schools with over 400,000 students and over 23,000 teachers.

Data Analysis Methods

The data collected for this study were analyzed using a data comparison matrix. The idea of using a data comparison matrix was found in the literature of Professors Helen M. Hazi and Daisy Arredondo Rucinski in 2009. Their paper used a data comparison matrix as a Fifty-State Review of Statute and Regulatory Action since NCLB, with teacher evaluation as the policy target. The data were collected for this study using states' department of education websites, teacher evaluation literature, and newspaper articles. The data comparison matrix was designed using the works of Professors James Popham, Linda Darling-Hammond, and James Stronge and the data comparison matrix designed by Professors Helen M. Hazi and Daisy Arredondo Rucinski.

Upon consultation with Dr. Bob L. Johnson, Jr., five primary anchors followed by sub-anchors were added to the matrix. The data comparison matrix consisted of several anchors that were used to analyze and compare/contrast the teacher evaluation policies enacted by the states in the study sample. These five primary anchors and subsequent sub-anchors are as follows:

- 1. Policy Origin**
 - a. What prompted the need for a teacher evaluation policy?
 - b. Who initiated it?
 - c. Who was involved in creating the new evaluation policy
- 2. Policy Purpose**
 - a. Is the purpose of the policy implicit or explicit?
 - b. Is the purpose of the policy clearly stated and logically consistent?
 - c. Is the policy being implemented/pursued in a manner consistent with its stated purpose?
- 3. Policy Design –**
 - a. Mandated for all districts or is voluntary?
 - b. If voluntary, are there state inducements/incentives?
 - c. Who is responsible for coordinating the policy in the state?
- 4. Policy Components, Mechanisms and Procedures –**
 - a. Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?

- b. Evaluative rubric
- c. Types of data used to evaluate?
- d. How data is collected? (i.e., By whom? How often? What means?)
- e. Mechanisms for training evaluators? Of what does it consists? How is it done?
- f. Procedures used to ensure validity and reliability of evaluation?
- g. How are data generated from teacher evaluation used? For formative or summative purposes? How specifically?

5. Policy Effects and Feedback: Pro and Con –

As noted from various sources and by various groups (e.g., evaluation studies, newspaper reports, teachers' union reaction, lawsuits, etc.), what positive and/or negative reactions and effects OR problems/successes has the policy produced?

It should be noted that consistent with the emergent nature of qualitative research, further anchors could be added or subtracted from the matrix as the study unfolded.

Standards of Rigor in Study

The validity of a qualitative study is paramount. Wolcott (1990) explained the need for maintaining validity of qualitative research. The researcher must do the following: be a listener, record accurately, initiate writing early, include all data in the final study, be candid, seek feedback, attempt to achieve balance, and write accurately.

Creswell (1998) explored and expounded on eight procedures for standards of quality and verification in qualitative research. He stated at least two of the eight approaches should be used to complete a qualitative study. Creswell listed the procedures as follows:

1. Prolonged engagement and persistent observation
2. Triangulation
3. Peer review and debriefing
4. Negative case analysis
5. Clarifying researcher bias
6. Member checks
7. Rich, thick description
8. External audits

Lincoln and Guba (1985) also gave a model to help with validity and rigor of a study so not to sacrifice relevance. The model gives four criteria, beginning with credibility. Creating a creditable study consists of prolonged and varied engagement, peer debriefing, triangulation, and member checking. Second is transferability, which asks can the findings be transferred to other studies. Next is dependability, the study must have an audit trail. The audit trail suggests that the research process has to follow a rational methodological and interpretative process. The audit trail leads the reader to follow the path of the research even if the reader does not agree with the findings. Finally, confirmability is defined as a degree of neutrality or the extent to which the findings of a study are shaped by the respondents and not researcher bias, motivation, or interest.

Because this study is a document analysis, some of the procedures given by both Creswell (1998) and Lincoln and Guba (1985) do not lend themselves for verification. Clarifying researcher bias, triangulation and rich, thick description (transferability) was used as the standards of rigor for this study.

Clarifying researcher bias was described by Creswell (1998) as when the researcher states his position from the outset of the analysis, by explain any biases or assumptions that may have any bearings on the findings. I have attempted to explain my bias and closeness to the research in the section that follows. As a public school educator for 17 years, I have been evaluated and an evaluator under Alabama's teacher evaluation policy. The researcher brought to this study a perception of how teacher evaluations should work and how they should be constructed to get the maximum effectiveness from teachers.

Rich, thick description or transferability allows the reader to transfer the findings to other studies because of the detailed description of the study. Due to the purposeful sample that was used in the study, transferability should be easily obtained. The state documents examined were

the latest information on state teacher evaluation policies, since the implementation of President Obama's Race to The Top competition. The scholarly peer reviewed literature was gathered from the leading researchers in the education field of teacher evaluation policy. Finally, newspaper articles were studied to present findings on the success and failures of these new policies. This should give a thickness to the study allowing the findings to be easily transferred.

An external audit is when an outside consultant examines the process and the overall content of a study for its accuracy (Lincoln & Guba, 1985). The consultant should have no connection to the study. The external audit for this study was conducted by the study's chairman and other committee members. The audit was conducted through a presentation of the proposed study to the chairman and the committee. During the proposal process the researcher presented the purpose of the study, the design of the study, and how the data would be collected and analyzed. Once the proposal was accepted, the final project included five chapters and was presented to the chairman and the committee for a final external audit. The final external audit was comprised of an analysis of the summary and findings, implications, conclusion, and recommendations for future research. The chairman and the committee evaluated the study by determining whether the findings, implications, and conclusion were supported by the data.

Triangulation of the data was used as the final test for rigor. Triangulation is when the researcher uses multiple sets of data to provide corroborating evidence. The researcher attempted to show how the interpretation of the data from the states' teacher evaluation policies is judged by teachers, policymakers, and other stakeholders. Triangulation helped give the interpretation of the data more reliability and the study attempted to show consistency in the data and the data interpretations.

Situating Self as Researcher

I began my professional education career after receiving my master's degree in secondary education with a concentration in social science. I began teaching United States History and Economics in a high school located in north Alabama in 1997. I was introduced to teacher evaluation during my first year of teaching under the newly develop Alabama teacher evaluation model called the Alabama Professional Education Personnel Evaluation (PEPE). As a novice teacher, I was enamored with the process of teacher evaluations. Gaining help and insight from veteran administrators was going to help advance my teaching career and make me a better teacher.

I continued my education by gaining an administration certification and Educational Specialist degree in administration. I was offered a position as an assistant principal at a middle school in the same north Alabama school district. All new administrators were required to attend a weeklong conference to be certified as a PEPE evaluator. Once I completed the training, I was tasked with evaluating teachers. This assignment as an evaluator gave me my first look into the process of evaluating teachers. The process sparked an interest in the evaluation of teachers, as PEPE's goal was to improve teaching and learning. I realized early on that using PEPE was not helping teachers improve instruction. The process was very cumbersome. The classroom observations that required scripting (writing every word the teacher or students said) and later going back assigning a numerical grade proved troublesome. The number of teachers each administrator was given to evaluate was too time consuming and no productive feedback followed. PEPE seemed like a paper trail for the district and not a tool to help students and teachers become the best and most effective students and teachers possible.

The process of evaluating teachers began to pique my interest as a research topic while I was completing my Educational Specialist degree. Also, I read Jonathan Kozal' *Savage Inequalities*. This text struck a chord with me heightening my sense of responsibility to the students and teachers with whom I worked. How can education be improved for the neediest and highest achieving students? Can this idea coexist in America's schools? I wanted to improve education by giving students the best teachers possible and give the teachers the input and encouragement to be successful.

The biases that I brought to the study is that teacher evaluation in its present form is not working and using these designs to judge or reward teacher performance is a travesty. As the study progressed, I allowed the research to guide the findings.

The personal goal for this research rested on helping students and teachers become the best they can be. Students deserve the most effective teachers and teachers deserve the tools to become effective. Growing as an administrator in a middle school, I have seen the frustration of both students and teachers as they try ways to achieve their goals. My belief is that through academic research and rigor and not for profit research we can find common ground to improve both teacher and student effectiveness. Professionally, I would like to help teachers grow professionally and to trust the academic research that is being produced daily. This professional goal has allowed me to delve into the research and gain many perspectives on not only teacher evaluation policy, but the value in research itself. Also, with this study I hope to contribute to the literature debate on teacher evaluation policy as a tool for improving schools.

Limitations of Study

There were multiple limitations to this study. First a purposeful sampling method was used and only four states were researched for an analysis of teacher evaluation policy. This gave a cursory look into teacher evaluation policies, as all 50 states began to evaluate their designs. Hence, the findings from this study cannot be generalized beyond the states included in the study.

The researcher began the study of documents dating back to the 1960s, knowing that formal public education began some time before. The researcher was also limited by the use of only a few attempts to improve education in America through a limited number of Presidential reforms, initiatives, and competitions.

Summary

In this chapter, the researcher described the methods of this comparative analysis of teacher evaluation policy concerning the use of new teacher evaluation policies, procedures, and methods. The unit of analysis for this study was a comparative multiple state teacher evaluation policy analysis. This was chosen as a way to examine new teacher evaluation policies in the hope of gaining understanding of the new and innovative ways of measuring effective teaching. This study anticipated adding to the body of knowledge of evaluation policy and school accountability research.

CHAPTER IV

NARRATIVE AND SUMMARY OF A COMPARISON DATA MATRIX OF STATE TEACHER EVALUATION MODELS

Chapter IV takes a comparative look at new teacher evaluation models in five states. The purpose of this chapter is to provide a descriptive analysis of the five states' teacher evaluation policies and models using a comparative data matrix. The matrix was constructed using the literature on teacher evaluation. The states compared in this study are Georgia, Tennessee, New York, Illinois, and Alabama. The evaluations created by Georgia, Tennessee, New York and Illinois were designed in an effort to win money under the Race to the Top initiative fashioned by the Obama administration.

Chapter IV also includes a data matrix to illustrate the narrative and summary of teacher in evaluation in Georgia, Tennessee, New York, Illinois/Chicago, and Alabama. The matrix is designed to give a visual comparison of the states being researched. The information contain in the matrix will come from the referenced data in the text.

Data Matrix Description, Definition and Decision-Rule Logic

The data comparison matrix consists of several anchors that will be used to analyze and compare/contrast the teacher evaluation policies enacted by the states in the study sample. A blank copy of the data matrix follows and is listed as Table 1.

Table 1

States Data Matrix

ANCHORS	GEORGIA	TENNESSEE	NEW YORK	ILLINOIS/ CHICAGO	ALABAMA
POLICY ORIGIN:					
(a) Prompted Need for Teacher Evaluation Policy					
(b) Who initiated?					
(c) Who created?					
POLICY PURPOSE:					
(a) Policy implicit or Explicit					
(b) Purpose Clearly Stated					
(c) Is policy being implemented/pursued in a manner consistent with its stated purpose					
Policy Design					
(a) Is policy mandated or voluntary?					
(b) If voluntary are there any inducements?					
(c) Who coordinates policy					
Policy components, mechanisms and procedures					
(a) Theoretical basis, literature, and causal assumptions					
(b) Evaluative rubric					
(c) Types of Data					
1. How data is collected					
2. Mechanisms for training evaluators					
3. Validity and Reliability					
(g) Formative or Summative					
(h) Legal Challenges/Issues					
(i) Distinctive Features					
5. Policy Effects and Feedback: Pro and Con					

These five primary anchors and subsequent sub-anchors are as follows:

6. Policy Origin

- a. What prompted the need for a teacher evaluation policy?
- b. Who initiated it?
- c. Who was involved in creating the new evaluation policy

7. Policy Purpose

- a. Is the purpose of the policy implicit or explicit?
- b. Is the purpose of the policy clearly stated and logically consistent?
- c. Is the policy being implemented/pursued in a manner consistent with its stated purpose?

8. Policy Design

- a. Mandated for all districts or is voluntary?
- b. If voluntary, are there state inducements/incentives?
- c. Who is responsible for coordinating the policy in the state?

9. Policy Components, Mechanisms and Procedures

- a. Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?
- b. Evaluative rubric
- c. Types of data used to evaluate?
- d. How data is collected? (i.e., by whom? How often? What means?)
- e. Mechanisms for training evaluators? Of what does it consists? How is it done?
- f. Procedures used to ensure validity and reliability of evaluation?

g. How are data generated from teacher evaluation used? For formative or summative purposes? How specifically?

10. Policy Effects and Feedback: Pro and Con – As noted from various sources and by various groups (e.g., evaluation studies, newspaper reports, teachers union reaction, lawsuits, etc.), what positive and/or negative reactions and effects OR problems/successes has the policy produced?

To help better understand the study, a data matrix was designed consisting of five anchors, each with sub-anchors. The first anchor examined the policy origin of the new models with the three sub-anchors that deals with the need for the new policy, initiation of the new policy, and the creation of the policy. This anchor investigated how the Race to the Top Fund played a part in the origin of the new policies and teacher evaluation models in each state studied. The specific origin of each new policy as the policy was enacted was found in this anchor and its sub-anchors. As Darling-Hammond (2013), Popham (2013), and Hull (2013) agreed, the more inclusive the team is for creating new teacher evaluation policies the better the policy will be. These three sub-anchors gave the study a base for examination; as the origin of the new policy push was explained, a further comparison of the policies themselves could be researched.

The second anchor and subsequent sub-anchors looked at the purpose of the policy. The implicitness or explicitness of each policy was sought out to determine if the policy purpose was simply implied or if it was clearly stated and easily found by all stakeholders. The clearness of the policy helped both the evaluator and the evaluatee understand what was being evaluated and why. Finally, the new policies were examined to determine if they were being implemented as they are written in a style that is unfailing according to the purpose of the evaluation?

To understand the makeup of the new teacher evaluation models, the third anchor and sub-anchors examined the policy design of each state. Looking at the policy design, this study investigated whether or not the models were mandated by the stated departments of education or if districts could voluntarily enter into the new models. Does the policy afford districts with incentives or inducements as a means to adopt the new policy and when adopted who will coordinate the implementation of the policy?

The fourth anchor and sub-anchors concentrated on policy components, mechanisms, and procedures for these new state evaluation policies. The first sub-anchor looked at the theoretical basis, the existing literature, and the working causal assumptions upon which the new policies were based. The theoretical basis leads the fourth anchor in the study to find which theories the new policies were based on as well as what new or old mechanisms were being introduced to the field of teacher evaluation. The causal assumptions tried to find the bases for the hope of the outcome of the new policies. What is the causation if new changes are made to the way teachers are evaluated?

The evaluative rubric gave evaluators and teachers a look at what was being evaluated and how much each component to the rubric counted toward the final evaluation numbers. A rubric by definition is a set of rules or conduct that helps to conduct a process.

The third sub-anchor consisted of two parts to include how the data were collected for each teacher evaluated, who collected the data in which companies, how often the data were collected, and by what means were the data being collected. This was followed by the mechanisms for training the evaluators. The study sought to identify the method of training as well as the frequency each state had introduced to train evaluators.

Sub-anchor four dealt with one of the most important components of any evaluation model, its validity and reliability. Validity is defined by Fraenkel and Wallen (2003) as the degree to which correct inferences can be made based on results from an instrument. Fraenkel and Wallen, 2003 also defined reliability as the degree to which scores obtained with an instrument are consistent measures of whatever the instrument measures. Sub-anchor four researched the methods used to determine both validity and reliability of each state's new evaluation model.

A summative evaluation is one that judges the worth of its intended value at the end of the process, and a formative evaluation is an ongoing evaluation throughout the process. This sub-anchor tells the audience whether or not the data generated were used as a summative, formative, or both, evaluation tool.

The final two sub-anchors dealt with legal challenges that may have been leveled against the new evaluation policies or any challenges from outside entities to the policies, and the distinctive features the states might have that were not found in any of the other states being researched.

The final anchor explored policy effects and feedback both, positive and negative.

Decision-Rules-Logic

The decision-rule logic is explained in this section, to give a purpose to how the information in the data matrix was chosen. Decision rules give a synthetic, easily understandable, and generalized representation of the knowledge contained in a data set organized in a matrix. According to Johnson and Kruse (2009), "data driven decisions are decisions directed toward specific goals and informed by data aligned with the realization of these goals" (p. 100). Johnson

and Kruse (2009) stated that this brings logic to decision making and helps make decision making a complete and reliable process from the beginning of the process until the final decisions are made.

The decision-rule logic for the first anchor and sub-anchors was to discover in the literature if there was a causal assumption for the need of a policy change and what encouraged the need. The logic to the origin also lies in the assumption that there was a specific trigger that led to the creation of the policy and if there was a trigger, how prominent a role did the trigger play in the origin of the policy? Was the trigger a federal government initiative or incentive, a state led idea; or was it educator or business directed? So, in looking at the origin of each state's new teacher evaluation policy, the policy was studied from its inception. The initiation of the policy and the creators of the policy were found in the policy itself. The initiation of the new policy and the creation centered on legislation passed in each state studied.

Anchor two, the policy purpose was sought out in each state's teacher evaluation policy. The implicitness or explicitness and the purpose being clearly stated were determined by what was written in each policy. The study specifically looked for statements that began with, "the purpose of this policy."

To determine how the policy was being implemented, research was done to compare the stated policy purpose with the components of each policy and the way they were being applied. Each policy was dissected to understand its clarity or its ambiguity. The purpose should be easily identified within the policy, with clear guides to its meaning and how its purpose is going to be accomplished. Is there a guide for the stakeholder to follow? Does the stakeholder have a hard time locating and determining the purpose of the new policy?

The policy design of each state's evaluation policy was determined to be mandated or voluntary by examining the legislation that created each policy, which also laid out the coordination strategy of each state.

To determine the policy components, mechanisms, and procedures, the enacted legislation was studied, along with the policies themselves and teacher evaluation literature. The theoretical bases was determined by what was written in the policies and the existing literature on teacher evaluation policy, the evaluative rubric, types of data, data collection, mechanisms for training evaluators, validity and reliability, type of evaluation comes from the policy components and procedures. Issues with the policies, distinctive features, the final anchor policy effects and feedback were composed from sources such as newspapers, teacher organizations, think tanks, stakeholder organizations, and other sources with a stake in educational accountability.

Georgia Teacher Evaluation Model: Teacher Keys Effectiveness System

As Georgia's department of education competes for funds provided by the Race to the Top Initiative, they have developed an evaluation system for teachers called the Teacher Keys Effectiveness System (TKES). TKES has several components. TKES looks to optimize student learning and growth along with improving the quality of instruction by ensuring accountability for classroom performance and teacher effectiveness. The model also strives to contribute to the successful achievement of the goals and objectives defined in the vision, mission and goals of Georgia Public Schools. The model also want to provides a basis for instructional improvement through productive teacher performance appraisal and professional growth, implement a performance evaluation system that promotes collaboration between the teacher and evaluator

and promotes self-growth, instructional effectiveness and improvement of overall job performance, with a primary focus on student learning (GADOE, 2012).

TKES consist of three components that will contribute to the Teacher Effectiveness Measure (TEM): teacher assessment on performance standards, student growth and academic achievement, and surveys of instructional practice. The amount each component will count towards the TEM will be determined over time as districts become proficient with the process (GADOE, 2012).

Teacher assessment on performance standards will comprise multiple data sources to include observations, both formal and informal, teacher documentation (i.e., student work and lesson plans), and surveys.

Student growth and academic achievement will be used to determine the student growth percentile (SGP) for those teachers in subjects that are tested and the student learning objectives (SLOs) for teachers whose subjects are not tested. The student growth percentile describes a student's academic growth compared to his or her academic peers. A growth percentile can range from 1 to 99. The higher the percentile scores, the higher the academic growth. The Student Growth Percentile will count as 50% of a teacher's TEM when the TKES is fully implemented. The SGP will be used only for the teachers who are teaching subjects that count toward accountability ("Teacher keys effectiveness," 2013).

