Spring 5-2017

Mississippi's Teacher Observation Rubric: Administrator Perceptions of Appropriateness by Grade Level

Danette Irvine Moore
University of Southern Mississippi

MISSISSIPPI'S TEACHER OBSERVATION RUBRIC: ADMINISTRATOR PERCEPTIONS OF APPROPRIATENESS BY GRADE LEVEL

by

Danette Irvine Moore

A Dissertation
Submitted to the Graduate School
and the Department of Educational Research and Administration
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved:

_____
Dr. Ann E. Blankenship, Committee Chair
Assistant Professor, Educational Research and Administration

_____
Dr. Stanley C. Benigno, Committee Member
Adjunct Professor, Educational Research and Administration

_____
Dr. David E. Lee, Committee Member
Associate Professor, Educational Research and Administration

_____
Dr. Kyna J. Shelley, Committee Member
Professor, Educational Research and Administration

_____
Dr. Lilian Hill
Department Co-Chair, Educational Research and Administration

_____
Dr. Karen S. Coats
Dean of the Graduate School

May 2016

Published by the Graduate School

ABSTRACT

MISSISSIPPI'S TEACHER OBSERVATION RUBRIC: ADMINISTRATOR PERCEPTIONS OF APPROPRIATENESS BY GRADE LEVEL

by Danette Irvine Moore

May 2016

The focus of this study was to measure elementary, middle, and high school administrators' beliefs regarding the appropriateness of the Mississippi Statewide Teacher Appraisal Rubric domains, as well as their perceptions regarding the Mississippi Statewide Teacher Appraisal Rubric's overall alignment with the Interstate School Leaders Licensure Consortium standards. This was a quantitative study that investigated whether a statistically significant difference existed between administrators' beliefs regarding the appropriateness based on their grade level assignment. A 48-statement questionnaire was developed using the current Mississippi Statewide Teacher Appraisal Rubric domains and standards to obtain the quantitative data. A five-point scale ranging from Strongly Agree to Strongly Disagree was distributed.

To test the hypotheses of this study, descriptive statistics were analyzed, and an analysis of variance (ANOVA) was utilized to determine statistical significance. The results of this study did reveal overall administrators' perceptions that the Mississippi Statewide Teacher Appraisal Rubric was aligned with Interstate School Leaders Licensure Consortium standards. It did not reveal a statistically significant difference in the administrators' perceptions of appropriateness of the Mississippi Statewide Teacher Appraisal Rubric domains based on their grade level assignment.

ACKNOWLEDGMENTS

The researcher would like to thank the committee chairperson, Dr. Ann Blankenship, for her encouragement and guidance as a professor, advisor, and mentor. Dr. Blankenship held high expectations for scholarly research and helped me to reach those standards.

I would also like to thank Dr. Kyna Shelley for her encouragement and support regarding the statistical elements of the study. She helped me to speak the language of numbers fluently.

Additionally, I would like to thank Dr. David Lee and Dr. Stanley Benigno for agreeing to serve on my committee and believing that I could finish this process. The confidence of my entire committee helped me to persevere despite a deep personal loss during the dissertation process.

DEDICATION

I would like to thank my husband, David Moore, my mom and dad, Judy and Donald Fleming, as well as my daughters, Earlie and Eimhear Davis for believing that I could accomplish this feat. Completion of this dissertation would not have been possible without the daily support and encouragement David and Eimhear, who understood when I was locked away writing or in class at the university. Thank you for your love and understanding.

I would also like to thank my grandmother, Earline Kendrick who impressed upon me the importance of education through her words and her example. She taught me that education, especially for women, is the key to independence both intellectually and practically. She demonstrated this when she went to night school at 48 years old to earn her high school diploma.

Finally, I would like to dedicate this work to my mother, Stephanie "Judy" Fleming. My mom taught me how to teach and reach the hearts of children through her daily example. She was a great support to me through the dissertation process-- from helping me to stuff envelopes when mailing out the survey instrument to speaking encouraging words when I wanted to procrastinate my work. Mama always doubted her own intelligence, creativity, and tenacity, but she never doubted mine. She never attended college, but she was one of the wisest people that I have ever known. Thank you, Mom.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I - INTRODUCTION

Overview

Most clockmakers are interested in precision. In every clock, from the famous Big Ben in London to the keepsake mantle clock that has been in some families for generations, each one takes many gears working together to measure out the success of the entire timepiece. If one widget's teeth begin to wear, it is simply replaced with another part machined to its exact specifications, and the work of the clock continues. Likewise, some research has shown that for many years school administrators seemed to view teachers like these gears in a larger machine (Levin, Mulhern, & Schunck, 2005). Although such an attitude seems to deny the impact of individual teachers, this was precisely the conclusion reached in *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness* (Weisberg, Sexton, & Mulhern, & Keeling, 2009). According to the authors, "94 percent of teachers receive one of the top two ratings and less than 1 percent are rated unsatisfactory" (p. 6). Their study found that the majority of teachers were rated *good* or *great* despite the limited success of their students. For these reasons, they concluded that principals viewed teachers as equally effective and interchangeable like widgets or cogs in a machine.

To differentiate among effective and ineffective educators, many states, including Mississippi, have moved toward more objective data-driven teacher evaluation systems. In fact, implementation of such a framework was a prerequisite for receiving federal education funds through Race to the Top grants (Harris, Ingle, & Rutledge, 2014). These evaluation models represent a significant shift in how teachers are observed and assessed.

Teacher Observation Rubrics

Framework for Teaching. Like many other states, Mississippi adopted some form of Framework for Teaching pioneered by Charlotte Danielson (Mississippi Department of Education, 2014a). The framework includes 4 domains, 22 components, and 76 smaller elements. The four domains are Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities (The Danielson Group, 2013). In the Planning and Preparation domain, teachers should demonstrate knowledge of subject matter and learning dynamics while designing both lessons and assessments. The second domain, Classroom Environment, again places the teacher in the role of designer by charging him or her with creating a culture of learning and an inviting physical space. The teacher must also demonstrate strong management skills in this domain in order to channel student behavior in the direction of learning. The Instruction domain assesses the teachers' interpersonal skills in effective communication to engage the students and check for understanding. Because teaching can be a rather isolated activity, domain four, Professional Responsibilities, requires that the teacher communicate with other professionals to learn new techniques and strategies. It also assesses the teachers on communicating with stakeholders and parents. In addition to the observation component, teachers in Mississippi and other states are also assessed on school performance, their own students' performance, or a combination of both in what is termed a "value-added model" (Harris et al., 2014).

M-STAR. Mississippi's adaptation of the *Framework for Teaching* is called Mississippi Statewide Teacher Appraisal Rubric (M-STAR). It is part of the Mississippi Teacher Evaluation System (MTES). M-STAR contains 20 standards divided into five

domains: Planning, Assessment, Instruction, Learning Environment, and Professional Responsibilities. Half of the standards are focused in the Instruction and Learning Environment domains. These two domains are the only ones assessed through classroom observation. The other three domains, Planning, Assessment, and Professional Responsibilities are assessed through the collection of artifacts and the pre- and post-observational conferences. Mississippi has also created an optional student survey, which may be used to inform the evaluator about assessment practices, instruction, and classroom climate (Mississippi Department of Education, 2014a). Although this survey's language and format is appropriate for junior high and high school age children, the process manual indicates, "School districts may create their own survey or use one that is appropriate" (Mississippi Department of Education, 2014a, p. 5).

Although the development of the teacher evaluation rubric was aimed at differentiating the effectiveness of educators, the rubric is identical regardless of grade level, educational ability level, or course format. For example, an elementary teacher is evaluated using the same rubric as an Advanced Placement Physics teacher. A drama teacher's effectiveness is judged according to the same criteria as a math teacher. Special education teachers are expected to display the same teaching strategies as any other teacher in the system according to M-STAR. It seems that in an effort to stop viewing teachers as interchangeable widgets, this rubric has further legitimized the concept of one-size-fits-all educational strategies by evaluating all teachers based on the same standard. This dissertation was designed to study the variations in perceptions of appropriateness among administrators at the elementary, middle, and secondary levels concerning M-STAR domains. The research was designed to isolate any areas that may

need differentiation based on the level of schooling. In addition to surveying participants regarding the M-STAR domain appropriateness, the questionnaire also assessed the administrators' perceived level of appropriateness of M-STAR's alignment to the School Leaders Licensure Consortium (ISLLC) standards. Because administrators are guided by ISLLC standards in their own practice, this research represents a vital point of cross-referencing policy measures with accepted standards of practice for school leaders.

Developmental Differences in Students

Even generic lists of educational best practices differ among the ages and stages of children. For example, the Alliance for Childhood (2002) suggested that child-initiated play is a building block of early-childhood education. Further, they report, "The rough and tumble of active play facilitates children's sensorimotor development" (p. 1). This stage may last up to age eight depending on the individual characteristics of the child. Adolescent learners, described by the National Middle School Association as students who are between ten and fifteen years old, need lessons that incorporate the senses and emotions (Wilson, Horch, & Wilson, 2002), as well as those incorporating movement and exercise. During this time, adolescents' bones are hardening, especially the tailbone, which makes it difficult to sit for hours. On the other hand, many of the sixteen, seventeen and eighteen-year-olds in high school, have transitioned to become adult learners. Knowles, Holton, and Swanson (2005) indicate that although adult learners enjoy engaging lessons and sharing their own experience, adults need immediately relevant and useful information. They are capable of comprehending abstract concepts and inferring deeper meaning. Clearly, students in all age brackets face different challenges and thrive with different types of learning opportunities at various

stages of development. However, many of the teacher observation rubrics or frameworks adopted by states in order to assess effective classroom instruction are the same for teachers of children ages four through eighteen (Danielson & McGreal, 2000; Marzano & Toth, 2013; Marshall, 2013). Likewise, the Mississippi Statewide Teacher Appraisal Rubric is identical for all levels of education.

Instructional Leadership and M-STAR

The implementation of the rubric and the judgment of teacher effectiveness is most often the responsibility of school administrators (Weisberg et al., 2009). Ingle, Rutledge, and Bishop (2011) point out that traditionally, it is the principal who observes individual teachers and assesses their effectiveness. However, when using M-STAR, not all principals believe it represents achievable standards (Moore, 2014). For example, when referencing the highest range of scores, some administrators use terms like "visiting" a four or "floating" up to a four occasionally (Moore, 2014, p. 14). Comments such as these make it very clear that the expectation of a perfect score would be unreasonable. In fact, one administrator alluded to a possible conflict of interest between administrators and teachers regarding the scoring. Dr. Robinson commented, "I don't think the state is discouraging you from doing that [giving a four], but when you assign a teacher a rating, you have to quantify and qualify what makes this teacher effective or distinguished" (Moore, 2014, p. 15). From her comment, it seems that administrators are scrutinized if they award the highest scores. Given the historically high ratings of most teachers uncovered in *The Widget Effect* (Weisburg et al., 2009), it appears that school districts are steering clear of this perception. However, the avoidance of high scores still makes the process disingenuous. One of the items on the research instrument in this

dissertation specifically asked administrators if a perfect score was achievable in each of the various M-STAR domains. Since this research explored the perceived levels of appropriateness of administrators from different schooling levels, their responses indicated any differences in the appropriateness of the rubric for various educational contexts.

Theoretical Framework

In addition to indicating a possible bias toward awarding lower scores, the statements of administrators in the Moore (2014) interview reveal competing interests in teacher evaluations, which set stakeholders in opposition to one another. This dynamic leads to "agendas of power being contested" (Cranston, 2013; Educational Broadcasting Corporation, 2004). Horng and Loebb (2010) describe the traditional view of instructional leaders as those who view themselves "as "hands-on" leaders, engaged with curriculum and instruction issues, unafraid to work directly with teachers, and often present in classrooms" (66). Because instructional leaders are characterized as exhibiting "strong, directive leadership focused on curriculum and instruction" (Hallinger, 2003, p. 329), as well as data-focused and ultimately accountable for student achievement (Hallinger, 2005), these leaders bring an involved, direct-impact perspective to teacher evaluation. Therefore, their perceived levels of appropriateness regarding the individual domains of M-STAR and the alignment with ISLLC standards could be very compelling.

As well as employing the identical rubric for all teachers regardless of context, this rubric represents a single learning theory. The Danielson Framework for Teaching is rooted in the constructivist theory of learning (The Danielson Group, 2013) and has become somewhat institutionalized as the official method of evaluating teachers. This

construct affects teachers in the broader political sense by defining knowledge and beliefs on how learning is created (Educational Broadcasting Corporation, 2004). Further, the diction surrounding the observation rubric constitutes a type of epistemological power because it is based on one learning theory.

If the evaluation tool only values teaching strategies from one theoretical perspective, then teachers can be penalized for making decisions to use other strategies. Coding the diction used in the rubric for teacher evaluations revealed the pedagogical priorities (Moore, 2014). For example, when analyzing the verb choices in the rubric, one will see words like *supports*, *engages*, and *facilitates*. These teacher behaviors indicate a student-centered approach, which is aimed toward learning independence. Student-centered pedagogy has been the focus of much educational reform for the past 20 years (Estes, 2004; Hopkins, 1994; Lord, 1999; Simpson, 2002). It stands in contrast to a teacher-centered or transmissive approach. Constructivist theory underlies student-centered teaching, and its discourse represents the idea that students create knowledge through interaction and reflection (Educational Broadcasting Corporation, 2004; Mascolo & Fischer, 2004). Likewise, student-centered discourse emphasizes that a teacher should become a coach or facilitator rather than an expert in the subject (Educational Broadcasting Corporation, 2004). In a recent interview conducted of three high school principals, each one characterized the distinguished teacher as a "facilitator" (Moore, 2014). Many presenters in professional development sessions almost vilify a teacher-centered approach with the admonition against being the "sage on the stage." This discourse casts teaching and learning in one type of theoretical framework. Studying the administrators' perceived level of appropriateness within various domains of the rubric

indicated areas for further study regarding the use of a single underlying learning theory in the evaluation rubric.

Because the rubric used for evaluating teachers is biased toward certain learning strategies, even teachers who contribute to their final evaluation through self-reflection are subject to an epistemological power imbalance (Towndrow & Tan, 2009). Epistemology seeks to define the nature of knowledge and how learning happens. There are multiple competing and complimentary theories on this topic (Steup, 2005). In modern education reform, constructivism is the dominant theory (Simpson, 2002). The essence of constructivism is that people build their own knowledge through the lens of their individual and collective experiences (Moscolo & Fischer, 2004). When following this line of reasoning, it follows that in said paradigm, knowledge is created rather than discovered. It disavows an objective ontological reality apart from individual interpretation (Simpson, 2002). This is the instructional shift that requires teachers to become guides, coaches, or facilitators rather than experts and authorities of a particular subject. Towndrow and Tan (2009) found that self-evaluation in this context can actually disempower teachers because of the restricted paradigm in which they are operating. The epistemological power has already been defined by policy makers and others who adopted the assessment framework. Again, researching the instructional leaders' perceived level of appropriateness with M-STAR domains helped to add the local administrators' voice to the discussion on epistemology.

The Danielson Framework for Teaching is rooted in the constructivist theory of learning (Danielson & McGreal, 2000) and when used in isolation, can devalue other theories of learning and pedagogical practices (Simpson, 2002). Further, some principals

often most value characteristics not even mentioned on the evaluation rubric (Harris et al., 2014). For example, in one study where principals rated characteristics of effective teachers, the one characteristic rated most highly was a caring demeanor (Harris et al., 2014), yet this affective characteristic is one of the 77 small characteristics within one of the 22 components under one of the domains in the Danielson rubric. Frameworks like this one create procedures that can become like an orthodoxy centering power outside of the educational institutions (Cranston, 2013; Papa, 2011). Researching the administrators' perceived level of appropriateness for M-STAR may help to empower educational professionals by adding their voice to the conversation regarding teacher evaluation and the uniqueness of their particular context.

Statement of the Problem

Although differentiation among educators is one goal of teacher evaluation, surprisingly the identical rubric is used to evaluate teachers regardless of grade level, ability level, or subject taught in Mississippi. Further, the M-STAR observation instrument has never been assessed for its alignment with professional practice standards for school administrators. This research challenged the one-size-fits-all rubric implemented in Mississippi on its ability to truly differentiate among educators and cross-checked this rubric with administrators' perceptions of alignment with professional practice standards.

The research in this study attempted to uncover any needed differentiation in the evaluation criteria for teachers according to their level of school placement, as well as any areas of divergence with the ISLLC standards based on the administrators' perceived levels of appropriateness for these components. In order to identify these possible

problem areas, the researcher surveyed Mississippi school administrators at the elementary, middle and high school levels regarding their perceived levels of appropriateness of the individual M-STAR domains and their perceptions regarding M-STAR's alignment to the ISLLC standards. In analyzing the data, the researcher specifically looked for differences among administrators according to their school grade levels, elementary, middle, or high school.

Research Questions

This study was designed to answer the following questions:

1. To what degree do elementary school administrators believe the overall M-STAR evaluation tool, its various domains, and its alignment to ISLLC standards are appropriate for teachers at the elementary level?
2. To what degree do middle school administrators believe the overall M-STAR evaluation tool, its various domains, and its alignment to ISLLC standards are appropriate for teachers at the middle school level?
3. To what degree do high school administrators believe the overall M-STAR evaluation tool, its various domains, and its alignment to ISLLC standards are appropriate for teachers at the high school level?
4. How do perceived levels of appropriateness between elementary, middle, and high school administrators differ?

Definition of Terms

The terms used within this study have been defined by the researcher or within the context of the literature.

M-STAR-- Although the acronym indicates the Mississippi Statewide Teacher Appraisal Rubric, the term can encompass an entire system of processes and tools used to evaluate teachers in Mississippi (Mississippi Department of Education, 2014a). However, for the purposes of this study, M-STAR refers to the actual rubric used when administrators observe teacher practice for the purpose of evaluation.

MTES-- Mississippi Statewide Teacher Evaluation System is the terminology regarding the entire evaluation system including value-added test scores (Mississippi Department of Education, 2014c).

ISLLC Standards-- The Interstate School Leaders Licensure Consortium (ISLLC) Standards as developed by the Council of Chief State School Officers are the basis for school administrator licensure in the state of Mississippi (Mississippi Department of Education, 2012). At the time of this study, these six standards guided best practices for school administrators when interacting with faculty, students, and other stakeholders (Council of Chief State School Officers, 2008).

Delimitations

The delimitations of this study involved the limited context. First, this study was restricted to schools in Mississippi that were using the M-STAR instrument to evaluate teacher performance. Although non-administrators may evaluate teachers in Mississippi, the study surveyed administrators only. All evaluators had received training regarding the evaluation system; however, responses were limited by their actual experience and familiarity with the rubric.

Assumptions

This study included several assumptions regarding the integrity of respondents. First, it was assumed that the participants would record their honest answers to the questions on the survey. Further, it was assumed that they were familiar with the M-STAR instrument because of training provided by the state of Mississippi to all administrators. The degree of familiarity differed according to their years of experience using the instrument.

Justification

This study examined how M-STAR, a teacher evaluation rubric based on the Danielson Framework, was functioning as a tool for differentiating teachers in varied educational contexts. Since the same rubric is used for teachers in elementary, middle, and high schools, this study was designed to indicate areas of varied perceptions of appropriateness from instructional leaders. Since the instructional leaders of these schools were largely responsible for assessing teacher performance, it made sense that their beliefs regarding appropriateness regarding the various M-STAR domains would indicate areas for improvement. This study was the first on this topic and added to the body of literature on teacher evaluations.