Student guidelines have been established to calculate a teacher's TEM score. First, for a student's score to count toward a teacher's TEM score a student must be enrolled at least 65% of the instructional days. There must be a minimum of 15 students in a teacher's class. Georgia uses a Roster Verification Model that attaches students to teachers, and retained students will be judged like students who have not been retained ("Teacher keys effectiveness," 2013).

Georgia's student growth model (GSGM) is designed to track a student's growth in relation to other students statewide with similar prior achievement. The GSGM is used to improve teaching and learning as teachers and stakeholders are provided with another tool to help analyze student performance. The GSGM hopes to be able to answer the following questions: Did this student grow more or less than other students who are performing at the same level? Are students growing faster in math or reading? Are students headed for proficiency ("Georgia student growth," 2012)

In subject areas not being tested using the CRCTs or End-of Course Tests (EOCT), Georgia has decided to use student learning objectives (SLO). The purpose of SLO is to improve academic achievement at the classroom level and to determine a teacher's impact at the classroom level. The student learning objective (SLO) component looks to improve student achievement levels in subjects that are not being tested. The SLO requires teachers to use assessments to determine student growth. Georgia districts must use GADOE approved SLOs that uses district-identified achievement growth measures ("Student learning objectives," 2012).

Student surveys are used based on the student's grade level. Surveys will be given in different versions to Grades 3-5, 6-8, and 9-12 with the readability being validated by the Flesch-Kincaid Readability Levels of surveys. The surveys will reflect items that students have direct knowledge of and can give a quick response and feedback. Once TKES is fully implemented, the surveys will count toward a teacher's TEM (GaDOE, 2012).

The Georgia Department of Education partnered with 26 school systems to implement the first phase of teacher evaluations. These school districts were chosen and given half of Georgia's Race to the Top funds based on their Title I formula. These districts made up 40% of Georgia's student population with a demographic breakdown of 46% poverty, 53% African American, 48%

Hispanic, and 68% of the lowest achieving school districts (Office of School Improvement Teacher and Leader Effectiveness Division, 2013).

Concerns of Teacher Keys Effectiveness System in Georgia

A consortium of Georgia college professors, researchers, and education advocates articulated the concerns of many stakeholders over the implementation of Georgia's new value added model for evaluating teachers. The consortium is referred as GREATER. GREATER expresses several concerns with Georgia's new Teacher Keys and Leader Keys. It is expressed by GREATER that the use of a value-added model (VAM) is not a good instrument for evaluation purposes. GREATER wants more high quality research completed and believes that the value-added model will have an adverse effect on student achievement (Strauss, 2012b).

GREATER identified four concerns of Georgia's new evaluation system. The concerns include the validity of the model, its feasibility, unintended consequences, and timing.

According to Strauss (2012a) the validity concern centers on the notion that VAMs do not give consistent or stable ratings for teachers, scores often vary from class to class, test to test, and year to year. As school systems are searching for ways to improve student achievement, an emphasis is placed on fiscal responsibility. GREATER believes that spending so much money on an untested and unreliable model of evaluation is a waste of money that could be used in different ways. Also, there are unintended consequences that will flow to the students if VAMs are used, such as focusing on the content areas that are being tested and not focusing on the non-tested content areas. The authors foresee civil rights litigation from low-socioeconomic students because of the focus on the test and students not being afforded a well-rounded education as well as teachers avoiding certain students because of preexisting conditions such as English language

learners, special needs students and others perceived to be slow learners, and the possibility of cheating. According to Shearer (2014), Georgia's new evaluation model might become a magnet for lawsuits. The validation process may not hold up to the scrutiny of litigation. Also Shearer (2014) is concerned with the availability of the test results, because normally the results do not arrive to the districts until after teachers are released for the summer. The final concern is that Georgia is not ready to implement a VAM.

GREATER offers two recommendations for their concerns. First is for another pilot study on a small scale and second, reduce or eliminate the percentage of student growth (Strauss, 2012b).

According to Sawchuk (2013b), Georgia teachers are being rated as being at or above average (94%) using TKES. The highest scores are usually given during the classroom observations. This has led Georgia to focus on training evaluators to ensure the reliability of in-school evaluators as to have a more accurate evaluation across the state.

Following the pilot study, lawmakers along with other stakeholders questioned the validity of the statistics. This sentiment was echoed by Dr. Stronge, who stated the number of teachers needing improvement was too low a number (Badertscher, 2013).

Another concern following the pilot study was creating incentives that will hinder education and educators instead of helping. With the proposal to base a teachers pay on students' standardized test scores, for those teachers in non-test subjects using student outcome will pose a problem. The fear is real that teachers may compete for the jobs that look like they will pay the most money and leave the less desirable jobs vacant or understaffed (The Savannah Morning News, 2014).

Georgia school districts are facing a challenge as they are being required to develop new evaluations for teachers that will cost money many districts do not have or want to spend (Morgan, 2014). According to Morgan (2014), districts have to create evaluations that are custom-made for the non-tested subjects that are taking up many man hours and resources. With the advent of all of these requirements as money from the American Recovery and Reinvestment Act begins to dry up, districts will be left to find new revenue streams to complete the reforms (Sharec, 2013).

Georgia: Data Matrix Summary

1. Policy Origin

(a) What prompted the need for a teacher evaluation policy?

According to the Georgia Department of education as part of Georgia's Race to the Top grant application, a new teacher evaluation system was needed. Georgia was awarded \$400 million under the American Recovery and Reinvestment Act of 2009 (ARRA). This act supported new approaches to improving schools in America ("Teacher keys effectiveness," 2013). This prompted the need for a new teacher evaluation policy.

(b) Who initiated it?

The move toward creating a new teacher evaluation system in Georgia was initiated by a partnership among the Governor's Office, the Georgia Department of Education, and the Governor's Office of Student Achievement and education stakeholders ("Teacher keys effectiveness," 2013).

(c) Who was involved in creating the new evaluation policy?

Georgia created the Georgia State Evaluation Steering Committee, which worked along with project consultant, Dr. James K. Strong, Professor College of William and Mary; Virginia Caine Tonneson, Ph.D., College of William and Mary; Xianxuan Xu, Ph.D., College of William and Mary; Leslie W. Grant, Ph.D., Old Dominion University; and Lauri M. Leeper, Ph.D., College of William and Mary (“Teacher keys effectiveness,” 2013).

2. Policy Purpose

(a) Is the Policy implicit or Explicit?

Georgia’ policy is explicit as the visions for the new system is outlined as one to

To equip all Georgia students, through effective teachers and leaders and through creating the right conditions in Georgia’s schools and classrooms, with the knowledge and skills to empower them to 1) graduate from high school, 2) be successful in college and/or professional careers, and 3) be competitive with their peers throughout the United States and the world. (“Teacher keys effectiveness,” 2013, p. 10)

(b) Is the purpose clearly stated?

Georgia’s new evaluation has a very clear purpose as stated in their evaluation handbook.

The handbook identifies several components of its purpose. The first purpose is to optimize student learning and growth; next it looks to improve the quality of instruction by ensuring accountability for classroom performance and teacher effectiveness; contribute to successful achievement of the goals and objectives defined in the vision, mission, and goals of Georgia Public Schools; provide a basis for instructional improvement through productive teacher performance appraisal and professional growth; implement a performance evaluation system that promotes collaboration between the teacher and evaluator and promotes self-growth; instructional effectiveness and improvement of overall job performance; and, finally focus on student learning (“Teacher keys effectiveness,” 2013).

(c) Is the policy being implemented/pursued in a manner consistent with its stated purpose?

According to the Georgia Department of Education (2012), the Teacher Keys Effective System is designed with multiple components that will help with providing information and feedback. This information and feedback will provide all stakeholders with the tools to help teachers improve professionally, which, in turn, should improve student achievement (“Teacher keys effectiveness,” 2013).

3. Policy Design

(a) Is the policy mandated or voluntary?

The Georgia state legislature passed House Bill 244, which mandated the new teacher keys evaluation system for all Georgia public K-12 schools beginning with the 2014-2015 school year (“Teacher keys effectiveness,” 2013).

(b) If voluntary are there any inducements?

Not Voluntary as noted by the passage of House Bill 244.

(c) Who coordinates Policy?

The Teacher Keys Effectiveness system was designed so that the state would have a consistent and comparable evaluation across districts. Each district will coordinate TKES and report its data to the state (“Teacher keys effectiveness,” 2013).

4. Policy Components, Mechanisms and Procedures

(a) Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?

Advanced Research on Teacher Effectiveness and Data Quality has allowed school organizations to identify teacher effectiveness to show what affects student achievement both

positively and negatively. Also, funds from the Federal Government have encouraged states to develop new evaluation policies (“Teacher keys effectiveness,” 2013).

(b) Evaluative rubric

Qualitative (rubrics-based) evaluation will count: 30% for teachers of tested subjects, 60% for teachers of non-tested subjects class-level value-added/growth score will count 50% for teachers of tested subjects. Teachers of non-tested subjects will not receive a value-added/growth score. Student achievement gap reduction will count 10% for teachers of tested subjects. Teachers of non-tested subjects will not receive a student achievement gap reduction score. Other quantitative measures (such as parent and student surveys): 10% for teachers of tested subjects, 40% for teachers of non-tested subjects teacher assessment on performance standards (TAPS) (“Teacher keys effectiveness,” 2013).

(c) Types of data used to evaluate?

Classroom observations, teacher documentation, surveys, CRCT Grades 4-8, and EOCT high school (for student academic growth) (“Teacher keys effectiveness,” 2013).

(d) How data are collected?

Principals, assistant principals, and other school administrators who are responsible for evaluating teachers. All evaluators must be fully trained and credentialed by a state and/or district credentialed trainer in using the components of the Teacher Keys Effectiveness System (TKES). All administrators who are responsible for evaluating teachers must be credentialed prior to using TKES (“Teacher keys effectiveness,” 2013).

(e) Mechanisms for training evaluators?

Evaluators will be trained by partnering Georgia Department of Education specialists and school district staff. Central office personnel who are responsible for evaluating principals will

be trained by Georgia Department of Education specialists. District personnel will provide an orientation to the Leader Assessment on Performance Standards for building principals. Building principals will provide an orientation to the Teacher Assessment on Performance Standards for teachers. Webinars and regional sessions will be scheduled by the Georgia Department of Education to assist with the orientation process for the Teacher Assessment on Performance Standards. Georgia Department of Education specialists will also provide training on the other measures included in the comprehensive evaluation systems during the 2011-2012 pilot study. For the 2012-2013 implementation of the validated Teacher Keys Evaluation System and Leader Keys Evaluation System, all personnel will be retrained.

A minimum of four walkthroughs, two formative observations, two formative assessments and a summative assessment, and a summative conference are required (“Teacher keys effectiveness,” 2013).

(f) Procedures used to ensure validity and reliability of evaluation

Georgia uses a performance appraisal rubric to help provide for increased reliability. The rubric is a behavioral summary scale that will be used to guide evaluators in collecting data on the individual teacher. The rubric provides a qualitative description of performance at each level (“Teacher keys effectiveness,” 2013).

(g) How are data generated from teacher evaluation used? For formative or summative purpose? How?

Formative/Summative approach: provide continual data on teacher and student performance that can be used to inform individual, team, school, and system professional learning plans and support ongoing professional growth (“Teacher keys effectiveness,” 2013).

(h) Legal challenges / issues associated with policy in state

A consortium of Georgia college professors, researchers, and education advocates articulated the concerns of many stakeholders over the implementation of Georgia's new value added model for evaluating teachers. The consortium is referred to as GREATER. GREATER expresses several concerns with Georgia's new Teacher Keys and Leader Keys. It is expressed that the use of a value-added model (VAM) is not a good instrument for evaluation purposes and it has been expressed by many assessment experts. GREATER wants more high quality research completed and believes that the value-added model will have an adverse effect on student achievement. GREATER identifies four concerns of Georgia's new evaluation system.

The validity concern centers on the notion that VAMs do not give consistent or stable ratings for teachers; and scores often vary from class to class, test to test, and year to year. As school systems are searching for ways to improve student achievement, an emphasis is placed on fiscal responsibility. GREATER believes that spending so much money on an untested and unreliable model of evaluation is a waste of money that could be used in different ways. Also, there are unintended consequences that will flow to the students if VAMs are used, such as focusing on the content areas that are being tested and not focusing on the non-tested content areas. The authors foresee civil rights litigation from low-socioeconomic students because of the focus on the test and students not being afforded a well-rounded education as well as teachers avoiding certain students because of preexisting conditions such as English language learners, special needs students, and others perceived to be slow learners, and the possibility of cheating. The final concern is that Georgia is not ready to implement a VAM.

GREATER offers two recommendations for their concerns. First is for another pilot study on a small scale and second, reduce or eliminate the percentage of student growth (Strauss, 2012b).

(i) Distinctive features of teacher evaluation policy found in this state and not in others.

Georgia uses student growth percentile (SGP) for those teachers in subjects that are tested and the student learning objectives (SLOs) for teachers whose subjects are not tested

5. Policy Effects and Feedback: Pro and Con--As noted from various sources and by various group (e.g., evaluation studies, newspaper reports, teachers' union reaction, lawsuits, etc.), what positive and/or negative reactions and effects OR problems/successes has the policy produced?

Georgia education leaders are in a quandary about the new piloted teacher evaluation policy. The pilot showed only 1% of the 5,800 teacher evaluated as being ineffective. The pilot data did not include student progress, as Georgia is still working on analyzing that component. Professor Stronge, who has been a consultant for Georgia, believes the number of teachers needing assistance was too low in the pilot study. As teachers wade through the new evaluation policy, Tennessee, like New York, is burdened by the implementation of the new Common Core standards along with new evaluation procedures (Badertscher, 2013). However, one positive to Georgia's new teacher evaluation is the inclusion of advice from teachers and administrators and it has become more flexible (Downey, 2014).

Tennessee Teacher Evaluation Model: Tennessee Educator Acceleration Model

The state of Tennessee embarked on a journey to revamp its education policies in an effort to compete for funds under the federal government's Race to the Top competition initiative. Tennessee became the first state to implement a comprehensive student outcomes-based, statewide evaluation model. The Tennessee State General Assembly passed the First to the Top Act. The First to the Top Act created the 15-member Teacher Evaluation Advisory

Committee (TEAC) to collaborate and recommend guidelines for a Tennessee new teacher evaluation model. The new system requires 50% of the evaluation be based on student achievement data, with 35% based on student growth as represented by the Tennessee Value-Added Assessment System (TVASS) or any comparable growth model, and the remaining 15% would be based on other forms of student achievement adopted by the State Board of Education. The remaining 50% would be gathered from teacher observations, conferences, and review of prior evaluations and work. Tennessee believes, with multiple observations that in turn lead to meaningful feedback, coupled with the use of student data and professional development that is significant to the individual teacher, the new policies that are being implemented will provide student and teacher growth (TDOE, 2012):

Through a combination of frequent observation, constructive feedback, student data and meaningful professional development, the new system is designed to support all educators so they can do their best work in the classroom and help every student learn and grow. (TDOE, 2012, p. 1)

The Tennessee Educator Acceleration Model (TEAM) has been approved as the evaluation model for the majority of the state of Tennessee. Three additional models were approved also: Project Coach for Hamilton County Schools, Teacher Effectiveness Measure for Memphis City Schools, and the Teacher Instructional Growth for Effectiveness and Results for 12 municipal school systems (TDOE, 2012). TEAM is designed to offer teacher accountability through frequent observations, constructive feedback, student growth, and professional development. TEAM looks to give teachers recurrent and productive feedback by having at least four classroom observations for all teachers. Tennessee's evaluation systems also will help identify and develop strengths and weaknesses a teacher might have and design professional development for support and use effective educators to help both novice and ineffective

educators. TEAM will use both qualitative and quantitative data to determine a teacher's effectiveness score (TDOE, 2012).

The qualitative portion of TEAM will consist of a teacher's performance inside the classroom, while the quantitative portion will be based on student data. The 35% student growth portion will be measured using the Tennessee Value-Added Assessment System (TVAAS). The TVAAS measures a student's actual growth to their projected growth. Teachers teaching subjects that are tested will have individual TVAAS data, but teachers who teach subjects that are not tested will use school-wide TVAAS data. The student growth model that will determine the remaining 15% of the quantitative section of TEAM may consist of one of the following approved measurements that are closely aligned with the educators' field: state assessments that are discipline-specific, Pre-K-12 diagnostic or achievement commonly used throughout the state, postsecondary matriculation, participation in advanced coursework, and ninth grade promotion and retention rate.

The qualitative components of TEAM will primarily come from teacher observations. Teachers who hold a state teaching certificate will be observed four times per year: two 15-minute observations and two lesson length observations. Non-tenured teachers will have a total of six observations per year, with three-minute observations and three lesson length observations. Teachers will be observed by principals, assistant principals, or other instructional leaders. The evaluators will focus on four areas during the observation: planning, instruction, environment, and professionalism based on the TEAM rubric. Evaluators will be trained using expert trainers during a 4-day training session with the evaluators being required to pass an end-of-course test for certification. The overall evaluation will be calculated based on the three-part process of 35% student growth, 15% student achievement, and 50% observations. The three parts will be

combined to get a final score, which will be scored using the effectiveness scale: 1 = The lowest level of significantly below expectations, 2 = below expectations, 3 = meeting expectations, 4 = teaching above expectations, and 5 = the highest level or significantly above expectations. The scores of TEAM will be used for personnel decisions such as tenure attainment, professional development, and dismissal (TDOE, 2012).

Concerns with TEAM

According to the Tennessee Department of Education, the TVAAS provides a reliable, objective measure of a school's influence on student growth (TDOE, 2012). The department also stresses the benefits that TVAAS will provide teachers, such as, monitoring student growth, modifying instruction for all students, developing professional development, and learning practices that work from teachers who have received high effectiveness scores (TDOE, 2012). However, as teachers and principals implement TEAM, there has been much discussion as to the merit of the new model. Teachers in subjects that are not being used for accountability testing are required to be judged on a school-wide student growth model. These teachers are having 50% of their evaluation based on students they may or may not have taught. SCORE has taken this into consideration, making a recommendation to the Tennessee Department of education to reduce the percentage from 50 to 25% (McCoy, 2012). According to Schelzig (2012), other concerns for TEAM include improving training for evaluators, linking evaluation results with training opportunities for teachers, accelerating a system for evaluating principals, redoubling efforts in districts where teacher evaluation systems have faltered in the first year, and integrating the evaluation system with the new common core state standard.

Sawchuk (2013b) stated that many of Tennessee's teachers are receiving the same or better rating as before TEAM was introduced. His research found that 98% of Tennessee teachers were rated at expectations or above by their principal. Also the data shows that TEAM's observations of teachers rated only 0.2% of the teachers as below expectations, while the quantitative portion of TEAM rated 16.5% in the below expectations level.

Tennessee: Data Matrix Summary

1. Policy Origin

(a) What prompted the need for a teacher evaluation policy?

In an effort to receive money from the federal government under the Race to the Top fund, Tennessee chose to complete an application for the competitive grant with one of the requirements being to develop a new teacher evaluation model using multiple measures to determine teacher effectiveness ("Tennessee First to the Top," 2013).

(b) Who initiated it?

The Tennessee state legislature passed the Tennessee First to the Top Act of 2010, which paved the way for their RTTT application submission ("Tennessee First to the Top," 2013).

(c) Who was involved in creating the new evaluation policy?

The Tennessee First to the Top Act created the Teacher Evaluation Advisory Committee to recommend guidelines and criteria for the new evaluation system. Committee members include teachers, principals, legislators, business leaders, and Tennessee Department of Education officials ("Tennessee first to the Top," 2013).

2. Policy Purpose

(a) Is the Policy implicit or Explicit?