Additionally, the survey instrument assessed school leaders' perceptions regarding M-STAR's alignment with the ISLLC standards. Such an alignment had not been studied previously, yet it was very relevant to the discussion of teacher evaluation. Since administrators are guided in their practice by the ISLLC standards, the tools mandated by the state department of education should align to those standards. This study provided a cross-reference between the standards for administrators and the

practice of administrators. The results of this research indicated areas for revision and future research in both the practice of school leaders and the differentiation of M-STAR based on educational context

CHAPTER II – LITERATURE REVIEW

Introduction

Teachers make a difference. Research consistently highlights evidence suggesting students with highly effective teachers not only perform better on the current year's standardized test, but also build momentum that impacts achievement for at least three years (Jordan, Mendro, & Weerasinghe, 1997; Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997). Conversely, students with ineffective teachers show a loss of achievement in the current year and improvement resistance even when placed with an effective teacher in subsequent years (Jordan et al., 1997; Sanders & Rivers, 1996; Wright et al., 1997). Compelling findings such as these contributed to the educational context surrounding the landmark study, *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness (*Weisberg et al., 2009), which revealed that despite evidence to the contrary, administrators tended to evaluate teachers as if they were equally effective.

Although accurately evaluating teacher effectiveness is challenging, it has become the focus of much educational reform. Educating children is a collective effort by teachers, in concert with parents, administrators, and even district and state personnel (Tucker & Stronge, 2005). Likewise, student achievement depends on many variables within the classroom. Student learning may be impacted by attendance, resources, class size, etc. (Tucker & Stronge, 2005). Challenges in accurately measuring student learning, as well as the collective and conditional nature of learning (Schalock, 1998), makes teacher evaluation a complex endeavor. Despite these challenges, teacher accountability has become a cornerstone of modern educational reform (Guilfoyle, 2013;

McGuinn, 2012). This review of the literature will present conceptual and theoretical frameworks for teacher evaluation, a history of teacher evaluation, a review of teacher evaluation systems, as well as literature regarding teacher effectiveness.

Conceptual Framework

Epistemological Power

Whether on T-shirts or billboards, many have seen the popular axiom "Knowledge is Power." Nowhere is this statement truer than in teacher evaluations. The ways in which knowledge is constructed truly situates the nexus of power when evaluating teacher performance with those who define how knowledge should be transmitted. Epistemology is the broad philosophical category that, at its most basic level is the study of knowledge and understanding. However, the underlying issues include how knowledge is constructed, what are legitimate sources of knowledge, and what are the limits of understanding (Steup, 2005). Educators have long debated these questions, as seen with the waxing and waning popularity of programs like Whole Language, the Magic Circle, and Glaser's Educational Systems.

However, in the current accountability movement, teachers are observed and evaluated using a rubric that defines legitimate teaching strategies and pedagogy. This rubric was vetted and adopted in Mississippi by the Mississippi Department of Education for use statewide. The Mississippi Department of Education represents a powerful entity that has the ability to define effective teaching at every level and demographic in the state. Bocock (1986) explains that when the leadership of the ruling class achieves power by manipulating the overarching outlook or worldview of a society, hegemony is created. According to Towndrow and Tan (2009), the ability to define how knowledge is

developed in students represents an intellectual and philosophical hegemony by establishing epistemological power. As Gramsci (1971) pointed out, hegemonies maintain control through the consent of members of the group and their acceptance of the underlying assumptions it proposed. From this conceptual framework, teachers and administrators are disempowered because they lack the opportunity to develop or revise the teacher observation rubric for their specific educational setting. Even when offered the opportunity to self-evaluate, teachers are disempowered because they must reflect on their practice from a rubric that may not reflect the needs of their specific students (Towndrow & Tan, 2009). This framework will be discussed in detail along with the challenges to the modern evaluation reform in upcoming sections.

Theoretical Framework

Constructivist Learning Theory

Constructivist Learning Theory is a dominant theory in modern education and reflects specific assumptions about how knowledge is developed (Danielson & McGreal, 2000). Prior to the 1980's, most educators based their practice in the theories of behaviorist like B.F. Skinner who proposed that children were like a *tabula rasa*, ready to be written upon (1954). Although the term *tabula rasa* traces back to Aristotle, Skinner used it anchor his own theory of behavior. This philosophy of learning depends on shaping behavior through stimulus and response (Skinner, 1954; Woolfolk, 2006). It does not account for individual differences or any internal paradigms; it essentially constitutes a form of environmental determinism.

Although, he had been publishing for 30 years, by the 1960's American universities began to recognize the work of Jean Piaget who directly challenged this

approach (Campbell, 2006). Piaget (1954) saw himself as a genetic epistemologist because he focused on how children processed knowledge and developed understanding according to physically determined development stages. He articulated four distinct stages: sensorimotor, preoperational, concrete operations, and formal operations (Piaget, 1954; Woolfolk, 2006). The first stage may span two years, the second may span up to five years while the third stage was theorized to occur within four years. The final stage continued into adulthood. Each stage was associated with different cognitive structures that he called *schemes*. Piaget believed that the developmental stage must be reached before learning associated with that stage could occur (Piaget, 1954; Woolfolk, 2006).

Piaget was especially interested in how sensory-motor behaviors influenced learning. He was a clear proponent of learning-by-doing and focusing on process rather than product (Hopkins, 2011). In fact, Piaget (1954) criticized trying to accelerate teaching certain concepts prior to children reaching particular developmental milestones. However, he did acknowledge the struggle of learning new ideas even when a child has reached the appropriate developmental milestone. Piaget (1954) believed that humans either assimilate new knowledge based on previous knowledge or accommodate new knowledge by changing previously held assumptions. Either assimilation or accommodation may involve frustration and struggle; he described this phenomenon as disequilibrium. According to Piaget, when the new information had been processed and reconciled with one's previous 'scheme' or concept of reality, equilibrium was achieved (Piaget, 1954). This idea of schemes could be viewed as the beginning of constructivism in that children build new understanding by reconciling it with previous knowledge.

Although he acknowledged a developmental impact of peers and adults, Piaget believed this impact to be mainly social (Woolfolk, 2006).

Like Piaget, Vytgosky (1978) believed that knowledge was internalized to form new understanding; however, the researchers differed in important ways. Unlike Piaget, Vytgosky theorized that learning could occur prior to reaching a certain developmental stage and actually assist a child in progressing to a new developmental level. Further, he believed that learning was influenced by social interactions (Vytgosky, 1978), an important dimension of constructivism. In his Socio-Cultural Theory of Development, Vytgosky proposed that children learn from their culture and social interactions. Not only do they acquire knowledge from this cultural context, this setting is essential for making meaning of new information (Vygotsky, 1978; Woolfolk, 2006). He later hypothesized that interaction with adults or peers may help students to reach their next stage of development in his 'Zones of Proximal Development' (Vygotsky, 1978). His research in this area began to develop the concept of scaffolding or presenting students with problems slightly above their ability to help them "reach" for the next developmental level (Woolfolk, 2006), another key component of constructivism. Vytgosky, therefore, added to the growing momentum toward Constructivist Learning Theory by exploring the interaction of community and context with the internal cognitive structures proposed by Piaget.

Finally, Jerome Bruner articulated this emerging school of thought and learning theory in The Process of Education (1960). In this landmark book, he explored Cognitive Constructivism as a theory of learning, which did rest on developmental stages, but unlike Piaget's model, Bruner's stages were asynchronous. Like Piaget (1954) and

Vygotsky (1978), Bruner (1960) saw learning as active and building on previous knowledge. He also believed, like Vygotsky (1978), that learning occurred through interaction with others, and teachers could assist students in reaching new levels of development. Bruner's constructivist model represented a student-centered classroom, ample opportunity for collaboration, and personal interpretation of information based on one's own experience. He also emphasized a spiral curriculum that continually revisited basic concepts until students fully internalized the understanding (Bruner, 1960).

Bruner (1960) believed that human cognition essentially rests on the idea of categorizing knowledge much like the schemes proposed by Piaget. However, he developed four themes that have influenced much educational theory. First, he proposed that children are predisposed to learn through exploration. He emphasized the importance of caregivers and teachers to encourage and direct explorations. He believed that problem-solving ability emerges from this natural curiosity. Next, he valued the structure of knowledge. Bruner (1960) indicated that children should be exposed to learning in a specific contextual relationship rather than discrete facts. He also began the discussion of differing modalities of expression to reach different learners, which were later expanded upon by Gardner (1983) who identified distinct learning styles. Finally, Bruner emphasized the cognitive mapping ability of children by categorizing new information. Bruner even extrapolated a hierarchy of categories that indicated levels of learning and hinted at the work of psychologist Benjamin Bloom and his cognitive taxonomy (Bloom, Engelhart, Frost, Hill, & Krathwohl, 1956). Based on his theory, Burner (1960) articulated several characteristics of the constructivist classroom: the teacher should personalize instruction, content should be structured with categorization in

mind and sequenced appropriately, and reinforcement should be used for positive motivation.

Although Constructivist Learning Theory does take into account the needs of the individual and the role of cultural context in learning, it has been challenged for its lack of objectivity. For example, Simpson (2002) asserts that taken to its logical ends, constructivism denies the existence of an objective truth. He writes, "…no statement can be taken as true beyond reasonable doubt, and like existentialism, is open to many interpretations" (p. 2). In recent years, scholars have described it as the "secular religion" (Phillips, 1995, p. 5) of modern educational theory. Like any religion, there are different sects under the constructivist umbrella, but Phillips argues that to assume this theory is simply a philosophical debate on whether knowledge is created or discovered would be naïve. He asserts that there are larger social and political concerns underlying the argument. Again, this debate leads back to the power of epistemology. If all knowledge is created in context, then, as Simpson (2002) points out many of mankind's greatest discoveries would have never happened because they defy personal observation like the idea of Earth not being the center of the universe or flat. Despite the limitations of Constructivist Learning Theory, modern teacher observation rubrics are rooted in its philosophy (The Danielson Group, 2013). Since teachers at all levels are evaluated using constructivist rubrics, the implication is this learning theory is equally appropriate for children of all ages, from four to eighteen. Further, it also implies that teachers at all levels agree to use constructivist strategies as best practices. As will be discussed in the Effective Teaching section below, there is no such consensus among educators regarding one best approach for all children; rather, children and teachers engage in an organic

process of learning, which demands a variety of learning frameworks to best meet the needs of children.

Human Capital Theory

Human Capital Theory is an economic theory that assumes that people and their capacities are essentially the building blocks of the global economy. This is not a new idea-- Sir Walter Petty (1899) and Adam Smith (1937) both connected the state's economy with the available labor pool. Modern interpretations of this theory assume that all human activity is based on economic self-interest and that education is essential to increasing productivity. Theodore Schultz, a Noble prize- winning economist coined the term *human capital* and linked an increase in national income to investment in education (1961). Papa (2011) acknowledged that there are many economic interests within the business of education from human capital to private profiteering. The myriad of standards and frameworks that are used in education today were created to make students ready for college and careers. These standards were essentially created to ensure human capital for the country's economy. The U.S. Secretary of Education, Arne Duncan, commented at the release of the 2012 Program for International Student Assessment, "In a knowledge-based, global economy, where education is more important than ever before, both to individual success and collective prosperity, our students are basically losing ground. We're running in place, as other high-performing countries start to lap us" (U.S. Department of Education, 2013, para. 12).

The concept of human capital in the field of education applies to not only the students but the educators as well. The New Teacher Project (2009) showed that in order to increase the achievement of students and the quality of human capital produced,

attention should be focused on developing the human capital of educators. In this study of the Cincinnati Public Schools, researchers found that teachers were not differentiated through a rigorous evaluation process, supported through professional development to improve, or compensated for outstanding performance. As a result, the recommendations included, "A district-wide human capital strategy centered on teacher effectiveness that produces improved student learning outcomes" (p. 10).

The Federal and State Action Theory proposed by the Mississippi Department of Education in a 2014 presentation regarding their teacher evaluation system echo a similar sentiment. Leaders stated in the presentation that an improved teacher evaluation system would lead to improved educator quality and ultimately result in improved student outcomes (Mississippi Department of Education, 2014a). It is these underlying assumptions regarding human capital that have led to the current Mississippi Statewide Teacher Appraisal Rubric (M-STAR).

History of Teacher Evaluation

In the early 19th century, schooling was viewed as an opportunity for those who were capable, not a right for all (Ravitch, 2002). Teachers were often required to pass a certification exam and vetting by the local school board and religious leaders, but once hired they were assumed to be competent. Teachers taught the material using direct instruction and tested students to see if they had learned the material. If a student failed the test, it was viewed as entirely the student's responsibility (Ravitch, 2002).

However, during the 20th century, public education began to change as it was influenced by the newly emerging field of educational psychology. Edward L. Thorndike was an educational psychologist who influenced modern education to a similar degree as

the well-known educational theorist, John Dewey. Thorndike wanted to move psychology away from the field of philosophy and more toward hard science (State University, 2015). He wrote a definitive three-volume manual titled *Educational Psychology* and the first textbook on social statistics, *An Introduction to the Theory of Mental and Social Measurements.* He is well-known for his emphasis on tests and measurement evidenced by the fact that he taught the first university class on educational testing. Thorndike hoped to move educational outcomes into quantifiable terms so that they could be replicated and reinforced through habit (Ravitch, 2002; State University, 2015). Although Thorndike was not interested in testing for the purposes of accountability, rather professional efficacy, he instigated the shift toward student achievement tests (Ravitch, 2002).

During this same period, the philosophy of progressive education began to take hold. With rapid industrialization and a growing divide between the rich and poor, educational theorists like John Dewey and Margaret Naumberg pushed for a more child responsive educational system (University of Vermont, 2002). This approach was termed "progressive" because it tended to see education as a right of all children rather than the privilege of those who could navigate the coursework. These progressive educators were heavily influenced by the work of psychologists and came to see school as a place to nurture future citizens; as a result, they encouraged social promotion of students when necessary (State University, 2015). Although this practice was partially motivated by the Great Depression and the need to keep children out of the burdened job market, it was also motivated by the idea that it would harm children's sense of self if they were held

back-- as social promotion increased the level of academic rigor decreased (Ravitch, 2002).

In October of 1957, Americans made drastic changes in the educational system. It was during the height of the Cold War, and Russia had just successfully launched the first unmanned satellite, Sputnik, into orbit. Experts agree this was a moment that focused the country on the need for increased educational rigor as a means of national defense (Johanningmeier, 2010; Powell, 2007). As a result, one year later Congress passed the National Defense Education Act (NDEA). According to the U.S. Department of Education (2012), "…the NDEA included support for loans to college students, the improvement of science, mathematics, and foreign language instruction in elementary and secondary schools, graduate fellowships, foreign language and area studies, and vocational-technical training" (para. 7). This emphasis on human capital for the defense of the nation carried with it a sense of urgency and began to shift the focus of the educational system onto product rather than process.

In the landmark Civil Rights Act of 1964, Title VI, prohibited discrimination based on race, color or national origin in any organization receiving federal funds. In the words of President John F. Kennedy (U.S. Department of Justice, 2013), "Simple justice requires that public funds, to which all taxpayers of all races contribute, not be spent in any fashion which encourages, entrenches, subsidizes or results in racial discrimination" (para. 2). President Johnson followed through with these sentiments after the assassination of John F. Kennedy and enacted the Civil Rights Act of 1964. Because of this emphasis on equity and the investment of all taxpayers, the accountability movement

began. It can be traced to the 1966 report titled *Equality of Educational Opportunity*, commonly known as the Coleman Report.

Sociologist James Coleman researched the distribution of educational resources among different races and the achievement test scores of students. When Coleman and his colleagues studied 600,000 school children, they found that the physical plants and teacher quality were very similar for black and white children. However, the achievement levels were very different because they showed that black children were behind their white counterparts by one to three grade levels in first grade and three to five grade levels by twelfth grade. This study represented a significant shift in accountability because of its focus on results, and it began a discussion about decreasing the achievement gap between races.

In 1983, the National Commission on Excellence in Education presented a study of the nation's educational system and the result was the report *A Nation at Risk: The Imperative for Educational Reform.* With the same sense of urgency seen following the launch of Sputnik, the authors implored:

> Our Nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world. This report is concerned with only one of the many causes and dimensions of the problem, but it is the one that undergirds American prosperity, security, and civility. We report to the American people that while we can take justifiable pride in what our schools and colleges have historically accomplished and contributed to the United States and the well-being of its people, the educational foundations of our society are presently being eroded by a

> rising tide of mediocrity that threatens our very future as a Nation and a people (U.S. Department of Education, 1983, para. 1).

As evidence for their concern, they demonstrated that high school achievement scores were now lower than when Sputnik had been launched. They reported that 17% of all high school seniors were functionally illiterate and the number could be as high as 40% among minorities. In addition to basic reading skills, the authors also looked at students' ability to solve problems; they reported that about 40% of 17-year-olds lacked the ability to draw the inferences necessary for critical thinking and could not solve math problems that required several steps (U.S. Department of Education, 1983). Finally, they warned of the frustration from military and business leaders with the inferior employment pool. Again, the human capital was found lacking and education was charged with the remedy.

Because education was (and still is) the single largest expenditure of most states, the accountability movement intensified. A growing body of literature shows that the best chance America has at increasing student achievement is through effective teachers (Nye, Konstantopoulus, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). However, a seminal report revealed that administrators made little distinction between teachers. *The Widget Report* (Weisburg et al., 2009) helped to ignite reform in teacher evaluation because it demonstrated a lack of oversight and assessment of effective teaching strategies. This study found that the majority of teachers were rated good or satisfactory despite the limited success of their students. For these reasons, the authors concluded that principals viewed teachers as equally effective and interchangeable like widgets or cogs in a machine. Teachers who were both underperforming and exceptional

were unidentified. According to the authors of *The Widget Effect*, prior to evaluation reform, follow-up training was inadequate. Approximately 75% teachers did not receive any specific feedback on improving their performance (Weisburg et al., 2009). Further, they were not given the opportunity to reflect on their own performance or set professional goals.

As a result of this lack of distinction among effective and ineffective teachers, many changes were proposed. Traditionally, in order to ensure quality teachers, policy makers increased credentialing requirements. For example, No Child Left Behind required that teachers be certified and high qualified. However, a growing body of research showed little connection between professional credentials and effective teaching (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006;) After studying 150,000 Los Angles school children and their teachers, Gordon, Kane, and Staiger (2006) found that the effectiveness of the teacher significantly impacted student achievement, but the teachers' effectiveness was not correlated with professional credentials. Further, they found that the teachers' effectiveness in the first two years was a reliable predictor of future success. They created a five-point plan to identify effective teachers in their report for the Brookings Institute: reduce entry barriers; make it harder to tenure least effective teachers; give financial bonuses to highly effective teachers willing to teach in disadvantaged schools, establish systems to measure teachers' job performance; track student performance and teacher effectiveness over time (Gordon et al., 2006).

All of these proposed changes rested on the need for reliable evaluations of teachers using objective data. The critical situation was confirmed with the report *A*

Nation Accountable: Twenty-Five Years After a Nation at Risk by the US Department of Education in 2008. This report concluded

> If we were "at risk" in 1983, we are at even greater risk now. The rising demands of our global economy, together with demographic shifts, require that we educate more students to higher levels than ever before. Yet, our education system is not keeping pace with these growing demands. (p. 1)

Between the release of this report and 2011, 32 states and the District of Columbia made substantive changes in their teacher evaluation systems (National Council on Teacher Quality, 2011). Most of these states used a combination of student growth data in the form of a value-added score, district created test scores, state end-of-course tests, surveys from students, and administrator observation data (Kane, Taylor, Tyler, & Wooten, 2011). Typically, half of the teachers' overall rating came from some type of measurement of student growth and half from surveys and observation data (National Council of Teacher Quality, 2011). The Bill and Melinda Gates Foundation (2009) worked with approximately 3,000 teachers to compare their value-added measures with student surveys and principal observations. They found that the most significant predictor of effectiveness was the prior year's value-added measure. For this reason, weighting state scores from 33% to 50% of a teacher's overall effectiveness rating provides a high rate of predictability (Bill and Melinda Gates Foundation, 2013).

Although there is a great deal of focus on using quantitative test data to distinguish effective and ineffective teachers, principal observations can prove equally reliable without devoting more resources to testing (Lefgren & Jacob, 2006). Many policy makers and stakeholders believed that principals could not distinguish among

teachers after *The Widget Effect* (Weisberg et al., 2009) reported that 94% of teachers were rated in the top two categories. However, in studies that asked principals to rate teachers on specific domains like a teacher's ability to raise math or reading achievement, administrators' scores correlated highly with test data. Like *The Widget Report*, Lefgren and Jacob (2006) found that teachers' overall ratings were high, but the subcategory ratings showed a wide range of scores. Likewise, the Bill and Melinda Gates Foundation (2013) found that principals discerned significant differences in teaching strategies. They also found that including observations helped to make a more stable and reliable effectiveness rating than merely using a state test score alone.