Tennessee's policy is explicit as the Tennessee First to the Top Act changed the law concerning teacher evaluations in the state. The new Act created annual evaluations for teachers, a new teacher and principal evaluation framework that requires 50% of evaluations be based on student achievement measures, including 35% of TVAAS where available, and created a 15-member Teacher Evaluation Advisory Committee to recommend guidelines and criteria for the new evaluation

(b) Is the purpose clearly stated?

The purpose is clearly stated as quoted from the Tennessee department of education: Tennessee can only successfully rise to the top, preparing every child for college and career in the 21st century, by significantly improving student achievement. First to the Top will give the State and districts resources to provide more support and development opportunities for teachers and principals to ensure everyone has the opportunity to grow and succeed. (TODE, 2012 p. 1)

(c) Is the policy being implemented/pursued in a manner consistent with its stated purpose?

In an effort to implement the new evaluation policy in a way consistent with its purpose, Tennessee will evaluate teachers by having four observations per year for professional teachers and six observations for apprentice teachers followed by feedback conferences: required domains for the 50% qualitative instrument, possible options for the 15% student achievement component, and a process for identifying and/or developing alternative growth measures for educators in non-tested subjects and grades ("Tennessee first to the top," 2013).

3. Policy Design

(a) Is the policy mandated or voluntary?

Tennessee's new evaluation policy is mandated by the First to the Top Act ("Tennessee first to the top," 2013).

(b) If voluntary are there any inducements?

Mandated

(c) Who coordinates Policy?

The Tennessee Department of Education will coordinate the new policy with each school district implementing the policy. Scores will be reported to the TDOE (“Tennessee first to the top,” 2013).

4. Policy Components, Mechanisms, and Procedures

(a) Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?

Tennessee’s First to the Top plan outlined a need for a new teacher evaluation policy. The new policy will help teachers identify weaknesses and strengths and help in ways to improve a teacher’s ability to improve student achievement.

(b) Evaluative rubric

Fifty percent will come from student achievement (35% student growth measure and 15% other student achievement). The other 50% will come from other criteria (classroom observations, review of prior evaluations, and other components). TEAM rubric. The TEAM Rubric is the result of a successful collaboration with the National Institute for Excellence in Teaching (NIET). TDOE chose to partner with NIET to develop the 50% qualitative component for TEAM based on several factors. Those factors were positive field test results, research that links the instrument to increases in student achievement, capacity of NIET to provide expert trainers for high-quality direct training and certification of all observers and evaluators statewide, and the availability of the NIET Best Practices Portal to provide immediate access to training and development resources for teachers (“Tennessee first to the Top,” 2013).

(c) Types of data used to evaluate?

TVAAS: 35% student growth measure and 15% other student achievement and classroom observations, review of prior evaluations, and other components. The other 50% will be comprised of classroom observations, teacher conferences, prior evaluations and portfolios.

(d) How data are collected? (i.e., By whom? How often? What means?)

Principals, assistant principals, or other instructional leaders trained in the observation protocol. Observers use the TEAM Rubric, which focuses on four areas: instruction, planning, environment, and professionalism observations will be conducted twice per year for licensed teachers

(e) Mechanisms for training evaluators? Of what does it consist? How is it done?

Principals, assistant principals, and other instructional leaders are eligible to serve as certified observers. All designated observers must participate in required certification training (3 days) and demonstrate proficiency in the NIET observation process by successfully completing an online, annual certification test to be certified.

(f) Procedures used to ensure validity and reliability of evaluation?

Principals and assistant principals and other instructional leaders will be trained and must pass a certification test on the observation rubric before they can conduct evaluations. The State will provide training to ensure consistent scoring and calculations and the state will analyze the evaluation procedures and results each year and make changes as needed.

(g) How are data generated from teacher evaluation used? For formative or summative purposes? How specifically?

Both formative and summative. The First to the Top Act requires that evaluations play a factor in personnel decisions to include dismissal, granting of tenure, and professional development (“Tennessee first to the top,” 2013).

(h) Legal Challenges/Issues Associated with Policy in State

According to the Tennessee Department of Education, the TVAAS provides a reliable, objective measure of a school’s influence on student growth (TDOE, 2012). The department also stresses the benefits that TVAAS will provide teachers such as, monitoring student growth, modifying instruction for all students, developing professional development, and learning practices that work from teachers who have received high effectiveness scores (TDOE, 2012). However, as teachers and principals implement TEAM, there has been much discussion as to the merit of the new model. Teachers in subjects that are not being used for accountability testing are required to be judged on a school-wide student growth model. These teachers are having 50% of their evaluation based on students they may or may not have taught. SCORE has taken this into consideration, making a recommendation to the Tennessee Department of education to reduce the percentage from 50 to 25% (McCoy, 2012). According to Schelzig (2012), other concerns for TEAM are a need to improve training for evaluators, linking evaluation results with training opportunities for teachers, accelerating a system for evaluating principals, redoubling efforts in districts where teacher evaluation systems have faltered in the first year, and integrating the evaluation system with the new common core state standards.

(i) Distinctive Features of Teacher Evaluation Policy found in this state and not the others.

Fifty percent of the evaluation is to be based on student achievement data

5. Policy Effects and Feedback: Pro and Con

As noted from various sources and by various group (e.g., evaluation studies, newspaper reports, teachers union reaction, lawsuits, etc.), what positive and/or negative reactions and effects OR problems/successes has the policy produced?

A positive to Tennessee's evaluation policy is that it has been fully implemented in the state. However, the preliminary results show teacher observations scores are still being scored high by onsite evaluators (Sawchuk, 2012).

New York Teacher Evaluation Model: Annual Professional Performance Review

The New York State Assembly and Senate passed the teacher and principal evaluation law proposed by the Governor on March 14, 2012, and this legislation was signed into law by the Governor on March 27, 2012. Chapter 21 of the Laws of 2012 amended Education Law 3012 in an effort to change the nature of teacher evaluation in the state of New York. New York stated it wanted a highly effective teacher in every classroom and to promote a culture of continuous professional development for all teachers. Teachers will be given one of four ratings during the evaluation process: highly effective, effective, developing, and ineffective (NYSED, 2012).

The evaluation will give teachers an Annual Professional Performance Review (APPR), which will be given in one composite score using multiple measures of teacher effectiveness. The composite score will reflect in a systems determination for promotions, retention, tenure, termination, professional development, and supplemental compensation (NYDOE, 2012).

New York's teacher evaluation model requires 40% of a teacher's rating come from student achievement. Twenty-five percent will come from a value-added model for the subjects tested and 15% will come from system-level measures that are deemed to be rigorous and comparable across classrooms. The remaining 60% will come from multiple measures of teacher

effectiveness as prescribed by the Commissioner of Education. Teacher observations are allowed to make up the 60% of the multiple measures portion of the evaluation.

Teachers who do not teach subjects that have state assessment given must rely on Student Learning Objectives (SLO) for their value-added score. New York State defines SLO as an academic goal for students that is determined at the beginning of the school year that represents what a student should learn during the school year. The objective must be measurable and specific based on prior student data and being aligned with Common Core, State, or National standards. New York's SLO should include student population, learning content, interval of instructional time, evidence, a baseline, HEDI criteria, and a rationale for the objective (NYSED, 2012).

The instrument used to determine New York's teacher value-added score was created by economists at the Value-Added Research Center, a nonprofit housed at the University of Wisconsin-Madison's School of Education. The model will use a design to control for variables like past performance, poverty, and class size. Its results allow school administrators to compare the effectiveness of each teacher against his or her peers (Rutkoff, 2012).

Concerns with New York Teacher Evaluation Model

Concerns with the Annual Professional Performance Review of teachers in New York are similar to those of other states who are introducing VAMs to determine a teacher's effectiveness in schools. A group of New York state high school, middle school, and elementary principals provided a list of concerns. The concerns began with educational research and researchers strongly cautioning against teacher evaluation like New York State's APPR Legislation. Stating that students will be adversely affected by New York State's APPR, and that tax dollars are

being redirected from schools to Testing Companies, Trainer and Outside Venders (“An open letter,” 2013).

Taking a look at each concern, we will try and articulate the negative feels held by educators in New York State. The research provided by the principals suggests VAMs offer an unstable rating for teachers with some ratings changing from year to year and class to class. They also offer evidence that student test scores used for evaluation do not produce student achievement gains (“An open letter,” 2013).

The idea of VAMs negatively affecting students is another piece of evidence offered. Principals believe that with the use of VAMs you will have a narrowing of the curriculum and non-tested subject areas will begin to diminish in importance. Students may be placed in lower level classes that have no standardized assessments, and the encouragement of students to take higher level course will also diminish. Finally it is suggested that teachers will not want to teach classes with lower level students and students with disabilities, and collaboration among teachers will turn into competition among teachers.

The training of evaluators will cost in the millions of dollars for the first round of training and continue to increase as more and more supervisor need training. Also, as the demand for speedy test results increase, this will incur another outside cost that could have been used for students.

The principals offered three recommendations they believe will help the process of ensuring all classrooms have an effective teacher, but that are fair to both educators and policymakers. The first recommendations is to have school-wide achievement results used as a part of every teacher’s evaluation, next they recommend piloting and adjusting the system before

using it statewide, and, finally, use four broadbands (ineffective, developing, effective, and highly effective) for the evaluations of teachers (“An open letter,” 2013).

Other concerns in New York City centered on the administration of Charlotte Danielson’s Framework for excellence and the development of measures for all content areas. In 2011, the New York teacher’s union filed suit against the state claiming the new evaluation model weighed too heavily on student test data and this disagreement led to the loss of nearly \$60 million in school improvement grant money. However an agreement was reached in 2012 between the union and the state (Sawchuk, 2013c).

New York: Data Matrix Summary

1. Policy Origin

(a) What prompted the need for a teacher evaluation policy?

The U.S. Department of Education announced that New York State had been awarded \$696,646,000 as a winner in the second round of the federal Race to the Top competition. A component for receiving funds requires states to create new teacher evaluation policies.

(b) Who initiated it?

Governor Cuomo proposed budget legislation that would amend Education Law Section 3012-c to resolve the issues in *New York State United Teachers, et. al v. Board of Regents*, the litigation challenging these regulations. The Department will be proposing amendments to Subpart 30-2 of the Regulations of the Commissioner to conform to the statutory changes and reflect the agreement to settle the litigation. We anticipate that amendments will be made that will substantially impact the conduct of Annual Professional Performance Reviews for the 2012-2013 school year.

(c) Who was involved in creating the new evaluation policy?

Development and adoption of the plan:

1. Except as otherwise provided in subparagraph (o)(1)(ii) of this subdivision, by September 1, 2011, the governing body of each school district and BOCES shall adopt a plan, which may be an annual or multi-year plan, for the annual professional performance review of its teachers providing instructional services or pupil personnel services, as defined in section 80-1.1 of this Title, that meets the content requirements prescribed in clause (b) of this subparagraph.

2. Each superintendent and, in the case of the City School District of the City of New York, the chancellor, in collaboration with teachers, pupil personnel professionals, administrators, and parents selected by the superintendent or in the case of the City School District of New York, the chancellor, with the advice of their respective peers, shall develop the professional performance review plan, which shall be approved by the governing body of each school district or BOCES, filed in the district or BOCES office, as applicable, and available for review by any individual no later than September 10th of each year. The governing body of each school district and BOCES shall provide organizations representing parents and the recognized representative of the teachers' bargaining unit with an opportunity to comment on such plan prior to its adoption.

2. Policy Purpose

(a) Is the Policy implicit or explicit?

The New York State Board of Regents has committed to the transformation of the preparation, support, and evaluation of all teachers and school leaders in New York State. Chapter 21 of the Laws of 2012 amended Education Law §3012-c to fundamentally change the way teachers and principals are evaluated.

(b) Is the purpose clearly stated?

The purpose of the evaluation system is to ensure that there is an effective teacher in every classroom and an effective leader in every school. The evaluation system will also foster a culture of continuous professional growth for educators to grow and improve their instructional practices.

(c) Is the policy being implemented/pursued in a manner consistent with its stated purpose?

The policy is being implemented in a manner consistent with its stated purpose; however, according to Stern (2013), many districts have concerns in several areas of the new policy. The value added component is causing many to question its merit, with the first round of scores not giving teachers the correct credit for the value they have added (Stern, 2013).

3. Policy Design

(a) Is the policy mandated or voluntary?

New York has mandated that the new teacher evaluation policy be followed by all teachers in the state of New York. "The New York State Board of Regents has committed to the transformation of the preparation, support, and evaluation of all teachers and school leaders in New York State" ("Guidance on new," 2013, p. 6).

(b) If voluntary, are there any inducements?

Mandated for all New York teachers

(c) Who coordinates Policy?

The policy is coordinated by the State Department of Education, and all districts have to submit a complete APPR using a State-prescribed form, for commissioner approval. Each district will implement the plan ("Guidance on new," 2013).

4. Policy Components, Mechanisms and Procedures

(a) Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?

New York's new teacher evaluation policy is research based. The policy is based on standards with data based on student progress to include standardized tests. The policy calls for data to be gathered from multiple sources and support for teachers being ongoing ("Learn more about," 2012).

(b) Evaluative rubric

Districts can use one of several rubrics to include CLASS (Danielson, Marzano, Marshall, etc.), 40% will be determined by student achievement. Beginning in 2013, 25% will be based on student growth on State assessments or comparable measures, and 15% on other locally-selected measures that are rigorous and comparable across classrooms, in accordance with standards prescribed by the Commissioner. "Other Measures" (60%), which include observations and evaluations of professional practice by either the supervising administrator, or, if the teacher so chooses, by the administrator (31 points) and a Lead Teacher/Peer Evaluator (up to 29 points). Teachers also have the option to select the revised PART (Performance Appraisal Review for Teachers) for up to 29 points in lieu of the Lead Teacher/Peer Evaluator.

APPR requires that at least 31 points of the "Other Measures" subcomponent shall be conducted by trained administrators, using the TEACHSCAPE Rubric, and based upon multiple observations, at least one of which shall be unannounced. Two formal observations are required ("Teacher and principal," 2013).

(c) Types of data used to evaluate?

The type of data used in the APPR will be based on 40% student achievement, which will include subjects and grades where there is an approved “value-added” model by the Board of Regents for such subject and grade: 25% on student growth on State assessments or comparable measures, and 15% on other locally-selected measures that are rigorous and comparable across classrooms, in accordance with standards prescribed by the Commissioner. The remaining 60% will come from multiple measures of teacher/principal effectiveness consistent with standards prescribed by the Commissioner in regulation. This will include the extent to which the educator demonstrates proficiency in meeting New York State’s teaching or leadership standards (“Guidance on new,” 2013).

(d) How are data collected? (i.e., by whom? How often? What means?)

Data will be collected by trained evaluators. The student growth data will be based on state assessments, and in grades or subjects that do not have a state assessments student learning objectives will be used. Multiple classroom observations (two or more) will be collected by the trained evaluator (“Guidance on new,” 2013).

(e) Mechanisms for training evaluators? Of what does it consists? How is it done?

Section 30-2.9 of the Rules of the Board of Regents provides that in order to be certified as lead evaluators, administrators must be trained in nine elements. The first element is the NYS Teaching Standards, and their related elements and performance indicators or ISLLC standards and their related functions, followed by evidence-based observation techniques grounded in research, along with the application and use of the student growth percentile model and the value-added growth model, and the application and use of approved teacher or principal practice rubric(s) selected by the district or BOCES for use in evaluations, including training on the effective application of such rubrics to observe a teacher’s or principal’s practice. The

application and use of any assessment tools that the school district or BOCES utilizes to evaluate its classroom teachers or building principals including, but not limited to, structured portfolio reviews; student, parent, teacher, and/or community surveys; professional growth goals and school improvement goals, etc.; application and use of any State-approved locally-selected measures of student achievement used by the school district or BOCES to evaluate its teachers or principals; use of the Statewide Instructional Reporting System; scoring methodology utilized by the Department and/or the district or BOCES to evaluate a teacher or principal under this Subpart, including how scores are generated for each subcomponent and the composite effectiveness score and application and use of the scoring ranges prescribed by the Commissioner for the four designated rating categories used for the teacher's or principals' overall rating and their subcomponent ratings; and specific considerations in evaluating teachers and principals of English language learners and students with disabilities.

(f) Procedures used to ensure validity and reliability of evaluation?

All evaluation plans must be approved under the APPR standards, with all evaluators trained to ensure a consistent rating procedure and all evaluation models must follow one of the research-based rubrics outline by the state ("Teacher and principal," 2013).

(g) How are data generated from teacher evaluations used? For formative or summative purposes? How specifically?

The policy calls for a summative evaluation model. Evaluators are required to complete summative evaluation reports, calculate SLOs and LATs, and conduct end-of-the-year meetings with teachers ("What have we," 2013).

(h) Legal Challenges/Issues Associated with Policy in State.

Concerns with the Annual Professional Performance Review of teachers in New York are similar to those of other states who are introducing VAMs to determine a teacher's effectiveness in schools. A group of New York state high school, middle school, and elementary principals provides a list of concerns. The number one concern is that educational research and researchers strongly caution against teacher evaluations like New York State's APPR Legislation. Students will be adversely affected by New York State's APPR, and tax dollars are being redirected from schools to testing companies, trainers, and outside vendors.

Taking a look at each concern we will try and articulate the negative feelings held by educators in New York State. The research provided by the principals suggests VAMs offer an unstable rating for teachers with some ratings changing from year to year and class to class. They also offer evidence that student test scores used for evaluation do not produce student achievement gains.

The idea of VAMs negatively affecting students is another piece of evidence offered. Principals believe with the use of VAMs there will be a narrowing of the curriculum and non-tested subject areas will begin to diminish in importance. Students may be placed in lower level classes that have no standardized assessments and the encouragement of students to take higher level course will also diminish. Finally, it is suggested that teachers will not want to teach classes with lower level students and students with disabilities, and collaboration among teachers will turn into competition among teachers

The cost of training evaluators will be in the millions of dollars for the first round of training and will continue to increase as more and more supervisor need training. Also, as the demand for speedy test results increase, this will incur another outside cost that could have been used for students.

The principals offer three recommendations they believe will help the process of ensuring all classrooms have an effective teacher that is fair to both educators and policymakers. The first recommendation is to have school-wide achievement results used as a part of every teacher's evaluation, next they recommend piloting and adjusting the system before using it statewide, and, finally, use four broad bands (ineffective, developing, effective, and highly effective) for the evaluations of teachers.

(i) Distinctive Features of Teacher Evaluation Policy found in this state and not the others.

New York and Tennessee have unique features to their new evaluation models. They both use their evaluation models for teacher compensation, granting tenure and promoting teachers (Hull, 2013).

5. Policy Effects and Feedback: Pro and Con--As noted from various sources and by various group (e.g., evaluation studies, newspaper reports, teachers union reaction, lawsuits, etc.), what positive and/or negative reactions and effects OR problems/successes has the policy produced?

In New York City, teachers are calling for a moratorium on the new evaluation system, stating it relies too heavily on standardized tests. The teachers union argues that using high stakes test is not an effective way to evaluate teachers ("Teachers call moratorium," 2013). Rochester, New York's teachers' association head and the superintendent have differing views on how well APPR performed. The majority of teachers were rated either effective or needs improvement. The union feels that APPR is ineffective and the superintendent views APPR as a great tool to ensure teacher accountability (Macaluso, 2013).

Illinois Teacher Evaluation Model/Chicago Teacher Evaluation Model:
Performance Evaluation Reform Act

Illinois Governor Pat Quinn gave his blessings to the Performance Evaluation Reform Act (PERA) by signing it into law in 2010. This act required all Illinois school districts to revamp how they determined teacher effectiveness in the classroom (ISBE, 2012). PERA requires school districts to design and implement teacher evaluation models that will judge teachers' classroom performance that is valid and reliable and will improve student learning. Professional excellence will be clearly defined in order that everyone will know what is expected of each classroom teacher. The new evaluation models must include student growth and in the 2012-2013 school year; all teachers will be graded and their performance will be labeled in one of four categories. These categories are excellent, proficient, needs improvement, and unsatisfactory (ISBE, 2012).

A new evaluation system was needed because the one used prior was considered just a checklist of items not related to student achievement or student progress. As time passed, this system was criticized by both evaluators and evaluatees. Not only was the desire for a new evaluation system needed for teacher growth, but with the implementation of the Obama administration's education competition Race to the Top, which tagged additional money for states that created and applied new teacher evaluation systems, school organizations now had an incentive to create new systems (Vevea, 2013).

The new evaluation models were to place emphasis on both teacher practice and student growth by focusing on consistent standards with clearer and more objective feedback, improved professional development, multiple measures of student growth, and improved student learning (ISBE, 2012).

Tenured teachers will be evaluated once every two years unless the teacher is rated needs improvement or unsatisfactory, then they will follow the model for non-tenured teachers, which will have them evaluated every year. Evidence of professional practice shall be collected through the use of multiple observations that include both formal and informal observations.

Classroom observations will be conducted to witness professional practices. The minimum required time for an observation will be 45 minutes in which the evaluator will judge a teacher's planning session, classroom management, and instruction. Non-tenured teachers are required to have three observations per year with at least two of the observations being formal. Tenured teachers who receive excellent or proficient need only have two observations on their next evaluation cycle. However, tenured teachers who are deemed unsatisfactory or needs improvement will be observed three times the following school year (ISBE, 2012).

The student growth portion of the Illinois teacher evaluation model will focus on three types of assessments. The state of Illinois labels these assessments as Type I, Type II, and Type III. Type I assessments are those that are created and graded by an outside entity that measures like students against each other. Type II are those assessments that are developed or adopted and approved by the school district and used on a district-wide basis that is given by all teachers in a given grade or subject area. Type III assessments are those that are rigorous and aligned with the subject's curriculum and that have been determined to measure a student's learning (ISBE, 2012).

PERA requires that the PEAC recommend, and ISBE adopt, a "model" teacher evaluation. By law, the "model" will have student growth comprising 50% of the overall performance evaluation rating (ISBE, 2012). As Illinois and its school district try and find a compromise and a valid teacher evaluation model that will eventually adopt the 50% requirement

by PERA, negotiations and concerns have come about between the teachers' union and policymakers.

Chicago Teacher Evaluation Model

As policymakers push for revised and new teacher evaluation methods, the Chicago teachers union is pushing back. Chicago school leaders proposed a new evaluation method that would require 25% of a teacher's evaluation to come from student performance on standardized tests (Omer, 2012). The Chicago Teachers' Union (CTU) predicts that over 6,000 teachers will lose their jobs under the new teacher evaluation policy because the CTU states 10 to 20% of Chicago students did not meet state standards in math and reading due to their neighborhood and home environments. Along with the CTU, college professors and researcher have come out with concerns with the new teacher evaluation models being implemented. There are three concerns the researcher have listed in a letter to Chicago Mayor Rahm Emanuel, Chicago Public Schools CEO Jean-Claude Brizard, and the Chicago School Board. The concerns are that CPS is not ready to implement a teacher evaluation system that is based on significant use of "student growth," that educational research and researchers strongly caution against teacher-evaluation approaches that use Value-Added Models (VAMs), and, finally, that students will be adversely affected by the implementation of this new teacher evaluation system.

Concerns with Performance Evaluation Reform Act

The concerns are straight forward with the researchers finding that the use of assessments that will be used to determine student growth (Type I, Type II) have not been identified and when identified they must be validated and deemed reliable to assess what needs assessing. The

assessment scores will have to be controlled for socioeconomic status of students, special needs determination of students, and other variables that may affect a student's ability to score at the rate of his/her peers. The use of VAMs as an indicator of teacher effectiveness does not work because of the instability of the models; the evidence does not show that the use of VAMs increase student achievement and the assessments used are not right to judge teacher effectiveness. Finally, the students will suffer if these models are used to determine teacher effectiveness. The researchers believes that teachers will teach the test, try and avoid what they may considers low performing students, and this model will create tension among teachers instead of a culture of collegiality (Strauss, 2012a).

The researchers have two recommendations for the implementation of a new teacher evaluation model. Pilot and adjust the evaluation system before implementing it on a large scale, and minimize the percentage that student growth counts in the teacher evaluation.

According to the researchers, they in no way want to allow ineffective teachers to remain in the classroom. They are for accountability and high standards, but their focus is on designing an evaluation model that will help increase the achievement of students (Strauss, 2012).

In 1987, President Reagan's Secretary of Education William Bennett called Chicago Public Schools the worst public schools in America. With this damnation, the Illinois legislature gave the city of Chicago unprecedented control of its public schools. The Chicago School Reform Act, which created local school councils, was passed in 1988. These new councils were made up of teachers, parents, community members, and principals and they were given powers to manage budgets, set standards, design curriculum, and hire and fire principals ("Chicago Magazine," 2012).

Chicago's new teacher evaluation model is known as **Recognizing Educators Advancing Chicago** (REACH). REACH was Chicago's first attempt at revamping their teacher evaluation policies in 40 years. As mandated by PERA, Chicago looks to create a system that will help educators grow professionally and to provide educators with tools needed to increase student achievement (CPS, 2012).

REACH students has six goals and three components of multiple measures. The goals are to establish a common definition and standards for teaching excellence, build principals' and/or teachers/leaders' expertise in observing and analyzing instruction to support teacher growth, provide teachers with information and guidance to inform their development, engage teachers in reflection and self-assessment regarding their own performance, differentiate support and accountability for teachers based on their experience and/or impact on student learning, and create a culture of continuous improvement among teachers, school leaders, system administrators and students.

The three multiple measure components are teacher practice, student growth, and student feedback.

REACH students was developed from suggestions from thousands of teachers who wanted an evaluation that would help them grow professionally. Also, Chicago Public Schools along with the Chicago Teachers Union spent countless hours negotiating to ensure REACH would be meaningful to both teachers and students (CPS, 2012).

The teacher practice component is based on CPS' Framework for Teaching, which is based on Charlotte Danielson's Framework for Teaching. The student growth component has two methods of measurement: standardized assessments and teacher designed performance tasks.

Finally, student surveys will be used to gauge teacher effectiveness beginning in the 2014-2015 school year (CPS, 2012).

The first year of implementation showed positive results as stated by non-tenured teachers. According to Sawchuk (2013), REACH evaluated 4,000 teachers with less than 3 years of teaching experience. CPS assigns teachers to one of four categories based on their evaluation: excellent, proficient, developing, and unsatisfactory. Of the teachers evaluated, 48% were evaluated as being proficient, 40% of teachers were in the developing phase, and only 2% of non-tenured teachers were viewed as ineffective (Vevea, 2013).

However, not all involved were as optimistic about the new system. The head of the Chicago Teachers Union (CTU) stated test data was weighted too much with student data counting 25% the first year and topping out at 30% of a teacher's evaluation percentage. Also, CTU does not believe the evaluators are properly trained to help advise teachers when asked about specific practices of teaching (Sawchuk, 2013). In an editorial by the Chicago Sun Times (2013), the new evaluation system places an extra burden on school administrators tasked with conducting the evaluations. Administrators were taking up to 3 weeks to complete observations for non-tenured teachers being evaluated, and it was estimated an additional 2 to 3 weeks when tenured teachers are added to the process.

Illinois/Chicago: Data Matrix Summary

1. Policy Origin

- a. What prompted the need for a teacher evaluation policy?
- b. Who initiated it?
- c. Who was involved in creating the new evaluation policy?

(a) What prompted the need for a teacher evaluation policy?

The state of Illinois received \$42.8 million in the third round of the RTTT competition. As part of this award, Illinois would have to create a new teacher evaluation policy (ISBE, 2012). A new evaluation system was needed because the one used prior was considered just a checklist of items not related to student achievement or student progress. As time passed, this system was criticized by both evaluators and evaluatees (Vevea, 2013).

(b) Who initiated it?

In 2010, Gov. Pat Quinn signed the Performance Evaluation Reform Act (PERA), which requires all schools in Illinois to change how teachers' and principals' performance is measured. PERA requires districts to design and implement performance evaluation systems that assess teachers' and principals' professional skills as well as incorporate measures of student growth.

(c) Who was involved in creating the new evaluation policy?

PERA established the Performance Evaluation Advisory Council (PEAC) comprised of teachers, principals, superintendents, and other interested stakeholders to advise ISBE on the development and implementation of improved performance evaluation systems and supports (ISBE, 2012).

2. Policy Purpose

- a. Is the Policy implicit or Explicit?
- b. Is the purpose clearly stated?
- c. Is the policy being implemented/pursued in a manner consistent with its stated purpose?

(a) Is the Policy implicit or Explicit?

The policy is explicit: Teacher evaluation systems will provide clear descriptions of professional excellence so everyone understands what great teaching means. The evaluations will be based on standards of effective practice, with evaluators trained and pre-qualified to conduct observations, collect evidence, and provide helpful, timely feedback. The new evaluations will add objectivity to a practice that almost universally was subjective (ISBE, 2012).

(b) Is the purpose clearly stated?

The purpose of Illinois' policy was to provide teachers with the support needed to improve the achievement of students.

Section 50.10: Purpose

This part establishes the minimum requirements for the establishment of valid and reliable performance evaluation systems for certified employees, pursuant to Article 24A of the School Code [105 ILCS 5/Art. 24A], that assess both professional competence or practice and student growth. The purposes of this part are to:

a) identify the minimum components, including those that address the use of data and indicators of student growth as a significant factor in rating performance, of a teacher performance evaluation system and of a principal and, as applicable, assistant principal performance evaluation system that each school district must implement;

b) provide a State model for the evaluation of teachers that addresses *the use of data and indicators on student growth as a significant factor in rating teacher performance*, some or all of which shall be required of a school district under certain circumstances outlined in Section 24A-4 of the School Code [105 ILCS 5/24A-4]; and

c) establish criteria for locally developed programs to prequalify and retrain evaluators, pursuant to Section 24A-3 of the School Code [105 ILCS 5/24A-3] (“23 Illinois administrative code 50,” 2013)

(c) Is the policy being implemented/pursued in a manner consistent with its stated purpose?

Yes

Section 50.20 : Applicability

Sections 24A-2.5 and 24A-15 of the School Code [105 ILCS 5/24A-2.5 and 24A-15] establish the dates for specific groups of school districts (or for schools within certain districts) to implement performance evaluation systems, including both professional practice and data and indicators of student growth, for teachers, principals, and assistant principals that meet the requirements of this Part and Article 24A of the School Code and, for City of Chicago School District 299 (CPS), Sections 34-8 and 34-85c of the School Code [105 ILCS 5/34-8 and 34-85c].

a) Each school district shall implement a performance evaluation system for principals by September 1, 2012. (See Section 24A-15 of the School Code.)

b) Each school district located outside of the city of Chicago shall implement a performance evaluation system for assistant principals by September 1, 2012. (See Section 24A-15 of the School Code.)

c) CPS shall implement a performance evaluation system for teachers *in at least 300 schools by September 1, 2012 and in the remaining schools by September 1, 2013.* (Section 24A-2.5 of the School Code.)

d) School districts that have received a grant under Section 1003(g) of Title I of the Elementary and Secondary Education Act (ESEA; 20 USC 6301 et seq.), as reauthorized by the

No Child Left Behind Act of 2001 (PL 107-110), shall implement a performance evaluation system for teachers in those schools that are covered by Section 1003(g) funds by the date set forth in the approved grants. (See Section 24A-2.5 of the School Code)

e) School districts located outside of the City of Chicago whose student performance ranks in the lowest 20 percent among school districts of their type (i.e., unit, elementary or high school) shall implement a performance evaluation system for teachers by September 1, 2015. (See Section 24A-2.5 of the School Code.) For purposes of this subsection (e), “student performance” shall be determined based upon a school district’s overall performance on the spring 2014 administration of the State assessments authorized under Section 2-3.64 of the School Code [105 ILCS 5/2-3.64].

f) Any school district not subject to subsection (c) or (e) of this Section and schools located in school districts subject to subsection (d) of this Section that are not covered by a grant under Section 1003(g) of Title I of ESEA shall implement a performance evaluation system for teachers by September 1, 2016.

g) In accordance with the provisions of Section 24A-2.5 of the School Code, a school district and either its exclusive bargaining representative of teachers or its teachers, if the teachers are not represented by an exclusive bargaining representative, may jointly agree to an implementation date that is earlier than the date specified in this Section for their district type. When an earlier implementation date is agreed upon, the school district shall provide to the State Board of Education, within 30 days after an agreement is executed, a dated copy of the written agreement specifying the agreed upon implementation date and signed by the district superintendent and the exclusive bargaining representative or teachers, as applicable (“23 Illinois administrative code 50,” 2013).

3. Policy Design

- a. Is the policy mandated or voluntary?
- b. If voluntary are there any inducements?
- c. Who coordinates Policy?

(a) Is the policy mandated or voluntary?

With the signing of the Performance Evaluation Reform Act (PERA) in 2010, all school districts in Illinois were mandated to create new teacher evaluation systems (ISBE, 2012).

(b) If voluntary are there any inducements?

No inducements

(c) Who coordinates policy?

The State Board of Education will coordinate the policy, but the individual districts will do the evaluating and report the scores to the ISBE (ISBE, 2012).

Section 50.130: Reporting

a) By no later than June 30th of each year, the State Board of Education shall identify the manner and timeline for the submission of data and other information relative to performance evaluations that each school district must submit. These data and information shall include, but not be limited to, data regarding the performance evaluation rating given to each tenured and non-tenured teacher and data about teacher retention, as well as other information specific to the locally adopted performance evaluation plan that will assist the State Board of Education in determining whether performance evaluation systems are reliable and valid, improve student achievement, and contribute to the development of staff. (See Section 24A-20 of the School Code.) (“23 Illinois administrative code 50,” 2013).

b) A school district shall not be required to submit its performance evaluation plan for teachers to the State Board of Education for review, comment, or approval, unless specifically requested by the State Board of Education.

4. Policy Components, Mechanisms, and Procedures

- a. Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?
- b. Evaluative rubric?
- c. Types of data used to evaluate?
- d. How data are collected? (i.e., By whom? How often? What means?)
- e. Mechanisms for training evaluators? Of what does it consist? How is it done?
- f. Procedures used to ensure validity and reliability of evaluation?
- g. How are data generated from teacher evaluation used? For formative or summative purposes? How specifically?
- h. Legal Challenges/Issues Associated with Policy in State.
- i. Distinctive Features of Teacher Evaluation Policy found in this state and not the others.

(a) Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?

The Illinois Administrative Code states:

In order to assess the quality of the teacher's professional practice, the evaluation plan shall include an instructional framework developed or adopted by the school district that is based upon research regarding effective instruction; addresses at least planning, instructional delivery, and classroom management; and aligns to the Illinois Professional Teaching Standards. (23 Ill. Adm. Code 50.120[a])

Also each evaluation system will be objective based on the Illinois Teaching Standard (“23 Illinois administrative code 50,” 2013).

(b) Evaluative rubric

PERA Administrative Rules: the school district is required to use an instructional framework that is based on research regarding effective instruction, addresses at least planning, instructional delivery and classroom management, and aligns to the Illinois Professional Teaching Standards. The framework shall align to the roles and responsibilities of each teacher who is being evaluated and contain a rubric that aligns to the instructional framework being used. See Proposed PERA Administrative Rules, Section 50.120(a). The teacher evaluation plan must, by statute, consider the teacher’s attendance and competency in the subject matter taught, as well as specify the teacher’s strengths and weaknesses and the reasons for identifying the areas as such (ISBE, 2012).

(c) Types of data used to evaluate?

Observations both formal and informal, along with multiple measures of student growth. The composite scores will be 20 points: Student growth on state assessments or a comparable measure of student growth using a Student Learning Objectives (SLOs) process for teachers in non-tested subjects increased to 25 points for teachers with an approved value-added growth model (VAM); 20 points: Other locally selected measures of student growth or achievement. Decreased to 15% for teachers with an approved VAM; and 60 points: Other multiple measures of teacher/principal effectiveness.

(d) How data are collected? (i.e., By whom? How often? What means?)

Observation will be completed for a tenured teacher who has received an “Excellent” or “Proficient” (or “Satisfactory” prior to switch to the four rating system) performance evaluation

rating in his or her last performance evaluation and been observed at least twice during the 2-year evaluation cycle, with at least one observation being formal, a tenured teacher who has received a “Needs Improvement” or “Unsatisfactory” performance evaluation rating in his or her last performance evaluation be observed at least three times during the school year following such evaluation rating, with at least two of the observations being formal, and a non-tenured teacher be observed at least three times, with at least two of the observations being formal.

The PERA Joint Committee is to identify at least one Type I or Type II assessment for each type of teacher with the understanding that the other assessment for the teacher will be a Type III.

Illinois Administrative Code defines a Type I assessment as a reliable assessment that measures a certain group or subset of students in the same manner with the same potential assessment items, is scored by a non-district entity, and is administered either statewide or beyond Illinois. Examples include assessments available from the Northwest Evaluation Association (NWEA), Scantron Performance Series, Star Reading Enterprise, College Board’s SAT, Advanced Placement or International Baccalaureate examinations, or ACT’s EPAS® (i.e., Educational Planning and Assessment System). A Type II assessment is any assessment developed or adopted and approved for use by the school district and used on a districtwide basis by all teachers in a given grade or subject area. Examples include collaboratively developed common assessments, curriculum tests, and assessments designed by textbook publishers. A Type III assessment is any assessment that is rigorous, that is aligned to the course’s curriculum, and that the qualified evaluator and teacher determine measures student learning in that course. Examples include teacher-created assessments, assessments designed by textbook publishers, student work samples or portfolios, assessments of student performance, and assessments

designed by staff who are subject or grade-level experts that are administered commonly across a given grade or subject. A Type I or Type II assessment may qualify as a Type III assessment if it aligns to the curriculum being taught and measures student learning in that subject area (see Section 50.110(b)(2) of this Part) (ISBE, 2012).

(e) Mechanisms for training evaluators? Of what does it consist? How is it done?

Evaluators are required to complete a series of training modules and pass two assessments. CPS also employs trained specialists to help with calibration and assigning evidence-based ratings aligned with the rubric (ISEB, 2012).

(f) Procedures used to ensure validity and reliability of evaluation?

CPS uses a modified version of the Charlotte Danielson Framework for Teaching (widely used and nationally respected). All other districts are required to use an instructional framework that is based on research regarding effective instruction; addresses at least planning, instructional delivery, and classroom management; and aligns to the Illinois Professional Teaching Standards (ISBE, 2012).

(g) How are data generated from teacher evaluation used? For formative or summative purposes? How specifically?

Summative / a rating will range from 100-400 based on all relevant measures of performance (ISBE, 2012).

(h) Legal Challenges/Issues Associated with Policy in State.

As policymakers push for revised and new teacher evaluation methods, the Chicago teachers union is pushing back. Chicago school leaders proposed a new evaluation method that would require 25% of a teacher's evaluation to come from student performance on standardized tests (Omer, 2012). The Chicago Teachers' Union (CTU) predicts that over 6,000 teachers will

lose their jobs under the new teacher evaluation policy because the CTU states 10 to 20% of Chicago students did not meet state standards in math and reading due to their neighborhood and home environments. Along with the CTU, college professors and researchers have come out with concerns with the new teacher evaluation models being implemented. There are three concerns the researchers have listed in a letter to Chicago Mayor Rahm Emanuel, Chicago Public Schools CEO Jean-Claude Brizard, and the Chicago School Board. The first concern is that CPS is not ready to implement a teacher-evaluation system that is based on significant use of “student growth,” educational research and researchers strongly caution against teacher-evaluation approaches that use Value-Added Models (VAMs), students will be adversely affected by the implementation of this new teacher-evaluation system.