National Policy

According to Goe, Holdheide, and Miller (2014), prior to the Elementary and Secondary Education Act (ESEA), the role of federal and state legislators regarding educator evaluations had been almost nonexistent. However, when President Johnson signed the ESEA into law in 1965, the role of legislative policy makers began to increase. This act provided federal funding for many state initiatives like special education, new text and library books, and increased funding for districts serving low-income students (U.S. Department of Education, 2012). This was later reauthorized under a new name, No Child Left Behind (NCLB), in 2002 by President George W. Bush (U.S. Department of Education, 2012). Ten years later, as the deadline approached for closing achievement gaps in order to retain federal funding under NCLB, the Obama administration offered schools flexibility if they implemented certain components of the voluntary Race to the Top initiative (U.S. Department of Education, 2014). Those components included adopting rigorous standards and a teacher evaluation system, which was based on student

achievement data. According to Glazerman, Golfhaber, Loeb, Raudenbush, Staiger, and Whitehurst (2011), there is little consensus on the reliability of state test scores as a measure of teacher effectiveness. However, many school systems have included some type of value-added measure in their teacher evaluation systems. Additionally, implementing an objective observational measure is often included to differentiate among teacher effectiveness (National Council on Teacher Quality, 2011). Under Race to the Top, states that complied with these requirements may apply for grants funding special projects. Some of the projects include the creation of specialized schools and alternative diplomas (U.S. Department of Education, 2014).

In "Setting the Pace: Expanding Opportunity for America's Students under Race to the Top," the U.S. Department of Education (2014) proposed that NCLB had in fact encouraged lower standards and an over-emphasis on standardized testing. It indicated that the Race to the Top portion of the American Recovery and Reinvestment Act of 2009 would, in fact, decrease the frequency of "one-size-fits-all remedies" (p. 2). Race to the Top grants required reform in the following areas:

- adopting standards and assessments that prepare students to succeed in college and the workplace to compete in the global economy;
- building data systems that measure student growth and success and inform teachers and principals about how they can improve instruction;
- recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and turning around lowest-achieving schools (Joint Legislative Committee on Performance Evaluation and Expenditure Review, 2014 p. 27).

Even though the Department of Education aspired to reduce a uniform reform solution and emphasis on standardized testing, 46 states adopted and implemented very similar standards commonly referred to as Common Core State Standards and a variety of teacher evaluation models that continue to emphasize standardized tests. Common Core State Standards were the result of state efforts enacted through the National Governors Association for Best Practices (NGA Center) and the Council for Chief State School Officers (CCSSO) in 2009 (Common Core State Standards Initiative, 2015).

State Policy

Like other states, Mississippi adopted Common Core State Standards (CCSS) and new testing procedures in 2009 under the Mississippi Teacher Evaluation System (MTES) umbrella. According to the Joint Committee on Performance Evaluation and Expenditure Review, Mississippi adopted CCSS and new testing procedures in order to be competitive for national Race to the Top funding. Although Race to the Top did not require adoption of CCSS– states could create new standards independently– it embedded incentives for doing so by assigning weighed points in the application process to those criteria. Rather than reducing the emphasis on standardized testing, Mississippi is slated to spend an additional $2.5 million dollars because of "an increase in the number of assessments to be administered" (Joint Legislative Committee on Performance Evaluation and Expenditure Review, 2014 p. 1). Despite Mississippi's investment and grant application, the state was not awarded a Race to the Top grant in phase one, two, or three of the grant process.

Localized Impact

Glazerman et al. (2011) acknowledge the unlikely success of national or state mandated policy reforms because they are so far removed from the practice of daily educators. The Organization for Economic Co-operation and Development (OECD) represents a coalition of countries whose mission it is to strengthen "economic and social well-being" (OECD, 2015, para. 1) of international citizens. It administers the Programme for International Student Assessment (PISA), an extensive knowledge and skills survey of 15-year-old international students. Although it was unlikely to see results so soon after Race to the Top grant awards, three years after the implementation of the ambitious Race to the Top program, this group published "Lesson form PISA 2012 for the United States," which found no significant changes in American student performance in math, reading, or science. Among the 34 OECD countries, United States students ranked 26 in math, 21 in science, and 17 in reading.

Interestingly, researchers found that socioeconomic disadvantage had a significant impact on American students compared to their worldwide counterparts, but the schools of those American students did not differ significantly in teacher-student ratio or the proportion of certified teachers. In the United States, more than other countries, student relationships with teachers positively correlated with achievement, as well as teacher morale. Although higher spending did not correlate to increased achievement, "schools with more autonomy over curricula and assessment," as well as, "greater teacher-principal collaboration in school management" (OECD, 2013, p. 8) showed higher gains. Glazerman et al. (2011) reiterated the notion of a "light hand from levels of government

above the school district" (p. 5) and the importance of local collaboration of educators to meet the needs of their unique context.

When trying to create an evaluation system to identify highly effective teachers, they wrote: "buy-in from teachers and utilization of their expertise are most likely if the design of an evaluation system occurs at a level at which they feel they have real influence" (p. 5). This premise defies the centralized authority of bureaucratic policy makers, which defines the modern educational landscape (Bush, 2009). Cranston (2013) proposed that school leaders, not bureaucratic policy makers, should lead the discussion on evaluation reform. He emphasized replacing the idea of accountability taken from a business perspective with the notion of professional responsibility unique to education.

The externally imposed standards and assessments operating from a business model have become like an orthodoxy of accountability (Cranston, 2009; Eacott, 2009; Hargreaves & Shirley, 2009; Papa, 2011), which has redefined the function educational leaders to be "the 'doers' of the bidding of others" (Cranston, 2009 p. 131). Hardy (2010) concurs that in top-down initiatives principals are directed on the processes to be implemented. They are often left without autonomy or power as professionals (Bruce, Ross, Dookie, & Beatty, 2010; Mitchell & Sackney, 2009). This dynamic can diminish professional efficacy and harm student outcomes (Mitchell & Sackney, 2015). Although educational leaders were disgraced by the *Widget Effect* (Weisberg et al., 2009) findings, it did reveal that overwhelmingly, administrators had failed to effectively evaluate teachers and distinguish among levels of effective teaching. However, despite these facts, school administrators should not become silent on the matter of defining effective teaching (Fitzgerald & Gunter, 2008; Michell & Sackney, 2015).

The Purpose of Evaluation

Teacher evaluation may serve a variety of purposes. Haefele (1993) listed some of the functions as:

- Assisting administrators in making personnel decisions
- Recognizing excellence
- Identifying topics for professional training
- Meeting legal obligations of supervision
- Creating a shared vision between administrators and teachers

Like any enterprise, the purpose of a system should guide how it is developed and which components are included. In modern evaluation reform, not all systems are designed to serve all the purposes outlined by Haefele (1993). Further, policymakers, educators, and stakeholders often disagree on the most important purpose of an evaluation system (Danielson & McGreal, 2000). The various purposes of teacher evaluation are generally broken into two broad categories: formative and summative.

Formative Evaluation

Formative evaluation is generally defined as assessment designed to identify areas for professional development (Danielson & McGreal, 2000). Educators tend to prioritize the formative purposes of teacher evaluation (Danielson & McGreal, 2000; Mississippi Department of Education, 2010). If the primary purpose of a teacher evaluation system is formative, it may sometimes appear unfair if viewed through a summative lens. For example, if the purpose were to identify areas for growth, then the rubric would likely be designed to avoid a perfect performance (Moore, 2014). In essence, it may represent an unachievable ideal. If the system, is also serving a summative purpose, this type of rubric

may decrease teachers' sense of self-efficacy because of its unachievable design (Bandura, 1982, 1997). However, Danielson and McGreal (2000) and Marzano (2007) have developed teacher observation rubrics with both a formative and a summative purpose.

Summative Evaluation

Summative evaluation is generally defined as an assessment for the purpose of quality assurance (Danielson & McGreal, 2000). Policy makers and stakeholders often prioritize the summative purposes of teacher evaluation because of the investment of both money and children's future (Danielson & McGreal, 2000). The summative evaluation is designed to distinguish among teachers. It provides administrators with a tool to make personnel decisions that are justifiable in court. It may also provide a basis for incentives like merit pay for excellent performance. Some evaluation systems are designed exclusively for this goal to satisfy legislative requirement in an efficient manner (Strong, 2011). Some research has demonstrated that the principal has very little impact on student achievement through classroom observation (Meister, 2010); so, a minimal investment of time required by a strictly summative evaluation may be wise.

Effective Instruction: Who Decides?

Although many recent definitions of effective teaching have relied on certification credentials as the defining characteristics, the literature reveals a more complex dynamic involving a variety of elements. Title II, Part A of No Child Left Behind deals with teacher and principal quality. An effective teacher was defined as "highly qualified" by holding a bachelor's degree from a four-year institution, attaining the necessary certifications and endorsements, as well as demonstrating competence in core subject

areas being taught (U.S. Department of Education, 2012). Education is the largest single profession in the United States with a reported 3.7 million full-time elementary and secondary teachers in 2012 (U.S. Department of Education). Of these full-time teachers, 52 percent hold advanced degrees beyond the bachelor's (U.S. Department of Education, 2012). Unfortunately, a survey of the literature reveals very little correlation between a teacher's level of schooling and student achievement scores (Gordon, Kane, & Staiger, 2006; Sahin & Adiguzel, 2014).

When defining effective teaching, Cranston (2009) argued that the purpose and recipient of schooling must first be defined. He proposed that because of rhetoric and funding, one might assume the purpose of school is to perform on achievement tests of core academic subjects for economic advantage in industry (Cranston, 2009). Another purpose may appear to be preparation for college or career by learning basic subjects and thinking skills. On the contrary, many researchers point to broader, somewhat softer purposes embedded in schools: a moral purpose, a social and emotional purpose, and a democratic purpose (Beachum, 2011; Brooks & Kensler, 2011; Cranston, Ehrich & Kimber, 2006). However, with the emphasis on test scores and rankings, external accountability has overshadowed the other purposes because they are not measured in value-added scores or teacher observation rubrics. When school administrators simply comply with externally imposed regulations, they become focused on what must be done for accountability sake rather than what may or may not be in the best interest of their unique students (Cranston, 2009; Firestone & Shipps, 2007).