The concerns are straight forward with the researchers finding that the use of assessments that will be used to determine student growth (Type I, Type II) have not been identified and when identified they must be validated and deemed reliable to assess what needs assessing. The assessment scores will have to be controlled for socioeconomic status of students, special needs determination of students, and other variables that may affect a student’s ability to score at the rate of his/her peers. The use of VAMs as an indicator of teacher effectiveness does not work because of the instability of the models; the evidence does not show that the use of VAMs increase student achievement and the assessments used are not right to judge teacher effectiveness. Finally, the students will suffer if these models are used to determine teacher effectiveness. The researcher feel teachers will teach the test, try and avoid what they may considers low performing students and this model will create tense among teachers instead of a culture of collegiality (Strauss, 2012a).

(i) Distinctive Features of Teacher Evaluation Policy found in this state and not the others.

Tenured teachers who are rated as “Excellent” or “Proficient only need one observation during 2-year evaluation cycle

5. Policy Effects and Feedback: Pro and Con

As noted from various sources and by various group (e.g., evaluation studies, newspaper reports, teachers union reaction, lawsuits, etc.), what positive and/or negative reactions and effects OR problems/successes has the policy produced?

According to Spote et al. (2013), Chicago teachers and administrators believe in the new process for evaluating teachers and are under the agreement that it will improve instruction. However, evaluators are spending an enormous amount of time on the evaluation process, which takes them away from doing other things need in schools.

Alabama Teacher Evaluation Model: PEPE to EducateAlabama

Description of Alabama’s Teacher Evaluation Policies

In 1988, the Alabama State Department of Education adopted a resolution calling for all professional educators to be evaluated based on new board criteria. The board allowed each school system to design their own evaluation system based on the board’s criteria or use the board’s new teacher evaluation system called The Alabama Professional Education Personnel Evaluation Program (PEPE). PEPE’s primary goal is to improve teaching and learning in our schools. PEPE’s emphasis was placed on knowledge/skills and competencies, instead of personal traits (“Alabama Pepe Manual,” 2002).

According to PEPE,

Unlike some assessment measures, the Alabama Professional Education Personnel Evaluation Program is predicated on the belief that because classroom teaching, leadership, and management processes in a school environment are extremely complex, multiple data sources and data collection tools are necessary to obtain an accurate appraisal of professional practices and needs. While no personnel evaluations are absolutely objective in the sense that they are completely devoid of supervisory judgment, the model used by PEPE does help to ensure that such judgments are not capricious and arbitrary, but instead are supported not only by reliable and valid data, but also defensible rationales derived through a structured process of data collection, interpretation, and inference. (“Alabama Pepe Manual,” 2002, p. 1)

EducateAlabama

The Alabama State Department of Education overhauled its teacher evaluation system in 2005. The change from PEPE to the new EDUCATEAlabama Collaborative Teacher Evaluation System was based on a need to examine data based on a teacher’s current performance to a new set of standards developed by the Governor’s commission on Quality Teaching.

EducateAlabama is a formative evaluation system. This system uses a teacher’s current practices and weighs them against the Alabama Quality Teaching Standards (AQTS). EducateAlabama is designed to help teachers grow professionally by using data to set expectations and goals for teachers. However, EducateAlabama does not use student achievement data to evaluate teacher’s performance (“Educatealabama,” 2012).

The 2011 school year began the process for EducateAlabama’s formative process to be implemented fully. The process calls for a self-assessment using the AQTS and the Alabama Continuum for Teacher Development, initial conversation between evaluator and teacher, and a completed Professional Learning Plan (PLP) that should be based on a teacher’s self-assessment.

Teachers will receive two unannounced classroom observations and will have an end-of-the-year PLP summary.

Alabama: Data Matrix Summary

1. Policy Origin

- a. What prompted the need for a teacher evaluation policy?
- b. Who initiated it?
- c. Who was involved in creating the new evaluation policy?

(a) What prompted the need for a teacher evaluation policy?

A group of stakeholders, comprised of Alabama instructional leaders and teachers/educators, worked diligently in 2009 to develop the EA process. As a result of that committee's work, the EducateAlabama Professional Learning Collaborative was chosen to replace the Professional Education Personnel Evaluation (PEPE) Program as the evaluation process needed to meet the state's 1988 requirement of personnel evaluation for Alabama's

(b) Who initiated it?

(c) Who was involved in creating the new evaluation policy?

Alabama instructional leaders and teachers/educators

2. Policy Purpose

- a. Is the Policy implicit or Explicit?
- b. Is the purpose clearly stated?
- c. Is the policy being implemented/pursued in a manner consistent with its stated purpose?

(a) Is the Policy implicit or Explicit?

The policy is explicit. EducateAlabama (EA) is a formative system designed to provide information about a teacher's/educator's current level of practice within the Alabama Continuum for Teacher Development, which is based on the Alabama Quality Teaching Standards (AQTS).

(b) Is the purpose clearly stated?

The purpose of this new method of evaluating teachers is to provide teachers with a collaborative, formative tool to enhance teaching and learning. EA supports sustained and collaborative activities for teachers/educators designed to increase the academic achievement of all students.

(c) Is the policy being implemented/pursued in a manner consistent with its stated purpose?

Yes

3. Policy Design

- a. Is the policy mandated or voluntary?
- b. If voluntary are there any inducements?
- c. Who coordinates Policy?

(a) Is the policy mandated or voluntary?

EA is mandated to all school districts in Alabama

(b) If voluntary are there any inducements?

There are no inducements

(c) Who coordinates Policy?

The policy is coordinated by the state department of education. LEA is responsible for collecting data and reporting it to the state department of education.

4. Policy Components, Mechanisms, and Procedures

- a. Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?
- b. Evaluative rubric
- c. Types of data used to evaluate?

- d. How data are collected? (i.e., By whom? How often? What means?)
- e. Mechanisms for training evaluators? Of what does it consists? How is it done?
- f. Procedures used to ensure validity and reliability of evaluation?
- g. How are data generated from teacher evaluation used? For formative or summative purposes? How specifically?
- h. Legal Challenges/Issues Associated with Policy in State.
- i. Distinctive Features of Teacher Evaluation Policy found in this state and not the others.

(a) Theoretical basis, literature, and working causal assumptions on which evaluation policy rests?

The adoption of the AQTS, which is the foundation for teaching, will help teachers' reflection, self-assessment, and goal setting for professional learning and growth.

(b) Evaluative rubric

The Alabama Quality Teaching Standards (AQTS) is the rubric used for evaluating teachers in Alabama. No numerical scores; teacher is now rated using emerging, applying, integrating, and innovating.

(c) Types of data used to evaluate?

Classroom observations (minimum of 2)

(d) How data are collected? (i.e., By whom? How often? What means?)

Data are collected by an online self-assessment, classroom observations, creation of professional learning plan, and dialogue between evaluator and evaluatee. At the local school level, principals and assistant principals evaluate teachers.

(e) Statewide via electronic correspondence, face-to-face trainings, and state publications

(f) Procedures used to ensure validity and reliability of evaluation?

Not clearly stated

(g) How are data generated from teacher evaluation used? For formative or summative purposes? How specifically?

EA is a formative system designed to provide information about a teacher's/educator's current level of practice within the Alabama Continuum for Teacher Development, which is based on the Alabama Quality Teaching Standards (AQTS), *Alabama Administrative Code* §290-3-3-.04.

(h) Legal Challenges/Issues Associated with Policy in State.

Issues with Alabama's teacher evaluation model can be traced to the idea that it does not use student data in any form to critique or analyze a teacher's effectiveness.

(i) Distinctive features of teacher evaluation policy found in this state and not the others.

EducateAlabama is a purely formative evaluation process, and there is no use of student data

5. Policy Effects and Feedback: Pro and Con

As noted from various sources and by various group (e.g., evaluation studies, newspaper reports, teachers union reaction, lawsuits, etc.), what positive and/or negative reactions and effects OR problems/successes has the policy produced?

Comparative Analysis of Teacher Evaluation Policy Across States

In an attempt to give further clarity to this study a comparative analysis narrative of the data matrix is given. Table 2 is a complete version of the state data matrix with all anchors and sub-anchors addressed.

Table 2

Completed State Data Matrix

	ANCHORS	GEORGIA	TENNESSEE	NEW YORK	ILLINIOS/ CHICAGO	ALABAMA
POLICY ORIGIN:						
(a) Prompted Need for Teacher Evaluation Policy		GA awarded \$400 million under ARRC of 2009	TN application for RTTT funds	New York being awarded money under the 2 nd round of ARRC	Flaws in the existing models along with the anticipation of receiving RTTT money	The need for a collaborative formative teacher evaluation
(b) Who initiated?		Governor, GADOE, Governor's office of Student Achievement and education stakeholders	Tennessee First to the Top Act of 2010	Education Law section 3012-c (New York Legislature)	Performance Evaluation Reform Act 2010 (PERA) (TN Legislature)	The adoption of the AQTs by the State Board of Education
(c) Who created?		Georgia State Evaluation Steering committee	Teacher Evaluation Advisory Committee	Governing body of each school district and BOCES shall adopt a plan	Performance Evaluation Advisory Council (PEAC)	Alabama instructional leaders and teachers/educators

(table continues)

ANCHORS	GEORGIA	TENNESSEE	NEW YORK	ILLINIOS/ CHICAGO	ALABAMA
POLICY PURPOSE:	Explicit	Explicit	Explicit	Explicit	Explicit
(a) Policy implicit or Explicit					
(b) Purpose Clearly Stated	Yes	Yes	Yes	Yes	Yes
(c) Is policy being implemented/pursued in a manner consistent with its stated purpose	Yes	Yes	Yes	Yes	Yes
Policy Design	Mandated	Mandated	Mandated	Mandated	Mandated
(a) Is policy mandated or voluntary?					
(b) If voluntary are there any inducements?	Not voluntary	Not voluntary	Not voluntary	Not voluntary	Not voluntary
(c) Who coordinates policy	State coordinates	State coordinates some district choice	State coordinates but mainly district developed	State coordinates Some district choice	State coordinates
Policy components, mechanisms and procedures	Advanced	“Through a combination of frequent observation, constructive feedback, student data and meaningful professional development, the new system is designed to support all educators so they can do their best work in the classroom and help every student learn and grow.”	Based on standards with data based on student progress to include standardized test	Illinois Administrative Code 50.120	Alabama Quality Teaching Standards
(a) Theoretical basis, literature, and causal assumptions	Research on Teacher quality and Funds RTTT				

(table continues)

ANCHORS	GEORGIA	TENNESSEE	NEW YORK	ILLINIOS/ CHICAGO	ALABAMA
(b) Evaluative rubric	Teacher Assessment on Performance Standards (TAPS)	TEAM rubric	Districts can use one of several rubrics	CPS Charlotte Danielson Framework / District can use or must use other Frameworks	Alabama Quality Teaching Standards (AQTs)
(c) Types of Data	Classroom observations, teacher documentation, surveys, CRCT (4-8), and EOCT (HS) GSGM and SLO	TVASS, classroom observations, teacher conferences, prior evaluations and portfolios.	Student achievement / an approved “value-added” model by the Board of Regents / locally-selected measures that are rigorous and comparable across classrooms / multiple measures of teacher/principal effectiveness	Observations both formal and informal along with multiple measures of student growth	Classroom observations (2) Post observation conference / online self-assessment and Professional Learning Plans
1. How data are collected	Principals, AP’s, Other school administrators	Principals, AP’s, Other school administrators	Trained evaluators: Administrator or designee	Trained evaluators: Administrator or designee	online self-assessment / Building Administrator

(table continues)

ANCHORS	GEORGIA	TENNESSEE	NEW YORK	ILLINIOS/ CHICAGO	ALABAMA
2. Mechanisms for training evaluators	GaDOE Educational Specialists and school district staff	3 Day training to acquire certification	administrators must be trained in the nine elements	complete a series of training modules and passing 2 assessments	Statewide via electronic correspondence, face-to-face trainings, and state publications
3. Validity and Reliability	Validity/Reliability study conducted after pilot / More training schedule for evaluators	All evaluators must be certified / continuous training to ensure consistent scoring and calculations	Approved under the APPR standards / consistent rating procedure/must use research based rubrics outline by the state	CPS uses Charlotte Danielson Framework for Teaching / other districts are required to use an instructional framework	Not stated / Uses Alabama Teacher Quality Standards
(g) Formative or Summative (h) Legal Challenges/Issues	Both Validity of VAM, More focus on tested areas, potential civil rights litigation Ratings still high	Both Weight of TVASS / Scores for teachers in Subject areas not being tested Ratings still high	Summative Use of VAMs	Summative Weight of student performance would count	Formative Teacher Eval not assessing classroom effectiveness / No use of student data / Tenure not connected to eval / bad teachers in the classroom

(table continues)

ANCHORS	GEORGIA	TENNESSEE	NEW YORK	ILLINIOS/ CHICAGO	ALABAMA
(i) Distinctive Features	GA uses student growth percentile (SGP) for those teachers in subjects that are tested and the student learning objectives (SLOs) for teachers whose subjects are not tested	State allows for the use of another evaluation, but it must be approved by the state	NYC's VAM controls for different variables (past performance, poverty and class size)	Tenured teachers who are rated as "Excellent" or "Proficient" only needs one observation during evaluation 2 year cycle	Purely a formative evaluation without the use of student data
5. Policy Effects and Feedback: Pro and Con	Pro: Teacher input allowed Con: teacher observations scores are still high / at risk to lose RTTT money due to weaknesses in implementing its teacher-evaluation system	Pro: TN has fully implemented its teacher evaluation policy Con: teacher observations scores are still high	Pro: VAM control for different variables Con: VAM data had a high margin of error	Pro: Teachers and Admin believe in new evaluation Con: Time to complete full evaluation	Pro: EA is designed to help teachers grow professionally Con: EA does not use student data to judge teacher effectiveness, merit pay is not tied to evaluation

The first anchor focused on policy origin to include who initiated the policy and who helped created the policy. Georgia, Tennessee, New York, and Illinois all received money under the federal Race to the Top fund, this was the prompt for these states to create new teacher evaluation policies. Alabama, however, did not apply for a grant under RTTT and wanted to create a collaborative formative evaluation model. The initial salvo into teacher evaluation policy was launched not by educators, but by politicians. The Georgia Governor's office, its department of education, and stakeholders were the initiators, while in Tennessee, New York, and Illinois legislation was passed requiring new teacher evaluation policies. Alabama's state board of education adopted the Alabama Quality Teaching Standards and this ushered in EducateAlabama. Creating the different policies shows a verity of stakeholders. Georgia and Tennessee created committees to help develop the new policy, while in New York it was left mostly to the individual districts; Alabama used instructional leaders and teacher leaders.

The second anchor examined the purpose of each policy. The first sub-anchor observed whether each state's policy was implicit or explicit. The idea through decision rule logic was to determine whether or not the state's policy was clearly stated or simply implied. All of the states in this study have a very clear statement within their policy that explicitly explains its purpose. With each policy being explicit in its purpose, each state is implementing its policy in a manner consistent with its explicit purpose.

Teacher evaluations are mandated for the states studied by legislative action. The Georgia legislature passed House Bill 244, Tennessee's governor signed the Tennessee First to the Top Act, New York has Education Law Section 3012-c, Illinois passed the PERA of 2010, and all Alabama schools districts are mandated to administer EducateAlabama. The policies of the states studied are all mandates; therefore, there are no inducements. Each state coordinates its teacher

evaluation policy, but in Tennessee, New York, and Illinois individual districts have flexibility in developing or choosing a teacher evaluation model.

The theoretical basis for each design is somewhat different. Georgia used research on teacher quality to design its new teacher evaluation model, while Tennessee used more of a causal assumption for the creation of its TEAM model. New York and Alabama based their model design on standards literature, and Illinois used its Administrative Code 50.120.

Each state used an evaluative rubric. Georgia has its TAPS, Tennessee created the TEAM rubric, and Alabama uses AQTs. New York and Illinois give its districts the option of choosing between several different rubric designs to include Charlotte Danielson's Framework for Teaching.

The types of data collected for each state were similar in some ways and different in others. All of the states, with the exception of Alabama, use multiple measures for its evaluation. Classroom observations are used in each of the states studied, with each state using different criteria for the number of observations and whether they will be formal or informal. Georgia has implemented the use of both grow models and student learning objectives. The GSGM is used for subjects that are tested on the CRCT and Student Learning Objectives are used for the non-tested subjects. High school students in Georgia are required to take the EOCT. Tennessee uses its TVASS and New York and Illinois use a VAM, while Alabama does not use student data for teacher evaluations. The data are collected by trained administrators or their designee. Each state requires their evaluators to be trained; however, the amount of training varies. Georgia uses State Department of Education Specialists to train its evaluators, Tennessee requires its evaluators to complete a 3-day training for certification, New York requires completion of a nine-element training, Illinois evaluators must complete training modules and pass two assessments, and

Alabama evaluators are trained via electronic correspondence, face-to-face training of state publications.

The validity and reliability are important in any evaluation model. This study looked at how each state was going to try and ensure validity and reliability of its teacher evaluation model. Georgia conducted a pilot evaluation using 26 districts, so after the pilot the state was going to evaluate the pilot data and schedule more training. Tennessee requires all evaluators to be certified and be given continuous training. New York and Illinois requires the use of research-based rubrics to ensure validity and reliability, whereas Alabama's teacher evaluation policy does not address the issue of validity and reliability.

Georgia and Tennessee use their evaluations as both a summative evaluation and a formative evaluation. New York and Illinois both have a summative evaluation model and Alabama's model is formative.

There are always issues when implementing something new, and the new evaluation models are no different. The use of student data is being questioned in Georgia, Tennessee, New York, and Illinois. The weight the data will count is causing concerns in Tennessee and Illinois. An issue in Georgia and Tennessee centers around the focus of the new model being on the tested subjects. The states using VAMs and growth models are seeing teacher ratings still too high. Alabama's issues are centered on not using student data, tenure not being connected to the evaluation, and how is EducateAlabama assessing teacher effectiveness in Alabama classrooms.

Each state in the study has a distinct feature about its new evaluation policy. For example, Georgia is using student growth percentile while Tennessee, New York, and Illinois use VAMs. Tennessee is giving more latitude to its district for use of different types of evaluation models and New York's VAM hopes to control for a student's past achievements,

socioeconomic status, and other variables that may affect a student's achievement level. Illinois districts have the option of lowering the number of evaluation if teachers score at a certain level, while Alabama has a purely formative evaluation model that looks help teachers grow professionally.

There are positives and negatives to each state's teacher evaluation policy. In Georgia one positive is that the state chose to pilot its evaluation in 26 districts and reevaluate before it was implemented statewide. However, after it was piloted, the results showed teachers still being rated at very high success rates and Georgia lost funding from RTTT because of the failure to fully complete its evaluation policy. Tennessee has fully implemented its new evaluation policy, but teacher observation scores are still too high. New York gets credit for using VAMs that control for different variables; however, there was a high margin of error when they were used. Teachers and administrators believe a new evaluation will help improve the quality of teachers, but the weight of student data and the time it takes to complete a full evaluation are concerns. Alabama's EducateAlabama teacher evaluation model does seek to grow teachers professionally, which is a positive step. Nevertheless, it does not include student data to judge teacher effectiveness.

The state data matrix was designed to compare five states' new teacher evaluation policies. The matrix used anchors and sub-anchors to compare different components of each state's new policy. The above narrative explains the matrix in written form, as to better understand the comparison of the anchors and sub-anchors.

Summary

Chapter IV examined teacher evaluation policies in five states and compared the states to one another. The overall consensus of Chapter IV is that every student needs an effective teacher and having an effective teacher evaluation model can and should be at the forefront of educational reform.

Georgia developed its Teacher Keys Effectiveness System (TKES) model in an effort to optimize student achievement and secure funds from RTTT. TKES consists of three components that will contribute to the Teacher Effectiveness Measure (TEM): teacher assessment on performance standards, student growth and academic achievement, and surveys of instructional practice. With much discussion, Georgia will fully implement its TKES to all districts beginning with the 2014-2015 school year.

The Tennessee state General Assembly passed the First to the Top Act in January 2010. The First to the Top Act created the 15-member Teacher Evaluation Advisory Committee (TEAC) to collaborate and recommend guidelines for the Tennessee new teacher evaluation model. The Tennessee Educator Acceleration Model (TEAM) has been approved as the evaluation model for the majority of school districts in Tennessee. Three additional models were approved also: Project Coach for Hamilton County Schools, Teacher Effectiveness Measure for Memphis City Schools, and the Teacher Instructional Growth for Effectiveness and Results for 12 municipal school systems (TDOE, 2012).

Tennessee's model uses a formative and summative approach to evaluation, with multiple measures of data used for assessing a teacher's effectiveness. The models are in place and being implemented throughout Tennessee.

New York is home to the United States largest school district, New York City School District. New York State, along with the Governor, State Board of Education, State Superintendent, and many other stakeholders felt a need to develop a standards-based teacher evaluation model. The model was created with the passage of Education Law Section 3012-c. This law paved the way for the creation of an Annual Professional Performance Review (APPR) for educators. The legislation wanted teachers ranked in one of four categories: highly effective, effective, developing, or ineffective.

Under the APPR, New York school districts have the option of developing their own evaluation model, but it must meet the guidelines outlined in APPR. Also districts may choose which rubric it is going to use; again, it must meet the guidelines that are laid out in APPR. There were concerns that were voiced by the teachers unions and a group of college professors. These concerns mainly addressed using student data to determine teacher effectiveness and how much of a percentage it would count toward a teacher's final evaluation score.

Illinois governor Pat Quinn approved of Performance Evaluation Reform Act (PERA) by signing it into law in 2010. This act required all Illinois school districts to revamp how they determined teacher effectiveness in the classroom (ISBE, 2012). The Governor, along with other stakeholders, wanted to produce a new way of determining teacher effectiveness by using multiple measures. Illinois, along with other states, applied for federal monies under RTTT. PERA requires Illinois school districts to evaluate teachers using growth models by the 2016-2017 school year.

Like New York State, Illinois's largest school district, Chicago, has provided challenges to the teacher evaluation debate. Chicago teachers, with the backing of the Chicago Teachers Union (CTU), went on strike to protest the percentage of student data that was going to be used

to determine a teacher's effectiveness. The Illinois teacher evaluations are summative in nature and would have the potential to be the key to dismissal of hundreds of teachers.

Alabama created a new formative evaluation system with the purpose of providing information about a teacher's/educator's current level of practice within the Alabama Continuum for Teacher Development, which is based on the Alabama Quality Teaching Standards (AQTS). Alabama was not one of the 12 states to receive RTTT funds as the Nation moves into a new era of public education. However, Alabama does use a framework of standards to label a teacher's skill level or progress and it outlines a path for improvement using its professional learning plan (PLP).

Taking a look at the newly designed teacher evaluation models should help the state of Alabama as the Nation moves into an era of educational reform. These states are some of the first to venture out to find ways to develop a policy/model for evaluating teachers and determining teacher effectiveness to increase student achievement.

CHAPTER V

SUMMARY, IMPLICATIONS, DISCUSSION, AND RECOMMENDATIONS FOR FUTURE RESEARCH

The purpose of this study was to examine teacher evaluation policies and models to ascertain their qualities and practices in an effort to recommend policy targets and reforms for the state of Alabama. The study was guided by researching the history of educational accountability. The study looked to problematize the casual assumptions of school accountability and frame a picture around teacher evaluations as a policy target to improve student achievement.

The study gives an outline for school accountability, starting with President Lyndon Johnson's War on Poverty that provided American schools with the Elementary and Secondary Education Act. The ESEA provided additional money to America's poorest school kids. A Nation at Risk studied the educational landscape of America and compiled a list of findings and recommendations for the American public. The findings and recommendations for teaching found in A Nation at Risk was to be a blueprint for schools as America looked to improve student achievement.

Following A Nation at Risk, Presidents have tried their hand in the education reform debate. President George H. W. Bush's administration released America 2000, which stated that there was "a rising tide of mediocrity" in America's educational system. President Clinton followed America 2000 with Goals 2000. Goals was a reauthorization of ESEA and when Congress reauthorized the ESEA it had a focus on school readiness, school completion, student

academic achievement, leadership in math and science, adult literacy, and safe and drug-free schools, which were holdover ideas from America 2000 (Paris, 1997).

President George W. Bush, in an effort to increase standards and in particular close the achievement gap between students from varying backgrounds and educational levels, reauthorized ESEA as No Child Left Behind. NCLB's goal was to have all school children proficient in reading and math by 2014.

A Race to The Top became the educational competition of the Obama administration and this fund was set with the federal government providing \$4.35 billion to fund this initiative. It was to be divided among the states that met certain criteria when they applied for funds (Race to the Top, 2009).

As educational reforms, initiatives, and competition became talking points for stakeholders, there has become a need for solutions and tools to help determine the effectiveness of schools. This study gives a brief overview of the standards movement, merit pay, high stakes testing, charter schools and teacher evaluation as a means to improve student achievement.

Questions Asked by the Study

1. What does the literature have to say about teacher evaluation policy and how evaluations are being used as an accountability tool for states and school district?

Teacher effectiveness has become a major focus of education policy, with RTTT funds becoming available for introducing new teacher evaluation models that include student data as a measure of teacher effectiveness; states are developing and piloting models for teacher evaluations. However, research indicates it has been difficult getting all stakeholders to agree on which design is best suited to evaluate teachers.

Designing an effective evaluation model is possible. Professor Darling-Hammond gives criteria for developing an effective model. She begins with professional teaching standards, which all the states studied include in their policies. It should have multifaceted evidence of teacher practice, student learning, and professional contributions; the evaluator should be knowledgeable and well-trained; feedback is essential; teacher collaboration should play a part in the evaluation; expert teachers should be included in the review process; and a panel of administrators and teachers should be the model's overseers (Darling-Hammond, 2013).

Teacher evaluations are now being used as an accountability tool by school districts across America. Districts are developing models that act as both formative and summative assessments. The focus on teacher evaluation as an accountability tool is becoming the norm as stakeholders look for ways to determine how effective teachers are and if those teachers are adding value to the students they teach. Student growth among students is essential as we hope to prepare all students for college and careers. The central idea is to have an effective teacher in every classroom, giving every student the opportunity to have a successful academic career.

2. What is the purpose and nature of the teacher evaluation policy in the state of Alabama?

The purpose and nature of Alabama's teacher evaluation policy is to provide for the professional growth of its teachers by using data to set expectations and goals for them. Alabama does not have a student growth or value-added component to its evaluation model but does hope with the professional growth of its teachers that student achievement will increase and the achievement gap will decrease (EducateAlabama, 2012).

EducateAlabama has provided Alabama school districts with a formative evaluation that looks to provide teachers with the necessary professional development through the use of the

Alabama Quality Teaching Standards to improve instruction in the classroom. However, EducateAlabama does not use student data in determining if teachers are effective or ineffective; this feature is a part of many of the new evaluation models beginning implemented as well as a key component of the Obama Administration's Race to the Top initiative for states.

3. What is the purpose and nature of the teacher evaluation policy enacted by a sample of four state designs other than Alabama?

The purposes outlined by the four states in this study are all similar in content and design. The studied states are looking for ways to improve teacher quality and student achievement. The use of multiple measures and improved classroom observation techniques are common themes in each new evaluation. There is a consensus that teachers should be judged using a rubric from a researched framework of excellence and the use of student data is paramount to an effective evaluation model.

Georgia began its new teacher evaluation implementation using 26 districts. Georgia's new evaluation model was piloted and what it found was similar results as previous evaluations. The results were that the majority of teachers were still being rated as very effective and very low percentages were rated as ineffective.

Tennessee has fully implemented its new evaluation models across the state, but like Georgia the results are showing a high percentage of teachers rated as effective and a very low percentage of teachers rated as ineffective. According to Stockard (2014), teacher evaluations are coming across as punishment for teachers and not a way to help improve the effectiveness of teachers. Some legislatures are looking for ways to make teacher evaluation policies easier to understand.

New York has struggled in its implementation of its teacher evaluation policies. The major concern has been with teacher ratings. The New York State Department of Education has acknowledged discrepancies and flaws with its teacher ratings (Hildebrand, 2014). The VAMs used in the many districts, including New York City, are cause for alarm along with the call by teachers' unions that the evaluations rely too heavily on standardized tests.

As teacher evaluations rolled out in Illinois and Chicago, there were concerns regarding the amount of time it takes to complete the evaluation process. This claim, although many teachers feel that new policies will help improve teaching, does take time away from other responsibilities teachers and administrators have to complete.

4. What can the state of Alabama and its school districts learn from (a) the literature on teacher evaluation policy and (b) a review of teacher evaluation policy in a select group of states that were examined in this study? What information do these data sources provide that could improve teacher evaluation policy in Alabama?

There are several fundamental issues that states/districts must address when considering a new or revised teacher evaluation policy. The issues range from which set of measures to use to how much each measure should count toward a teacher's final evaluation rating. In discussing the fundamental issues of today's new teacher evaluation, it should be noted that the literature gives several methods states/districts should use. The literature tells us that new evaluations should include multiple measure data. The data could range from classroom observations, surveys, and student achievement on various measures of student data. Issues with classroom observations are twofold: one deals with what is to be observed and the second is how many times a teacher should be observed. The frequency of classroom observations can run from 0 for some tenured teachers to as many as 4 per year for novice teachers or those not rated satisfactory.

Many states are using frameworks or standards to guide teachers in what is expected of an effective teacher.

Surveys pose another issue for new teacher evaluation policies. Georgia wants to use student surveys as a part of a teacher's rating, but some have concerns with the legal challenges that might arise if this method is used. Also, what information will the surveys look for and will the surveys be a valid and reliable tool to use in determining teacher effectiveness?

The proverbial elephant in the new teacher evaluation room is how to use student data as a measure of teacher effectiveness. Under the RTTT initiative, the Obama Administration has stressed the need to use student data as a measure of teacher effectiveness. All of the states compared in this study, except Alabama, use student data as part of their quantitative evaluation.

The issues that have been discussed center around using the new teacher evaluation policies to determine performance/merit pay and the cost of developing and implementing these new models. Stakeholders are debating ways to compensate the most effective teachers and many feel using teacher evaluation models that implement multiple measures to include student data is a good way to start. However, teachers and teachers' unions may push back, and merit/performance pay does not have a good track record for rewarding the most effective teachers. Concerns over how these new evaluations are going to be funded once the RTTT money runs out leaves some school districts to take a close look at the process.

As teacher evaluations become a higher priority for school organization, the development and implementation of these evaluations will take on a more prevalent role in today's schools. The literature shows that past teacher evaluations have been inadequate to say the least. The designs do not help teachers and, in turn, do little to nothing about improving student achievement or decreasing the achievement gap.

Teacher evaluation in its current state is viewed as a burden by the evaluators and treated as just a routine event that has to be completed each year. The overwhelming majority of teachers are rated as excellent or exceeding expectations, while student achievement is at a standstill. Teacher evaluation can improve teacher quality as stated by some of the leading professors (i.e., Linda Darling-Hammond, James Stronge, and James Popham).

Alabama has the ability to observe the states that are implementing new teacher evaluation design and judge what is working and what is not. The AQTs is a great start, as most of the states are incorporating teaching standards in their evaluation models. The key will be for Alabama to add a student data section to its evaluation model and, with that, determine what percentage it should count toward a teacher's final evaluation rating. Alabama will need to determine if a growth model design is warranted, a VAM, or a combination of both and with this how much of a percentage will it count toward a teacher's final evaluation score. Also, will Alabama tie personnel decisions or teacher compensation to any new teacher evaluation model developed?

Alabama has the capability to develop a teacher evaluation model that incorporates the ideas and input of all stakeholders. Teachers must play a vital role in any new evaluation policy that is designed by the state. The goal is reachable through communication and collaboration between all stakeholders.

Implications for States Looking to Design New Teacher Evaluation Policies

Teacher evaluation as a policy target became the focus of this study. The history of evaluation, to the most recent attempts at finding a tool to help provide every student with a

highly effective teacher, was researched and studied. Evaluation models are beginning to take shape as states transition from NCLB to Race to the Top competition for additional funds.

States are in a competition for Race to the Top funds and one of the criteria is to have in place a teacher evaluation system to measure the effectiveness of teachers. One of the criteria for great teachers and leaders is improving teacher and principal effectiveness based on performance. Each of the state evaluation systems researched in this study has developed a new method of evaluating teachers. The one common component to these systems is the use of multiple measures to evaluate teachers, with student data being a major component. This component is lauded by policymakers and some stakeholders while it is being questioned by educators. States will have to balance the wants of policymakers, concerns of educators, and the needs of students as they continue to revamp their teacher evaluation models.

According to Scott (2013), Tennessee's Race to the Top Grand award was over \$500 million to help with its implementation of new teacher evaluation policies and other RTTT reforms. New York was granted over \$690 million, Georgia was awarded over \$400 million and Illinois was awarded \$28 million in phase III of RTTT grants. The grants were awarded over a 4-year grant period and for the phase I and phase II states the end of the 4-year window will close in 2014. These are significant amounts of money states received from the federal government, but as the money goes away or a new administration takes office, states will have to find a new revenue source to maintain the consistency, reliability, and validity of the evaluation policies. This will be a major challenge for states as the lack of revenue sources may have been the driving force behind the application for RTTT funds. States will have to be creative if they are to maintain the teacher evaluation policies as they were created and modified.

Alabama attempted twice to win money from the RTTT fund. Phase I of RTTT, Alabama was awarded a score of 291.2 out of a possible 500 (USDOE, 2012c). There were several areas in which Alabama received less than 50% of the total number of points allowed. Alabama's low areas were in articulating states education reform agenda and LEA's participation in it, securing LEA commitment, using broad stakeholder support, improving teacher and principal effectiveness based on performance, developing an evaluation system, conducting annual evaluations, using evaluations to inform key decisions, ensuring equitable distribution of effective teachers, ensuring equitable distribution in high-poverty or high-minority schools, ensuring equitable distribution in hard-to-staff subjects and specialty areas, turning around the persistently lowest-achieving schools, making educational funding a priority, and enduring successful conditions for high-performing charter schools and other innovative schools. The final section of Alabama's application which was rejected by all five reviewers was Absolute priority-comprehensive approach to educational reform (USDOE, 2012c). The interesting take for this study revolves around the evaluation of teachers in section D of the RTTT application. The section awarded a total of 138 possible points and Alabama's take was only 69.2 (USDOE, 2012c).

Alabama reapplied for the RTTT grant during its phase II competition cycle. Alabama's application was awarded 212 of the possible 500 points (USDOE, 2012d). Again, the reviewers graded Alabama down in the same areas as the phase I cycle. The section of the application that addressed charter schools saw Alabama rated a 0 in all four categories, and in the section Great Teachers and Leaders, Alabama was awarded 38.8 points out of a possible 138 points (USDOE, 2012d).

Forty-one states are now required, or it is recommended, that their teacher evaluation policies include multiple measures, while more than 20 states are requiring student achievement data to be a part of the policy (Hull, 2013). As outside pressure mounts, along with the need for federal money, the implications for states to move toward these evaluation methods will increase.

The process, however, can become murky, with complications among stakeholders as to how these new policies should be written. According to Petrarca et al. (2012), teachers' unions will debate procedural matters to all new teacher evaluation policies. The procedural matters usually are concerned with deadlines for completing the evaluations, components of the evaluation plan, number of observations, documentation provided to the teacher, and when and how the documentation will be provided (Petrarca & Scariano, 2012). The inclusiveness of the new state policies will be tested as the debates between teachers' unions and state officials heighten like in Chicago and New York.

As teachers' unions and states banter back and forth over what makes a good teacher evaluation policy, the largest teachers' union in America has given its guidelines. First the National Educators Association (NEA) believes all teachers should be regularly evaluated by highly trained evaluators on the basis of clear standards. NEA states the standards should be high and rigorous to encompass a teacher's knowledge, expertise, temperaments, and responsibilities. NEA acknowledges evaluations must use multiple measures, to include indicators of teacher practice, indicators of teacher contribution and growth, and indicators for student growth ("New policy statement," 2013).

The American Federation of Teachers, the second largest teachers' union in America outlines four components they view as crucial to new teacher evaluation models. Professional

teaching standards that advance an all-inclusive vision of the teaching profession is first, followed by a set of standards for measuring teacher practice to include what is considered effective teaching and student growth, next there has to be standards for implementation which should cover the procedures for evaluations, and finally there needs to be standards for professional contexts. The final standard should explain the condition of a school in which teachers may or may not have control over (“Teacher development and,” 2012).

The implications for states is somewhat clear--with the call for effective teachers and effective teaching practices at the forefront of many education debates, states are bound to enter into the development of new ways to measure teachers. The new teacher evaluation policies designed by states should have and will have an effect on teachers and the teaching profession. As the states begin to implement their new evaluation models, each of the policies could be examined to determine what went right and what went wrong.

Implications for Teachers as New Teacher Evaluation Policy is Implemented

The National School Board Association (NSBA) poses an argument that state school districts cannot fire ineffective teachers. NSBA contends that teacher evaluations in their present form are unfocused, undifferentiated, unhelpful, and inconsequential. For teachers, this is saying that there is no clear purpose to what they are charged to do, especially with regard to student achievement. Most evaluations for teachers do not take into account the differences of teacher content or the type of student being taught. In the end, there is little to no feedback for teachers about their teaching habits and there is no connection to pay, tenure, and promotion, and very little professional development (“Getting value out,” 2012).

Creating a teacher evaluation model to judge a teacher on the quality of his or her job can be a daunting task. Popham (2013) gave several ideas and approaches to creating a viable teacher evaluation model. One of Popham's suggestions for an evaluation is that it must include human judgment. This comes at a time when most policymakers are pushing states to use more Value-Added Models or Student Growth Models to determine a teacher's effectiveness. Most policymakers are looking for an error free teacher evaluation system, but this hope or idea is not likely to be attained (Popham, 2013). VAM are being challenged daily for their lack of accuracy and miscalculations, as noted by Popham (2013) and Ravitch (2014).

The human element to teacher evaluation, as noted by Popham (2013), revolves around several tasks. The tasks begin with selecting criteria for the evaluation model. There has to be a clear understanding of what is to be used to judge a teacher's effectiveness. In bringing all stakeholders to the table to decide on what is the best criterion or criteria to judge a teacher's effectiveness can be difficult. Next is choosing the evidence sources to judge the criteria, followed by giving each source a weight amount. The difficulty in choosing evidence again could be problematic based on the stakeholders involved and once the source or sources have been decided how much should each count. Debates continue between stakeholders in the states compared in this study and in numerous other states that are undertaking the creation of teacher evaluations.

Are there difficulties in judging teachers' performances? The answer is a resounding yes. Marzano (2012) discusses how the purpose of evaluation models can affect the outcome of what a school district is trying to accomplish. Marzano (2012) tells us that a teacher evaluation system either fosters teacher learning or measures a teacher's competence. This creates a dilemma for both states and teachers alike. For states looking to improve teacher effectiveness, models that

develop teachers may not give the desired results. Taking the same approach, models that measure teachers' effectiveness may not provide the data to improve teacher effectiveness (Marzano, 2012). As states begin the process of developing new teacher evaluation models, emphasis must be placed on the purpose of the evaluation developed. States' evaluation models will be judged on the output of not only teacher practice but student performance. This creates a slippery slope for developers of these new designs and teachers who are being evaluated.