Horng and Loeb (2010) described the traditional view of instructional leaders as those who view themselves "as 'hands-on' leaders, engaged with curriculum and

instruction issues, unafraid to work directly with teachers, and often present in classrooms" (p. 66). Because instructional leaders are characterized as exhibiting "strong, directive leadership focused on curriculum and instruction" (Hallinger, 2003, p. 329), as well as data-focused and ultimately accountable for student achievement (Hallinger, 2005), these leaders bring an involved, direct-impact perspective to teacher evaluation. Unlike legislatively driven accountability, instructional leaders are close to the needs of individual learning communities and can better customize educational strategies for their student population.

Cranston (2009) argued that educational leaders must develop internal accountability as a profession and reverse the tide of mistrust of educators. Although there may be a sense among many administrators that enacting their own plans for success at the school-level is a fruitless endeavor because of the inevitable intervention of policy makers (Mulford & Edmunds, 2009), Fitzgerald and Gunter (2008) contends that administrators and teachers should lead the debate on reform in their own profession rather than relegating it to those with political agendas.

When identifying teacher effectiveness some studies show that principals value a mixture of affective and professional skills (Harris, Ingle, & Rutledge, 2014). Among the qualities that principals report as valuable are strong communication skills, content knowledge, and enthusiasm. However, Harris et al. (2014) found that the most important characteristic was caring. They report this attribute as ranked higher than teaching skill, content knowledge, and communication skill. Since, according to the OECD report (2013), secondary students in America who reported a positive relationship with their teacher achieved greater gains, it is not surprising that administrators would recognize the

caring teacher as a highly effective quality. In their 2013 study, Muguti and Mawere found that school administrators found effective teachers to maintain a supportive climate conducive to learning through effective classroom management and they maximize active learning opportunities. Likewise, when analyzing over thirty years of research, Wong and Wong (2012) isolated three basic characteristics of effective teachers. They are excellent classroom managers; they teach for the goal of learning and mastery; they expect success from their students.

Not surprisingly, when pre-service teachers were surveyed about their memories of effective teachers, the results echoed similar themes. Walker (2008) identified twelve characteristics defining effective teachers from students majoring in education based on their essays on the topic. First, he found they were always prepared, positive, and held high expectations for student success. Effective teachers were described as creatively engaging, fair, and personable- getting to know their students. They created a sense of belonging because they were caring, respectful, funny, forgiving, and humble enough to admit when they were wrong. Breault (2013) identified the five most frequent characteristics used to describe effective teachers as high expectations, genuine concern for students, content knowledge, variety of engaging activities, and enthusiasm.

Similarly, when 340 teachers were surveyed on their views of effective teaching many of the attributes listed were affective nature (Koutrouba, 2012). The most important characteristic identified was the teacher's ability to make sure that information was comprehensible and processed by the students while being flexible and responsive to various learners with tact and warmth. At the same time, teachers felt that maintaining a structured, orderly learning environment and an impartial attitude toward students was

key. Subjects in this study believed that effective teachers deviate from the official curriculum when appropriate to differentiate and individualize instruction. Finally, unlike the stereotype of the strict teacher who scares students into compliance, teachers believe that effective instruction most often happens through friendly exchanges with "open-heartedness and open-mindedness" (Koutrouba, 2012, p. 369).

In another study of mathematics, science, and computer teachers, researchers divided characteristics into three groups: personal, professional, and classroom management (Sahin & Adiguzel, 2014). The *personal* category consisted of items such as friendliness, flexibility, sense of humor, and willingness to admit mistakes. The *professional* category included items like explains clearly, makes material relevant to real life, engages students with content, and holds high expectations of students. The *skills* category was distinguished by descriptors such as grades student work fairly, keeps students on task, provides feedback on assignments, and good classroom organization (Sahin & Adiguzel, 2014, p. 636). Consistent with the other studies discussed, Sahin and Adiguzel (2014) found that personal characteristics were perceived as most important to effective teaching. In fact, the top three descriptors were "enjoys teaching," "respectful of all students," and "good communicator" (p. 643). Although personal qualities seemed to be a necessary foundation on which professional strategies and classroom management skills rest, these qualities also proved important. Teachers viewed as effective were prepared and disciplined, as well as those who explained material clearly. Subjects also reiterated the importance of grading student work fairly.

Although the definition of effective teaching varies, the attributes and cost of an ineffective teacher is clear. In Bridges (1986) iconic work, *The Incompetent Teacher*, he

identified several characteristics that prevent teachers from helping students to grow academically or socially. He describes these characteristics as:

1. Failure to maintain discipline;
2. Failure to treat students properly;
3. Failure to impart subject matter effectively;
4. Failure to accept teaching advice from superiors;
5. Failure to demonstrate mastery of the subject matter being taught; and
6. Failure to produce the intended or desired results in the classroom (p. 5).

Of these failures, research shows that failure to maintain a disciplined environment is the most common and can cause the most damage. Clearly, the balance between affability and compassion for students, and maintaining discipline highlights the nuances of effective teaching. It requires flexibility and responding to unique children in an organic learning environment that is ever-changing. Additionally, since teachers may be interacting with children age 3 to 18, any attempt to define effective teaching should consider developmentally appropriate techniques for specific ages.

Even a generic list of educational best practices differs among the ages and stages of children. For example, the Alliance for Childhood (2002) suggested that child-initiated play is a building block of early-childhood education. Further, they report, "The rough and tumble of active play facilitates a child's sensorimotor development" (p. 1). This stage may last up to age eight depending on the individual characteristics of the child. On the other hand, many of the sixteen, seventeen and eighteen-year-olds in high school, have transitioned to adult learners. Knowles (2005) indicated that although adult learners enjoy engaging lessons and sharing their own experience, adults need immediately

relevant and useful information. They are capable of comprehending abstract concepts and inferring deeper meaning. Conversely, adolescent learners, described by the National Middle School Association as students who are between ten and fifteen years old, need lessons that incorporate the sense and emotions (Wilson & Horch, 2002), as well as those incorporating movement and exercise. During this time, adolescent's bones are hardening, especially the tailbone, which make it difficult to sit for hours. Clearly, children face different challenges and thrive with different types of learning opportunities at various stages of development. However, many of the teacher observation rubrics or frameworks adopted by states in order to assess effective classroom instruction are the same for teachers of children ages four through eighteen (Danielson & McGreal, 2000; Marzano & Toth, 2013; Marshall, 2013).

Common Teacher Observation Rubrics

Framework for Teaching

Many states have adopted some form of the Framework for Teaching pioneered by Charlotte Danielson. Thousands of schools nationally and internationally are using the framework. The framework includes 4 domains, 22, components, and 76 smaller elements. The four domains are Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities (Danielson & McGreal, 2000; The Danielson Group, 2013). In the Planning and Preparation domain, teachers are expected to demonstrate knowledge of subject matter and learning dynamics while designing both lessons and assessments. The second domain, Classroom Environment, again places the teacher in the role of designer by charging him or her with creating a culture of learning and an inviting physical space. The teacher must also demonstrate strong management

skills in this domain in order to channel student behavior in the direction of learning. The Instruction domain assesses the teachers' interpersonal skills in effective communication to engage the students and check for understanding. Because teaching can be a rather isolated activity, domain four, Professional Responsibilities, requires that the teacher communicate with other professionals to learn new techniques and strategies. It also assesses the teachers on communicating with stakeholders and parents (Danielson & McGreal, 2000; The Danielson Group, 2013).

The Danielson Framework for Teaching is rooted in the constructivist theory of learning (The Danielson Group, 2013) and has become somewhat institutionalized as the official method of evaluating teachers. This construct affects teachers in the broader political sense by defining knowledge and beliefs of how learning is created (Educational Broadcasting Corporation, 2004). The one characteristic rated most highly by administrators was "caring," (Harris et al., 2014), yet this affective characteristic is one of the 77 small characteristics within one of the 22 components under one of the domains in the Danielson rubric. The Danielson Framework for Teaching is rooted in the constructivist theory of learning and when used in isolation, can devalue other theories of learning and pedagogical practices.

Marzano Protocol

Another popular teacher observation rubric was developed by Robert Marzano for the purpose of both measuring teacher quality and developing teacher potential (Marzano, 2007; Marzano & Toth, 2012). Like the Charlotte Danielson Framework for Teaching, Marzano's Teacher Evaluation Rubric contains four domains: Classroom Strategies and Behaviors, Preparing and Planning, Reflecting on Teaching, and Collegiality and

Professionalism. There are 60 elements in the entire rubric with 41 elements concentrated in Domain One: Classroom Strategies and Behaviors. This domain reflects the complexity of teaching with nine broad design questions, which are then divided into three segments: lesson segments involving routine events, lesson segments addressing content, and lesson segments enacted on the spot. According to Peggy Schooling, Director of Teaching, Learning, and Development at Learning Sciences International (2011), "Domain 1 is the most complex and has a direct causal link with student achievement" (para. 3). The components listed in this domain are observable teaching strategies like "examining similarities and differences, hypothesis generating and testing, noticing when students are not engaged, demonstrating value and respect for low expectancy students" (Marzano & Toth, 2012, p. 44).

Although Domain Two through Domain Four are important for teacher growth and quality evaluation, they represent fewer elements. Marzano emphasizes that Domain 2, Planning and Preparation, is directly linked to classroom strategies and behaviors by weighting this category heavily. There are eight elements organized into three segments: lessons and units, materials and resources, and the special needs of students. In Domain 3: Reflecting on Teaching, instructors have an opportunity to evaluate their own practice. Teacher self-reflection has been shown to enhance development and professional practice. This domain is divided into two categories with a total of five elements that encompass a self-evaluation and professional growth plan. The final domain, Domain 4: Collegiality and Professionalism contains three categories with two elements in each. Teaching can often feel like an isolated endeavor, but this domain assesses how the teacher reaches beyond the classroom to collaborate with colleagues, participates in

school or district initiatives, and interacts positively with parents and stakeholders. Despite the lack of a causal like between these activities and student achievement, there is evidence that they inform the school climate and professional culture, which is correlated to academic achievement (Glickman, Gordon, & Ross-Gordon, 2010; Hallinger, 2003; Sebastian & Allensworth, 2012).