As new evaluation models are produced some see too much emphasis placed on exceedingly sanguine teacher evaluation literature (Hallinger et al., 2014). In determining how to judge a teacher's effectiveness, certain factors that could have a direct effect on a student's performance should not be ignored (Hallinger et al., 2014). Teachers who educate students with disabilities should not be compared or judged with the same set of standards as those who teach students who are in advanced classes. That is the quandary when it comes to new teacher evaluation models. According to Hallinger et al. (2014), recent literature is counter to what the new evaluations are trying to accomplish. These authors found that teacher evaluation does not have a high impact on school improvement. This is another dilemma for the new evaluation models. How are educators going to buy into a system that the research is saying has little value? Developing a new teacher evaluation model still will hinge on the belief in causation or correlation of a teacher's impact on student achievement. As Hallienger et al. (2014) concluded, developing new teacher evaluations to judge a teacher's effectiveness is going to take a heavy investment of time and money.

The heavy investment that will be needed to create these new teacher evaluation models will also have to ensure that the models are defensible. Popham and DeSander (2014) discuss how teachers may need saving from the new evaluation models as many teacher will begin to

lose their jobs because of ratings given under these new evaluation policies. As the RTTT competition forces more and more states to use student growth as a major factor in the evaluation policies, and classroom observations become more unreliable, teachers will begin to look for relief from the courts (Popham & DeSander, 2014).

In their opinion, Popham and DeSander (2014) stated the courts have refused to substitute their judgment for that of an education organization in the realm of teacher dismissals.

The value-added approach to teacher evaluations has become a major topic. NSBA (2012) stated that a value-added approach looks at teacher effectiveness and raises questions for teachers who are not adding value to the students they teach. Using value-added data for teacher evaluations is lauded by many stakeholders, including Brown University Professor John Papay who endorses the value-added approach over the traditional teacher evaluation models (Papay, 2012). According to Papay (2012), value-added evaluations are gaining traction with policymakers; as the nation moves to using high-stakes testing as a way to judge student achievement these tests also give measurable data on student test score increases or decreases thus giving policymakers a look into which teachers are being effective and which teachers are not. Value-added evaluations are also seen as more objective than the traditional teacher evaluation models.

Merit pay, licensure, and tenure are three topics that can be discussed when addressing ways that policymakers and other stakeholders try to find to evaluate teachers. Merit pay affects the income of teachers, licensure affects the ability of teachers to gain employment, and tenure, if gained, is the tool that allows teachers protection from unlawful termination.

The National Council on Teacher Quality (2013) conducted a study to evaluate state teacher evaluation models. The study focused on teacher evaluation policies and gave

recommendations for policy improvement. The council identifies merit pay or teacher compensation as a tool for recognizing individuals who demonstrated effective teaching. Their position is explained as “in most other professions, performance matters and good performance is rightfully rewarded with promotions and salary increases” (NCTQ, 2013, p. iii). Teacher license renewal or advancement should include data to show teacher effectiveness is also a recommendation by the council, but the great big elephant in the room for teachers is the attack on teacher tenure (Doherty & Jacobs, 2013). As the push to tie the granting of tenure to teacher effectiveness gains momentum, with policymakers and think tanks pushing for a more effective teacher workforce, there is a grave possibility that tenure in its present status might become extinct.

How do teachers feel about evaluations? Should teachers have a stake in the decision making of these new teacher evaluation models? According to O’Pry and Schumacher (2012), teachers do not necessarily dislike the instruments themselves but the way the instruments are carried out at the school level. This can pose a problem at the school level, with building administrators being primarily responsible for executing the new systems. Several factors may cause teachers to feel a little uneasy or quite comfortable with the new evaluation models. A building administrator who is well-versed in the new model helps put many teachers at ease, ones who are confident in the system. Teachers stressed the need to feel supported and prepared for the new way of evaluation, trust, collegial relationships, and finally one who gives timely feedback and thoughtful reflection (O’Pry & Schumacher, 2012).

Popham and DeSander (2014) suggest that teachers are disinterested in the inner workings of the new evaluation policies and designs thus they are not taking a front seat in the development stage of these new policies. Furthermore, it is suggested by Popham and DeSander

(2014) that the lack of interest could come from the feeling that the courts will protect teachers from being unfairly dismissed by an unfair evaluation policy.

These implications, along with others, are going to become major talking points as states begin to pilot and develop new teacher evaluation policies. Do teachers find their own solution to the achievement problem or allow others to dictate what is best for their profession.

Discussions/Recommendations for Alabama

Teacher evaluation in its current form is inadequate to determine if teachers are highly effective. According to Varlas (2012), many of today's teacher evaluation models are not built to encourage teachers to perform at a highly effective level and thus do not create an atmosphere for improving student learning. Policymakers are finding that the vast majority of states using teacher evaluation systems have very disturbing results. Teachers are receiving the same high ratings, while student scores are declining. In school districts that use binary rating systems, it was found that 99% of the teachers were rated satisfactory, while systems that used a more complex form of teacher evaluation saw 94% of its teachers rated satisfactory (Glazerman, Leob, Goldhaber, Staiger, Raudenbush, & Whitehurst, 2010).

School administrators and teachers at the local and state levels are at the forefront of the reform movement. This movement has gained momentum over the last 30 years in education as America looks to improve student achievement as well as closing the achievement gap (NGA, 2010). State educational associations, local educational associations, and school administrators are given the task of ensuring that teachers are highly qualified and performing at a level that will maximize student learning.

Teacher evaluation policy has not been viewed as a high priority method for improving educational quality thus having little to no effect on school personnel, teaching, or professional development (Darling-Hammond, 2004). Hazi and Arredondo-Rucinski (2009a) explained that teacher evaluation as it is now does not help the teacher, which, in turn, does not improve student learning. Teacher evaluation in its present form is just a routine procedure that both teacher and evaluator are asked to perform yearly (Hazi & Arredondo-Rucinski 2009a). Zirkel (2010) addressed a problem with teacher evaluation from the aspect of the evaluator. He described a Lake Wobegon effect for administrators as his study pointed out, in the district that was used evaluators gave 99% of their teachers an excellent rating on their evaluation. Further research by Zirkel showed less than 1% of teachers received an unsatisfactory rating on their teacher evaluation rating scores. The problem that this study highlighted focuses on the state of teacher evaluation policy as a tool for determining effective teaching and how with the right communication, collaboration, and commitment, evaluation systems can be used as a tool for improving teacher performance.

The discussion question that needs addressing is, “How can Alabama benefit from this study?”

Recommendation 1: Alabama educators and policymakers must remember and address the fundamental issues of any teacher evaluation initiative. The first issue of any evaluation is a written purpose. The purpose needs to be clear without ambiguities and should state the reason for the evaluation and how its purpose will be accomplished. The next issue involves explaining to what ends the evaluation will be used. Some of the ends may be, whether it be summative or formative or both, will it be used for tenure, for teacher advancement, performance/merit pay, or teacher retention to list a few. The evaluation model development must have multiple levels of

participants, which include the following: teachers, administrators, policymakers, college professors, and parents. The recommendation for the process of any new teacher evaluation model should include multiple measures to gauge a teacher's effectiveness. The process should include at least four classroom observations and the use of student data that is comprised of a VAM which controls for student variables that a teacher has no control over. To ensure validity and reliability of the process, Alabama should have yearly training for evaluators; the use of valid and reliable standards of teaching practices should be used, such as the Alabama Quality Teaching Standards or Charlotte Danielson's Framework for Teaching or any other valid teaching standards framework, as a rubric. However, one recommendation would be to adjust the standards as not to have so many to use for a teacher's rating.

The evaluation model should be used as a tool to help improve student achievement through the process of molding and retaining effective teachers. In doing so, both teachers and evaluators need to be trained on the process. The teachers should have interment knowledge on the purpose and the process of the teacher evaluation model. Evaluators should initially be trained and certified, with continuous training and professional development on the processes used to evaluate teachers. The training should be significant, rigorous, and meaningful.

The evaluation process has many intertwining parts, and it is hoped that they will lead to an effective program for judging the behaviors of teacher as they relate to student achievement. In this process, developers must take into consideration an effective way to manage the time it takes to complete an entire evaluation cycle. The process should not overwhelm the teacher or the evaluator. Time is always a key factor and should be examined very carefully when developing teacher evaluation policies.

Recommendation 2: This study recommends Alabama incorporate student data into their teacher evaluation policy and in doing so use a value-added model based on the model used in New York City. Why New York City's value-added model? The VAM used in New York City has been said to be too convoluted, too hard to understand by the average teacher, and too complicated. This recommendation comes because the model used in New York City controls for variables that have a direct effect on a student's ability to learn. These variables, be it class size, socioeconomic status, or past achievement, will give stakeholders a look into the outside influences that may cause a student to score a certain way on a standardized test. However, when incorporating a VAM, the percentage it counts toward a teacher's final rating should fall between 1 and 5% until a valid and reliable VAM is adopted by all stakeholders. The use of statistical data should be done in a prudent manner, along with the use of statistical experts to design and implement the statistical models (American Statistical Association, 2014).

The VAMs are taking a beating as teachers in Houston, Florida, and Tennessee have filed suit over its use to judge teachers' effectiveness. The use of VAMs is essential in obtaining money under RTTT, but it should not be used as a tool to determine a teacher's effectiveness but rather as a tool to determine how students are performing on high stakes test (American Statistical Association, 2014).

Recommendation 3: Teachers and the Alabama Education Association should take the lead. Teachers should be proactive and devise an evaluation plan and not wait and be reactive as was the case in the states researched in this study. AEA is the

Recommendation 4: Alabama should add a summative score to its evaluation model and have a gradual increase in the percent the quantitative data counts toward a teacher's summative score. Georgia's evaluation model calls for a both formative and summative evaluation. This

approach allows for ongoing feedback to the teacher as well as an end-of-the-year assessment of their performance.

Recommendation 5: Alabama would be well-served if they adopted the TEAM approach of having four classroom observations. EducateAlabama required two unannounced observations, but the recommendation is the more valid qualitative data Alabama can include, the better the evaluation model. The classroom observations should be centered on a valid and reliable evaluative rubric. Alabama has a set of standards, AQTs, which would serve that purpose.

Recommendation 6: It is also recommended Alabama examine the Illinois model, as it uses three types of assessment to judge student achievement. Their assessments range from those created by an outside agency, district created, and curriculum aligned assessments. All of the assessments could fit Alabama's evaluation model either as a set standardized test like the ACT Aspire, assessments for subjects not tested under Aspire, or for end-of-the-year examinations.

Alabama is at a crossroads with teacher evaluation. The report issued by the National Council on Teacher Quality (2013) gives Alabama all negative marks in regard to teacher evaluation policy. Alabama has a unique opportunity to learn from the mistakes and success of states who are implementing new models for teacher evaluation.

Recommendations for Future Research

Research should be conducted on the original 12 states that received RTTT monies that designed new teacher evaluations, after the evaluations have been implemented statewide. Also, a look into how the student growth model has fared as a viable component for teacher evaluation

models and what percentage of the evaluation model works best for school systems and teachers needs to be undertaken.

Future research for Georgia should hinge on its full implementation of its model and compare its use of SGP to other states that are using VAM. Also, there needs to be research on the Georgia plan to introduce merit/performance pay as a component to its evaluation policy.

Tennessee uses the motto First to The Top to outline its reforms in education. Further studies should be conducted to evaluate if the TEAM model and the other models allowed by the Tennessee Department of Education are producing comparable results. Researchers should consider how teachers have been affected by the tenure provision of the models along with merit/performance pay systems that will gradually be implemented.

New York has implemented an evaluation model that uses VAMs that takes into account outside variables such as poverty and class size; however, during its pilot many problems were found with the VAM, so further study needs conducting to reexamine New York's VAM.

Illinois/Chicago, like the aforementioned states, has begun implementation of new teacher evaluation models. Once their evaluation has been implemented, statewide research should be conducted to get an understanding whether districts that are using different evaluative rubrics are getting like results. In addition, research will need to investigate the concern behind the amount of time it took evaluators to conduct a full evaluation.

Research on Alabama's teacher evaluation policy should start with a look into whether or not Alabama has chosen to include student data as a part of its policy, and how effective is the policy as it is written today.

All models should be reexamined after the common core standards are fully implemented and have become a part of the new teacher evaluation models.

Finally it is recommended that research be conducted to evaluate the models and their impact or lack of impact on student achievement. As a way to judge teacher effectiveness, the research should include data on the latest court battles in Houston and Florida concerning the use of Value-Added Models. The goal of these new evaluation models is to have an effective teacher in every class. If so, what is the research saying after full implementation by all the states in this study? A qualitative study should be conducted to gauge the attitudes teachers have toward these new models and their likes and concerns as teacher evaluation becomes a major part of the American education system.

REFERENCES

- A brief history of education in America.* (n.d.). Retrieved from www.cblpi.org/ftp/SchoolChoice/EdHistory.pdf
- A Nation Still at Risk. Retrieved . . . ProQuest . . . August 19, 2011.
- Alabama PEPE Manual.* (2002). Retrieved May 23, 2012, from <http://www.alabamapepe.com>
- Adams, S., Heywood, J., & Rothstein, R. (2009). *Teachers, performance pay, and accountability: what education should learn from other sectors.* Washington, DC: Economic Policy Institute. Retrieved from <http://www.epi.org>
- American Statistical Association. (2014, April). *ASA statement on using value-added models for educational assessment.*
- An open letter of concern regarding New York State's appr legislation for the evaluation of teachers and principals. Retrieved October 3, 2013, from <http://www.newyorkprincipals.org>
- Anderson, L. (2009,). Upper elementary grades bear the brunt of accountability. *Phi Delta Kappan*, 90(6), 413-418. Retrieved from <http://www.pdkintl.org/publications/kappan.htm>
- Baker, E., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., Ravitch, D., Rothstein, R., Shavelson, R., & Shepard, L. (2010, August). *Problems with the use of student test scores to evaluate teachers.* Washington D C: Economic Policy Institute. Retrieved from [epi.org](http://www.epi.org)
- Baker, M., & Johnston, P. (2010). The impact of socioeconomic status on high stakes testing reexamined. *Journal of Instructional Psychology*, 37(3), 193-199.
- Badertscher, N. (2013, Jan. 7). New evaluation pilot "skewed" with too few unsatisfactory teachers, officials say. *The Atlanta Journal-Constitution*. Retrieved from <http://www.ajc.com>
- Better than a meaningless teacher checklist. (2013, Sept. 18). *Chicago Sun-Times*. Retrieved from <http://www.suntimes.com/opinions/22627200-474/better-than-a-meaningless-checklist-cps>
- Bird, T. (1990). The school teacher's portfolio: an essay on possibilities. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers.* Newbury Park, CA: Corwin Press.

- Borman, G., & Kimball, S. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal*, 106(1), 3-20.
- Bohte, J. (2004). Examining the impact of charter schools on performance in traditional public schools. *Policy Study Journal*, 32(4), 501-521.
- Bouchamma, Y.(2007). Evaluating teaching personnel. Which model of supervision do Canadian teachers prefer? *Journal of Personnel Evaluation in Education*, 18, 289-308.
- Bradshaw, L. (2002). Local district implantation of state mandated teacher evaluation policies and procedures: the North Carolina case. *Journal of Personnel Evaluation in Education*, 16(2), 113-127.
- Chow, A., Wong, E., Yeung, A., & Mo, K. (2002). Teachers' perception of appraiser-appraisee relationships. *Journal of Personnel Evaluation in Education*, 16(2), 85-101.
- Common core state standards initiative*. (2010). Retrieved from <http://www.corestandards.org/>
- CPS. (2012, December 11). Retrieved from <http://www.cps.edu/Pages/reachstudents.aspx>
- Creswell, J. (1998). *Qualitative inquiry and research design choosing among five traditions*. Thousand Oaks, CA: Sage Publications.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106(6), 1047-1085.
- Darling-Hammond, L. (2010). Restoring our schools. *The Nation*, 14-19.
- Darling-Hammond, L. (2013). *Getting teacher evaluation right*. New York, NY: Teachers College, Columbia University.
- Dagley, D., & Veir, C. (2002). Legal issues in teacher evaluation legislation: A study of state statutory provisions. *B.Y.U. Education and Law Journal*, 1.
- Dagley, D., & Veir, C. (2002). Subverting the academic abstention doctrine in teacher evaluation: How school reform legislation defeats itself. *B.Y.U. Education and Law Journal*, 123-139.
- Dee, T., & Keys, B. (2004). Does merit pay reward good teachers? Evidence from a Randomized experiment. *Journal of Policy and Management*, 23(3), 471-488.
- Dee, T. S., Jacob, B. A., Hoxby, C. M., & Ladd, H. F. (2010). The impact of no child left behind on students, teachers, and schools: Comments and discussion. *Brookings Papers on Economic Activity*, (00072303), 149-207. Retrieved from <http://search.proquest.com/docview/859360413?accountid=14472>

- Dennis, R. (2000). *The role of the state government in public education in the united states*. Retrieved from <http://www.departments.bucknell.edu/edu/ed370staterole.html>
- Ding, C. (2009). Measurement issues in designing and implementing longitudinal evaluation studies. *Educational Assessment, Evaluation and Accountability*, 21, 155-171.
- Doherty, K., & Jacobs, S. (2013, October). *State of the states 2013 connect the dots: using evaluations of teacher effectiveness to inform policy and practice*. Retrieved from http://www.nctq.org/dmsView/State_of_the_states_2013_Using_Teacher_Evaluations
- Donlevy, J. (2003). Educational reforms and high-stakes testing: are public schools still for the public? *International Journal of Instructional Media*, 30(3), 225-230.
- Downey, M. (2013, October 10). New teacher evaluations: Less reliance on test scores alone. *Atlanta Journal Constitution*. Retrieved from <http://www.ajc.com/weblogs/get-schooled/2013/oct/10/new-teacher-evaluations-less-reliance-test-scores/>
- Drake, P., Shanks, G., Broadbent, M. (1998). Successfully completing case study research: combining rigour, relevance and pragmatism. *Information Systems Journal*. 8, (4), 273-289.
- Educatealabama*. (n.d.). Retrieved October 4, 2012, from <http://www.educatealabma.net/about.htm>
- Ehler, M., Koedel, E., Parsons, E., & Podgursky, M. (2013). Selecting growth measures for school and teacher Evaluations: should proportionality matter (Working Paper 80)? Retrieved from National Center for Analysis of Longitudinal Data in Education Research: <http://www.caldercenter.org>
- Ellett, C., & Teddle, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17(1), 101-128.
- Federal Education Policy and the States, 1945-2009. (2009). Retrieved February 20, 2012 from http://www.archivesnysed.gov/edpolicy/research/res_essay_bush_ghw_sysremform.shtml
- Field, L. (2013, Sept 16). *City moves ahead on performance pay for teachers*. Retrieved from <http://mdjonline.com>
- Finn, C., & Hess, F. (2004). On leaving no child behind. *Public Interest*, 157, 35-56.
- Fraenkel, J., & Wallen, N. (2003). *How to design and evaluate research in education*. (5th ed.). Boston, MA: McGraw-Hill Higher Education.
- Futrell, M. H., & Brown, W. (2000). Should African Americans support the current educational reform standards movement? *The Journal of Negro Education*, 69(4), 288-304.