Strong's Rapid Assessment of Teacher Effectiveness

Not all teacher observation rubrics seek to achieve the dual purpose of evaluation and development as in the Danielson and Marzano protocols (Danielson, 2007; Marzano & Toth, 2013). A rubric with this dual purpose is more complex, adding layers of strategies and teaching behaviors in order to adequately accomplish both summative teacher evaluation and formative professional development assessment (Marzano & Toth, 2013). However, some rubrics like Strong's Rapid Assessment of Teacher Effectiveness (RATE) function primarily as a summative evaluation tool and encompass a more streamlined list of observable characteristics of the learning environment (Strong, 2011). The latest version of the RATE rubric focuses on only "six items relating to the lesson objective, instructional delivery mechanisms, teacher questioning strategies, clarity of presentation of concepts, time on task, and the level of understanding" with each item being rated on a three-point scale (Gargani & Strong, 2014, p. 392). The purpose of this rubric is not a comprehensive picture of 'good' teaching, rather it is focused on evaluating observable strategies that predict student outcomes on standardized testing and meet the federal requirements for Race to the Top legislation (Gargani & Strong, 2014). When judging effective teaching based solely against student academic achievement, research suggests that the RATE system discriminates between teacher quality more

effectively that many other rubrics (Marzano & Toth, 2013; Strong, 2011; Gargani & Strong, 2014). Further, the RATE method requires less training of evaluators and fewer observations of teachers than those that Kane and Staiger (2012) evaluated in their MET study. Finally, with only one 20-minute observation, Gargani and Strong (2014) "were able to generate scores that were consistently more reliable, predictive, and inexpensive" (p. 391). When a state or district's purpose is to distinguish among levels of instructors for personnel decisions, Strong's RATE system represents a viable alternative.

Points of Divergence

Differences between the RATE system, Marzano's protocol, the Framework for Effective Teaching might suggest that different purposes for teacher evaluations lead to points of divergence among the protocols used. Educators disagree on one correct framework for teacher observations including who should assess teachers, how often, and under what conditions (Hull, 2013). For example, some points of difference include not only which rubric is most effective, but also the frequency and duration of observations. Unlike Strong's method using one 20-minute observation, many teacher evaluation models now include multiple observations by the principal (Danielson, 2012; Marzano & Toth, 2013), yet there is little evidence supporting the effectiveness of that practice (Sebastian & Allensworth, 2012; Whitehurst, Chingos, & Lindquist, 2015). Some programs require as many as eight walk-through visits defined as brief, informal observations of classroom activity. However, research indicates a negative impact of brief "walk-through" observations at the high school level (Sebastian & Allensworth, 2012). Whitehurst, Chingos, and Lindquist (2015) found that "observations conducted by evaluators from outside the building have higher predictive power for value-added scores

in the next year (0.21) than those done by administrators in the building (0.15)" (p. 63). These researchers also found that the predictive value of observations did not increase with more than two observations. In fact, the very nature of teacher observation and supervision outlined in Race to the Top guidelines "have consistently proven ineffective in raising student achievement" (Dufour & Mattos, 2013, p. 34).

Although there is disagreement on the predictive power of administrator observations (Whitehurst et al, 2015), research indicates that best practices can be measured by various methods (Hanover Research, 2012). Marzano (2007) contends that a rubric that essentially functions as checklist of best practices like the RATE system defies the organic spirit of reflective practice. Additionally, Dufour and Marzano (2011) conclude that if the principal is looking for a one-size-fits-all effective teaching strategy during the observations-- it does not exist. Often principals rely on a generic rubric to assess teachers because they do not have specialized content knowledge in every area (Dufour & Mattos, 2013). In the MET study, five different frameworks were evaluated that according to the authors, do not represent simple "checklists, focusing on easy-to-measure but trivial aspects of practice" (Kane & Staiger, 2012, p. 12). Instead, they represented a comprehensive look at teaching and learning requiring up to 76 items on a rubric. These tools required an extensive investment of time and resources in training, as well as implementation of the protocol (Gargangi & Strong, 2014; Hull, 2013; Kane & Staiger, 2012; The New Teacher Project, 2013).

Although designed to reflect the complexity of teaching and learning, many teacher observation rubrics present significant logistical challenges for administrators. The New Teacher Project (2013) research indicates that the process and time constraints

required by these in-depth observation rubrics are overly burdensome for administrators and though the purpose of such a lengthy rubric is to provide professional development opportunities, there is in fact little time left for focused feedback and development. Additionally, according to the 2013 State of the States report from the National Council on Teacher Quality, "… if observation rubrics are too detailed and try to capture too much, there is a danger that they can become unworkable instruments for differentiating teacher performance" (Doherty & Jacobs, p. 31). Finally, one of the key findings in the Center for Education's report, *Trends in Teacher Evaluation*, was that "Local school districts need flexibility in designing and implementing teacher evaluation systems so they are aligned to the needs of the district" (Hull, 2013, p. 3) while still receiving necessary support and resources from the state. These concerns reflect the common divergence between something that is excellent in theory yet very challenging in practice.

Mississippi Teacher Evaluation System

Development Process

In 2010, the Mississippi Department of Education formed the Statewide Teacher Evaluation Council (STEC) in order to make recommendations on a new statewide teacher evaluation framework aimed at complying with national initiatives like Race to the Top. This committee included 20 individuals representing a broad range of stakeholders. After some initial work defining guiding principles, 60 teachers who attended the Mississippi Delta Community College's Millennium Partnership Summer Institute for Secondary Teachers were surveyed along with the STEC stakeholders regarding implementation options. Both groups felt that using teacher evaluations as formative data for improving instruction should be the highest priority of the new

evaluation system (Mississippi Department of Education, 2010). They also agreed that thorough training of evaluators and timely feedback would be crucial. However, the two groups differed on how many observations were necessary and who should conduct them. The group of teachers believed strongly that no more than two observations would be necessary and that other teachers should conduct them—not administrators (Mississippi Department of Education, 2010). This sentiment is reflected in the literature as well. Some research has found that teachers are often ambivalent about the influence of their administrators (Meister, 2010). For many veteran teachers, they outlive the professional lives of principals they work with over the years. Teachers acknowledge that their colleagues are the most important influence on their practice—not administrators (Meister, 2010). On the other hand, the group of stakeholders from STEC indicated on their surveys that administrators should conduct the observations and more than two observations per year would be necessary (Mississippi Department of Education, 2010); the state of Mississippi concurred with Statewide Teacher Evaluation Council members. The report went on to indicate, "These differing responses may in some way relate to the perception by teachers of the effectiveness and utilization of evaluation results" (Mississippi Department of Education, 2010, p. 3).

MTES Purpose

Like many other states, Mississippi has actually implemented a new teacher evaluation system in order to comply with federal regulations. According to the Mississippi Department of Education (2015), "The Mississippi Statewide Teacher Appraisal Rubric (M-STAR) is an evaluation process designed to improve the

professional performance of all educators" (para. 1). With this purpose in mind, Mississippi invested in a system with professional development as its goal; therefore, it is designed to present a comprehensive picture of the educator as a whole within a dynamic learning environment and identify areas for improvement.

In 2011, Mississippi began a pilot program using the new rubric and collected focus group feedback in 2012. From July 2012 through July 2013, the state trained administrators as evaluators at professional development conferences, and during the 2013-2014 school year began a full-scale field test. After soliciting feedback from administrators, some changes were made and implemented in 2014-2015 (Mississippi Department of Education, 2014b).

MTES Revisions

Using the MTES framework, teachers are scored using data from testing, as well as data from direct classroom observations. During the first year of implementation, 2014-2015, teachers were divided into two categories: state tested and non-state-tested. For state tested teachers, 50% of their overall score depended on classroom observations, 20% of their score depended on school-wide growth, while 30% varied according to individual growth based on test scores. For non-state-tested teachers, 50% of their overall score was based on classroom observation, and the other 50 % depended on school-wide growth (Mississippi Department of Education, 2014c).

In coming years, beginning in 2015 -2016, the state plans to change the proportions somewhat and add an individual component for non-state-tested teachers. The classroom observation piece will remain constant for both categories of teachers with 50% of the score resting on the assessment by evaluators. For state tested teachers,

school-wide growth will be weighted 20%, while individual growth will increase to 30%. Instead of non-state-tested teachers' overall score consisting of only school-wide growth, it will also include an individual growth component weighted just as the state tested teachers. Non-state tested teachers will use student learning objectives that are measured on pre- and post-test measures for their particular students (Mississippi Department of Education, 2014c).

Although teachers surveyed by the Statewide Teacher Evaluation Council indicated the efficacy of only two peer observations per year, the council recommended a different course of action. While, this process was later revised to remove the summative evaluation, during the field test of 2013-2014, administrators conducted a formative observation in the fall and a summative observation in the spring; both were preceded by a pre-observation conference and followed by a post-observation conference. Additionally, administrators conducted five informal walk-through observations. The walk-through observations spanned between 10 and 30 minutes, whereas the formal observations were intended for an hour or more depending on the class period.

Following the field test, several changes were implemented. Originally professional growth goals were slated to account for 20% of a teacher's score, but that component was eliminated. The number of observations decreased from seven to three: one formal formative observation with a required post-conference and optional pre-conference and two walk-through observations. The length of the walk-through observations was left to the administrator's discretion and the formal formative observation was reduced from the entire class period to 30 minutes. Some of the

performance level indicator language was made more precise for discrimination purposes (Mississippi Department of Education, 2014c).