- Georgia Department of Education. (2012). *Georgia keys effectiveness system handbook*. Retrieved from <http://www.doe.k12.ga.us/>
- Georgia race to the top (rt3) steering committee on evaluation. (2011, August 25). Retrieved from <http://www.doe.k12.ga.us/RT3.aspx>.
- Georgia student growth model. (n.d.). Retrieved August 28, 2012, from <http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Pages/Georgia-Student-Growth-Model.aspx>
- Getting value out of value-added: Evaluating teachers based on student achievement*. (2012, August). Retrieved from <http://www.nsba.org/leadershipinsider>
- Gilroy, M. (2008). Teacher-pay reforms slowly gain momentum. *The Hispanic Outlook in Higher Education*, 18(14), 32-34.
- Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D., & Whitehurst, G. ((2010, November 17)2010, November 17). Evaluating teachers:the important role of value-added. Retrieved from [www://www.brookings.edu/research/reports/2010/11/17-evaluating-teachers](http://www.brookings.edu/research/reports/2010/11/17-evaluating-teachers)
- Goldrick, L. (2002). Improving teacher evaluation to improve teaching quality. National Governors Association, Issue Brief. Retrieved December 30, 2011, from <http://www.nga.org/cms/home/nga-center-for-best-practices/center-issues.html>
- Good, T., & Mulryan, C. (1990). Teacher ratings: a call for teacher control and self-evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin Press.
- Gordan, M. (1998). Qualitative comparative analysis. a dictionary of sociology. Retrieved May28, 2013 from Encycolopedia.com: <http://www.encyclopedia.com/doc/1088-qualitativecomparatvnylyss.html>
- Gordon, R., Kane, T., & Staiger, D. (2006). *Identifying effective teachers using performance on the job*. The Brookings Institution.
- Green III, P., Baker, B., & Oluwole, J. (2012). The legal and policy implications of value-added teacher assessment policies. *B.Y.U. Education and Law Journal*, 1-29.
- Guidance on New York state's annual professional performance review for teachers and principals to implement education law §3012-c and the commissioner's regulations*. (2013, August 30).
- Gutierrez, M. (2012, November 20). [Web log message]. Retrieved from www.ed.gov/blog/2012/what-is-eds-stance-on-testing-data-in-teacher-evaluation/

- Haertel, E. (1990). Performance tests, simulations, and other methods. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, California: Corwin Press.
- Hallinger, P., Heck, R., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of evidence. *Educational Assessment, Evaluation and Accountability*, 26, 5-28.
- Hanushek, E., & Raymond, M. (2005). Does school accountability lead to improved Student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Harrison-Jones, L. (2007). No child left behind and implications for black students. *The Journal of Negro Education*, 76(3), 346-356.
- Hazi, H., & Arredondo-Rucinski, D. (2009a). Teacher evaluation as a policy target: viable reform venue or just another tap dance? *Journal of Research and Information* 27(3), 31-38.
- Hazi, H., & Arredondo-Rucinski, D. (2009b). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *Education Policy Analysis Archives*, 14(5). Retrieved (August 23, 2010) from <http://epaa.asu.edu/epaa/v17n5/>
- Hildebrand, J. (2014, February 17). State officials admit problems with new teacher ratings. Retrieved from <http://www.newsday.com/long-island/nassau/state-officials-admit-problems-with-new-teacher-ratings>
- Hill, H., Kapitula, L., & Umland, K. (2010). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.
- Holtzapple, E. (2003). Criterion-related validity evidence for standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(3), 207-219.
- Hull, J. (2007). *Measuring student growth: a guide to informed decision making*. Retrieved from <http://centerforpublic.org/./measuring-student-growth-A-guide-to-decision-making.html>
- Hull, J. (2011, March 31). *Building a better evaluation system*. Retrieved from centerforpubliceducation.org
- Hull, J. (2013, October). Trends in teacher evaluation how states are measuring teacher performance. Retrieved from <http://www.centerforpubliceducation.org/>
- Illinois Department of Education. (2012). Performance evaluation reform act (pera). Retrieved from <http://www.isbe.state.il.us/>
- Illinois General Assembly. (2010). Public Act 096-0861. Retrieved from <http://www.ilga.gov/legislation/publicacts/96/PDF/096-0861.pdf>

- Iwanicki, E. (1990). Teacher evaluation for school improvement. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin Press.
- Jacob, B. (2004). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics*, 89, 761-796.
- Johnson, B. (1999). Great expectations but politics as usual: The rise and fall of a state-level evaluation initiative. *Educational Assessment, Evaluation and Accountability*, 13, 361-381.
- Johnson Jr., B., & Kruse, S. (2009). *Decision making for educational leaders* (pp. 71-124). New York, NY: State University of New York Press.
- Jones, B. D., & Egley, R. J. (2006). Looking through different lenses: Teachers' and administrators' views of accountability. *Phi Delta Kappan*, 87(10), 767-771.
- Kasper, B. (2005). Educational reform 1983-1994: New ideas or rebirth of Quintilian's ideologies. *American Educational History Journal*, 32(2), 175.
- Kelly, A., Orris, J. (2011). Assessing accountability in u.s. public education. *Journal of Public Budgeting, Accounting & Financial Management*, 23 (1), 1-30
- Kimball, S., White, B., & Milanowski, T. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54-78.
- Kimball, S., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.
- Klein, A. (2013, Sept. 15). [Web log message]. Retrieved from http://blog.edweek.org/edweek/campaign-k-12/2013/09/gao_race_to_the_top_states_have
- Kyriakides, L, Demetriou, D. (2007). Introducing a teacher evaluation system based on teacher effectiveness research: An investigation of stakeholders' perceptions. *Journal of Personnel Evaluation in Education*, 20, 43-64.
- Larsen, M. (2005). A critical analysis of teacher evaluation policy trends. *Australian Journal of Education*, 49(2), 292-305.
- Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The Future of Children*, 17(1), 87-109. Retrieved from www.futureofchildren.org
- Lawton, S. (2009). Effective charter schools and charter school systems. *Planning and Changing*, 40(1/2), 3 5-60.

- Learn more about teacher evaluations.* (2012). Retrieved from <http://learnmore.nysut.org/what-the-research-says-on-teacher-evaluations/>
- Lee, J., & Wong, K. (2004). The impact of accountability on racial and socioeconomic equity: considering both school resources and achievement outcomes. *American Educational Research Journal*, Winter(41), 797-832.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Macaluso, T. (2013, Sept. 11). *Rochester teachers bruised by review*. Retrieved from <http://www.rochestercitynewspaper.com>
- Madaus, G., & Mehrens, W. (1990). Conventional tests for licensure. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, California: Corwin Press.
- Mangiante, E. (2010). Teacher matter: measures of teacher effectiveness in low-income minority schools. *Educational Assessment, Evaluation and Accountability*, 23, 41-63.
- Marzano, R. (2012). The two purposes of teacher evaluation. *Educational Leadership*, 70(3), 14-19.
- McCoy, L. (2013, July 16). *Report suggest tweaks to weight of test scores on teacher evaluations*. Retrieved from <http://www.knoxnews/news/2012/jul/16/report-suggest-tweaks-to-weight-of-test-scores-on-teacher-evaluation/>
- McFadyen, D. (2001). The merit pay controversy. *Working USA*, 4(3), 131.
- Mehrens, W. (1990). Combining evaluation data from multiple sources. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin Press.
- Millman, J., & Darling-Hammond, L. (Eds.). (1990). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin Press.
- Morgan, H. (2014, January 24). New evaluation system to cost schools \$1m. Retrieved from <http://mdjonline.com>
- Movement-in-American-Education. Retrieved (January 13, 2012) from <http://education.stateuniversity.com/pages/2445/Standards->
- Natriello, G. (1990). Intended and unintended consequences: purposes and effects of teacher evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin Press.
- National Governors Association Center for Best Practices. (2002). Retrieved August 19, 2011, from <http://www.nga.org/cms/center>

- New assessments for improved accountability. (2011). *The Hamilton Project*. DOI: www.hamiltonproject.org
- New York Department of Education. (2012). Nysed resources and tools to support annual professional performance review (appr) implementation. Retrieved from <http://www.nysed.gov/>
- Newmann, F., King, M., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review*, 67(1),41-74.
- New policy statement on teacher evaluation and accountability. (n.d.). Retrieved June 1, 2013,from <http://www.nea.org/grants/46326.htm>
- Nickerson, N. (1984). Merit pay--Does it work in education? Retrieved August 19, 2011, from <http://bul.sagepub.com/content/68/470/65.citation>
- O'Day, J. (2002). Complexity, accountability, and school improvement. *Harvard Review*, 72, 3, 293-329.
- Office of School Improvement Teacher and Leader Effectiveness Division. (2013, July 22). *Teacher keys effectiveness system*. Retrieved from www.doe.k12.ga.us
- Omer, S. (2012). Questions at the heart of Chicago strike: How do you measure teacher performance. Retrieved September 12, 2012, from: <http://usnews.nbcnews.com>
- O'Malley, K., Murphy, S., McClarty, K., Murphy, D., & McBride, Y. (2011). *Overview of student growth models* [White paper]. Retrieved from <http://www.perasonassessment.com/>
- O'Pry, S., & Schumacher, G. (2012). New teachers' perceptions of a standards-based performance appraisal system. *Educational Assessment, Evaluation and Accountability*, 24, 325-350. doi: 10.1007/s11092-012-9148-4
- Organization of U.S. education: State role i-primary and secondary education*. (2008). Retrieved from <http://www.ed.gov/international/usnei/edlite-index.html>
- Ovando, M., & Ramirez, A.(2007). Principal's instructional leadership within a teacher performance appraisal system: Enhancing student's academic success. *Journal of Personnel Evaluation in Education*, 20, 85-110.
- Painter, S. (2000). Principals' perceptions of barriers to teacher dismissal. *Journal of Personnel Evaluation in Education*, 14(3), 253-264.
- Papay, J. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123-141.
- Paris, D. (1997). School reform: Too much talk, too little action. *Forum for Applied Research and Public Policy*, 12(3), 46-56.

- Peabody-led survey shows teacher evaluation effort viewed positively. (2013, October 09). Retrieved from <http://nashvillepost.com>
- Peters, W. (1994). The empowering process of accelerated schools: An essential link in school reform. *English Journal*, 83(7), 62-65.
- Peterson, K. (1987). Teacher evaluation with multiple and variable line of evidence. *American Educational Research Journal*, 24(2), 311-317.
- Petrarca, J., & Scariano, L. (2012, August). *Changing the teacher evaluation process*. Retrieved from <http://www.nsba.org/leadershipinsider>
- Polnick, B., & Reed, D. (2006). What's really at stake with high-stakes testing? *Advancing Women in Leadership*, 22,
- Popham, J. (2013). *Evaluating America's teachers: Mission possible?*. London, England: Corwin.
- Popham, J., & DeSander, M. (2014, February). Will the courts save teachers? *Educational Leadership*, DOI: www.ASCD.org
- Ravitch, D. (2014, March 31). [Web log message]. Retrieved from <http://dianeravitch.net/2014/03/31/us-doe-struggles-to-prove-that-vam-works-but-it-doesnt/>
- Race to the top program executive summary*. (2009, November). Retrieved from <http://www2.ed.gov/programs/racetothetop/index.html>
- Rebell, M. (1990). Legal issues concerning teacher evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin Press.
- Reform before the storm: A timeline of the Chicago public schools. *Chicago Magazine*. (2012) n. page. Print. <<http://www.chicagomag.com/core/pagetools.php?url=/Chicago-Magazine>>.
- Resnick, L. (2006). Making accountability really count. *Educational Measurement: Issues and Practice*, 25(1), 33-37.
- Ritter, G. W., & Jensen, N. (2010). The delicate task of developing an attractive merit pay plan for teachers. *Phi Delta Kappan*, 91(8), 32-37.
- Rockoff, J., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *American Economic Review & Proceeding*, 100, 261-266.
- Rutkoff, A. (2012, February 24). [Web log message.] Retrieved from <http://blogs.wsj.com>
- Salimaki, A., & Jansen, S. (2009). Perceptions of politics and fairness in merit pay. *Journal of Managerial Psychology*, 25(3), 229-251.

- Sawchuk, S. (2012, July 18). [Web log message]. Retrieved from <http://blogs.edweek.org>
- Sawchuk, S. (2013a). Chicago teachers See value in new evaluations, but eschew test scores. *Teacher Beat*. Education Week's Blogs, 18 09 2013. Web. 23 Sep. 2013. <http://blogs.edweek.org/edweek/teacherbeat/2013/09/chicago_teachers_see_value_in.html?....>
- Sawchuk, S. (2013b, February 5). Teacher's rating still high despite new measures changes to evaluation systems yield only subtle differences. Retrieved from http://www.edweek.org/ew/articles/2013/02/06/20evaluate_ep.h32.html?
- Sawchuk, S. (2013c, June 12). Race on the ready nyc teacher review. Retrieved from www.educationweek.org
- Schelzig, E. (2012, June 11). [Web log message]. Retrieved from <http://blogs.knoxknews.com/humphrey/2012/06/score-report-the-ap-story.html>
- Scherrer, J. (2012, May). What's the value of VAM (value-added modeling)? *Pi Delta Kappan*, 93(8), 58-60.
- Scott, G. United States Government Accountability Office, Office of Public Affairs. (2013). *Race to the top states implementing teacher and principal evaluation systems despite challenges*. Retrieved from website: <http://www.gao.gov/products/GAO-13-777>
- Sharec, C. (2013, December 28). Georgia schools prepare for end of some fed funding. Retrieved from <http://mdjonline.com>
- Siegal, M., & Worth, C. (2000). The impacts of trust and control on faculty reactions to merit pay. *Personnel Review*, 30(6), 646-656.
- Snyder, C. W. (2010). Standards and assessment in education. *Development*, 53(4), 540-546.
- Sporte, S. E., Stevens, W. D., Healey, K., Jiang, J., & Hart, H. (2013, September). Teacher evaluation in practice: Implementing Chicago's REACH students. <http://ccsr.uchicago.edu/publications/teacher-evaluation-practice-implementing-chicagos-reach-students>
- Staiger, D., & Rockoff, J. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, 24(3), 97-118.
- Stark, J., & Lowther, M. (1984). Predictors of teacher preferences concerning their evaluation. *Educational Administration Quarterly*, 20(4), 76-106.
- Stern, G. (2013, Oct 15). *N. y.'s teacher evaluations faulted in study*. Retrieved from <http://www.lohud.com/article/2013/NEWS/310150042/Study-faults-NY-s-teacher-evaluation>

- Stodolsky, S.(1990). Classroom observation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook on teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin Press.
- Stockard, S. (2014, February 20). *Legislators seek changes in evaluations, common core curriculum*. Retrieved from <http://www.murfreesboropost.com/legislators-see-changes-in-teacher-evaluations-common-core-curriculum>
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.
- Strauss, V. (2012a, March 08). Researchers blast Chicago teacher evaluation reform. *Washington Post*. Retrieved from <http://washingtonpost.com/blogs/answer->
- Strauss, V. (2012b, July 10). Georgia professors blast teacher evaluation system. *Washington Post*. Retrieved from <http://washingtonpost.com/blogs/answer->
- Strike, K. (2010). Charter schools, choice, and distributive justice: What evidence do we need? *Theory and Research in Education*, 8(1), 63-78.
- Stronge, J. (2010). Evaluating what good teachers do: Eight research-based standards for assessing teacher excellence. Retrieved from www.eyoneducation.com
- Stronge, J., & Tucker, P. (1999). The politics of teacher evaluation: A case study of new system design and implementation. *Journal of Personnel Evaluation in Education*, 13(4), 339-359.
- Strong, J., & Tucker, P. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. Larchmont, NY: Eye on Education. DOI: www.eyoneducation.com
- Stronge, J., Ward, T., Tucker, P., & Hindman, J. (2008). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education*, 20(1), 165-184.
- Stronge, J., Xianxuan, X., & Leeper, L. (2012). Teacher evaluation and teacher effectiveness using multiple sources of evidence. Retrieved from www.eyoneducation.com
- Student learning objectives. (n.d.). Retrieved April 13, 2013, from <http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Pages/Students>
- Sullivan, J. (2012). A collaborative effort: Peer review and the history of teacher evaluations in Montgomery County, Maryland. *Harvard Educational Review*, 82(1), 142-168.
- Sullivan, K., & Zirkel, P. (1998). The law of teacher evaluation: case law update. *Journal of Personnel Evaluation in Education*, 11, 367-380.
- Summary of Goals 2000: Educate America Act. (1994). Retrieved March 1, 2012, from <http://www.ncrel.org/sdrs/areas/issues/envrnmnt/stw/sw0goals.htm>

- Tarrou, A., Opdal, L., & Holmesland, I. (1999). Improving teacher evaluation through evaluation: an analysis of four Norwegian case studies. *European Journal of Teacher Evaluation*, 22(2/3), 135-157.
- Teacher and principal practice rubrics. (2013, September 30). Retrieved from NYSED.gov.
- Teachers call moratorium new teacher evaluation system. (2013). *Amsterdam News*. Retrieved from <http://amsterdammnews.com/news/2013/sep/12>
- Teacher development and evaluation*. (n.d.). Retrieved April 10, 2012, from <http://www.aft.org/issues/teaching/evaluation.cfm>
- Teacher keys effectiveness system implementation handbook 2013-*. (2013). Retrieved from <http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/TKES Handbook FINAL 7-18-2013.pdf>
- Tennessee Department of Education. (2012). Teacher evaluation in Tennessee: A report on year 1 implementation. Retrieved from <http://www.tn.gov/education/>
- Tennessee Department of Education, (2013). *Tennessee first to the top educators guide: New teacher and principal evaluation*. Knoxville, TN: Author.
- Tennessee First to the top*. (2013). Retrieved from <http://www.tn.gov/firsttothetop/about-legislation.html>
- The Savannah Morning News. (2014, January 02). Race to the top: Beware of strings that bind. Retrieved from <http://mdjonline.com>.
- Tucker, P., & Stronge, J. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- University of Wisconsin at Madison. Retrieved October 4, 2010, from <http://cpre.wceruw.org/paper/CoventryAERA04.pdf>
- U. S. Department of Education. (n.d.). Federal education policy and the states, 1945-2009. Retrieved May 23, 2012, http://www.archives.nysed.gov/edpolicy/research/res_essay_bush_ghw_edsummit.shtml
- U. S. Department of Education, (2012a). Esea flexibility. Retrieved from <http://ww2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>
- U. S. Department of Education. (2012b). National education leaders release shared vision for the next generation of teaching during the 2012 labor management conference. Retrieved May 23, 2012, <http://www.ed.gov/news/press-releases/national-education-leaders-release-shared-vision-next-generation-teaching-during>
- US Department of Education. (2012c, February 15). *Race to the top fund*. Retrieved from <http://www2.ed.gov/programs/racetothetop/phase1-applications/index.html>

- US Department of Education. (2012d, February 15). *Race to the top fund*. Retrieved from <http://www2.ed.gov/programs/racetothetop/phase2-applications/index.html>
- U. S. Department of Education, (2013). *Race to the top: Georgia year 2 report: School year 2011-2012*. Washington, DC: Author.
- VanderHoff, J. (2008). Parental valuation of charter schools and student performance. *Cato Journal*, 28(3), 479-493.
- Varlas, L. (2012). Rethinking teacher evaluation: Leaders advocate for more meaningful measures. *Education*, 54(12), 1-6.
- Vevea, B. (2013, 09 18). *New teacher rating not dramatically different*. Retrieved from <http://www.wbez.org/print/108703>
- What have we learned from APPR? (2013, June 27). Retrieved from <http://ocmbocesis.wordpress.com/2013/06/27/what-have-we-learned-from-appr/>
- White, B. (2004). *The relationship between teacher evaluation scores and student achievement: evidence from Coventry, RI*. Working paper series TC-04-04. Madison, WS: Consortium for Policy Research in Education (CPRE-UW),
- William, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 4(2), 107-122.
- Zepeda, S., & Kruskamp, B. (2012). Teacher evaluation and teacher effectiveness: classroom observations. Retrieved from www.eyeducation.com
- Zhang, X. (2002). Exploring the impact and process of merit pay for public school teachers. (Doctoral dissertation).
- Zirkel, P. (2010, September). Teacher tenure is not the real problem. *Phi Delta Kappan*, 92(1), 76-77.
- 23 Illinois administrative code 50. (n.d.). Retrieved May 23, 2013, from <http://www.isbe.net/rules/archive/pdfs/50ARK.pdf>