MTES Policy Decisions

When comparing Mississippi's progress on implementing teacher effectiveness policies, clearly things have changed. Mississippi is among the 27 states that require annual teacher observations and among the 42 that consider heavily students' achievement data as a criterion for teacher evaluation. These 42 states count student test data as at least 50% of the total teacher evaluation score. However, Mississippi does not consider teacher evaluations in awarding tenure or releasing ineffective teachers. In fact, there is no official use for the summative evaluation. Twenty-nine states have a policy of dismissing ineffective teachers as identified through teacher evaluations, yet Mississippi has no such explicit policy. Additionally, Mississippi does not use teacher evaluations in consideration for licensure advancement, reciprocity, layoffs, teacher preparation, program accountability, or student teacher placement (National Council on Teacher Quality, 2014). These policy decisions are consistent with Mississippi's stated goal for the evaluation system of improving teacher quality. Unlike some states that are seeking to distinguish among levels of effectiveness for personnel decisions and identify areas of needed improvement, this data suggests that Mississippi is singularly focused its teacher evaluation system on distinguishing among teachers to determine their professional development needs.

Mississippi Statewide Teacher Appraisal Rubric

The Mississippi Statewide Teacher Appraisal Rubric (M-STAR) is the tool administrators use when assessing classroom teachers. It is based on Charlotte

Danielson's Framework for Teaching with the addition of one domain, Assessment. The M-STAR rubric contains 20 standards divided into five domains: Planning, Assessment, Instruction, Learning Environment, and Professional Responsibilities. Half of the standards are focused in the Instruction and Learning Environment domains. These two domains are the only ones assessed through classroom observation. The other three domains, Planning, Assessment, and Professional Responsibilities are assessed through the collection of artifacts and the pre- and post- observational conferences. Mississippi has also created an optional student survey, which may be used to inform the evaluator about assessment practices, instruction, and classroom climate (Mississippi Department of Education, 2014a). Although this survey's language and format is appropriate for junior high and high school age children, the process manual indicated, "School districts may create their own survey or use one that is appropriate" (Mississippi Department of Education, 2014a, p. 5).

M-STAR Domains. Each domain represents a broad category of assessed standards. The Planning domain contains four standards while the Assessment domain contains only two. In the Danielson Framework for Teaching, these domains are combined into one domain titled Planning and Preparation. The standards in these domains indicate that teachers should have extensive content and pedagogical knowledge while using this information to design lessons aligned with Mississippi Curriculum Frameworks and College and Career Readiness Standards. The expectation in this domain is that teachers also demonstrate knowledge of their students as individuals and differentiate lesson when appropriate. Further, they should incorporate assessments into lesson planning and use the data resulting from assessments to inform future planning, as

well as give feedback to students. The standards in these two domains are evaluated during the optional pre-observation conference, the mandatory post- observation conference, and through the collection of artifacts such as lesson or unit plans (Mississippi Department of Education, 2014a).

The Instruction and Learning Environment domains encompass the largest number of standards assessed with ten equally distributed standards. In order to provide a climate conducive to learning with high expectations for all, teachers are expected to manage student behavior and to create and maintain a safe environment where respect and support are always available. They should make effective use of space and resources including instructional time (Mississippi Department of Education, 2014a).

When providing instruction, the standards indicate that again teachers should demonstrate thorough knowledge of the content and pedagogy. They should actively engage students in the learning process by using questioning and discussion techniques, as well as bringing multiple perspectives to bear in their delivery. Furthermore, the standards dictate that teachers should communicate clearly and effectively when delivering instruction. Both the Instruction and Learning Environment must be assessed via classroom observation by an administrator, but may also be informed by student surveys if the school or district chooses (Mississippi Department of Education, 2014a). Whether to distribute the student survey and how to use the results of the survey is left to the discretion of local educational leaders (Moore, 2014). Many schools do not use this tool, but some find it a valuable resource. For example, one school leader indicated that faculty collaborated to make minor revisions to the survey instrument before disseminating it. Further, it was used as a 'reality check' for teachers and one part of the

formative puzzle that helped teachers better meet the needs of their students (Moore, 2014).

Finally, the Professional Responsibilities domain contains four standards, which indicate a high level of collaboration with both colleagues and stakeholders. These standards indicate that teachers should be continuously learning about new research in their profession and implementing what they learn in the classroom. Teachers should communicate with parents and colleagues in the best interest of their students. Following the Mississippi Code of Ethics is also delineated in this domain. The standards in Professional Responsibilities are evaluated during the optional pre-observation conference, the mandatory post- observation conference, and through the collection of artifacts such as a professional learning community agendas and parent communication logs (Mississippi Department of Education, 2014a).

M-STAR scoring. Teacher performance in each of the five domains is scored in one of four levels: distinguished, effective, emerging, and or unsatisfactory. Although teachers may be evaluated as distinguished in individual domains, it is unlikely any teacher would be evaluated as distinguished in all domains. The purpose of the M-STAR rubric is to

- Provide formative assessment information about the performance of individual teachers to highlight areas of strength and identify areas for growth.
- Serve as a guide for teachers as they reflect upon their own practices.
- Provide shared understanding regarding priorities, goals, and expectations of quality practice.

- Serve as a tool to help structure instructional leadership and feedback. (Mississippi Department of Education, 2014a, p. 3)

It is not intended to reveal perfection, rather areas for improvement. When referencing the highest range of scores, some administrators use terms like "visiting" a four or "floating" up to a four occasionally (Moore, 2014). Perhaps, this is one reason that Mississippi no longer requires a summative evaluation or uses this data to inform personnel decisions. Teachers at the distinguished level are considered exemplary in that domain and consistently exceed expectations. Effective teachers are considered those who consistently meet expectations. According to the Mississippi Department of Education (2014a), "Effective: Level 3 is the expectation for all teachers" (p. 4). Level two teachers are considered emerging and may indicate new teacher status or someone who is not consistently meeting expectations. Finally, an unsatisfactory, level one teacher rarely meets expectations. It is recommended that these teachers "receive immediate and comprehensive professional development" (p. 10).

CHAPTER III - METHODOLOGY

Participants

All elementary, middle and high school principals of standard public schools in the state of Mississippi were invited to participate in this study, although they could select a designee from their school to complete the questionnaire. Standard public schools were defined as those institutions not used for exclusively specialized education such as career technical schools, special education only schools, or alternative schools. Principals were defined as those school officials designated by the Mississippi Department of Education on the list "Principal Contact Information SY 2015- 2016" (Mississippi Department of Education, 2015). Other school officials such as assistant principals may have had expertise and duties regarding teacher evaluation, and they may have participated if designated by the principal. School principals were contacted through the publicly available name and address list provided on the Mississippi Department of Education's website (Mississippi Department of Education, 2015).

Instrument

A search for an instrument to measure administrator beliefs regarding teacher evaluation rubrics yielded minimal results. Therefore, the researcher created the instrument used to determine differences among principals of elementary, middle, and high schools concerning their appropriateness perceptions of M-STAR. This instrument was used for both the pilot study and the final instrument. Initially, respondents were asked to indicate their school context as either elementary, middle, or high school. Other demographic data requested included the respondent's position at the school, his or her experience as a teacher, his or her experience as an administrator, and his or her

experience with M-STAR. For the purposes of this study, responses from principals, assistant principals, or other evaluators experienced with M-STAR were included in data collection.

The pilot instrument included 48 questions divided into six major sections. In the first section, the researcher rephrased the ISLLC standards as questions and related them to the M-STAR instrument in order to determine the administrators' beliefs regarding M-STAR alignment with the standards of certification for school administrators. The next five sections corresponded to the domains included in M-STAR: Planning, Assessment, Instruction, Learning Environment, and Professional Responsibilities. The questions in each section ranged from four to twelve. The quantity of questions varied because they were based on the indicators listed in the teacher observation rubric domains, which also varied in number.

The survey instrument asked respondents to rate their beliefs regarding appropriateness of M-STAR domains in the context of their instructional level and the overall alignment with ISLLC standards. It employed a positively packed five-point rating scale in order to elicit variable responses. The possible responses included: Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree. In order to gain further insight into the perceptions of efficacy for M-STAR, one identical question was included in each domain section. Administrators were asked to rate their belief in the ability of any teacher at their school level to achieve a perfect score in each domain.

Design

A cross-sectional design drove the development of this study. According to research methodology, a cross-sectional design allows the researcher to sample a variety

of participants who are similar in certain characteristics at one moment in time (Shanahan, 2010; Williams, 2007). In this study, the dependent variable of interest was the beliefs regarding appropriateness regarding individual domains of M-STAR, as well as beliefs regarding M-STAR alignment with ISLLC standards. The common characteristics shared were those associated with being a school administrator of any educational level. For example, all school administrators in Mississippi are certified according to the ISLLC standards, and all administrators must use M-STAR to evaluate teachers. Accordingly, the independent variable for this study was group membership. This independent variable had three levels: elementary administrators, middle school administrators, and high school administrators.

The cross-sectional design enabled the researcher to ascertain differences between the levels of the independent variable. Use of the researcher-created survey instrument allowed the researcher to isolate differences in how administrators at various educational levels perceived the appropriateness of M-STAR as a teacher evaluation tool at their specific level. This design allowed the researcher to draw inferences regarding the efficacy of M-STAR at all levels of education.

Procedures

After receiving IRB approval, the researcher distributed a pilot questionnaire using a Likert scale that assessed administrators' perceived levels of appropriateness for M-STAR. The pilot survey was administered in one district consisting of six schools. For the purposes of the pilot study, a survey was mailed to each administrator at every school in this district along with an open-ended feedback form for a total of twenty

participants. The pilot questionnaire consisted of 48 questions divided into six sections. The results of the pilot study were used to establish content validity and reliability. This project utilized a traditional distribution of the questionnaires in order to increase the overall responses. The researcher mailed a copy of the survey to principals of each school in Mississippi. The Mississippi Department of Education makes those names and addresses available to the public on the Mississippi Department of Education website (Mississippi Department of Education, 2015). A hard copy of the survey was mailed to each non-specialized school on the list totaling approximately 900 schools. Special schools such as attendance centers, alternative schools, and career centers were not included in the mailing. The Mississippi Department of Education makes a list of all administrators, their school addresses, and email addresses available on its website for anyone to access. Although many schools have multiple administrators, the researcher only sent one survey to the school's head principal. However, the researcher requested the position of the respondent in the possibility that the principal was not the respondent. In addition to the survey, the researcher included informed consent and a self-addressed stamped envelope for easy return.

Data Analysis

The researcher analyzed the data from the responses received on the survey instrument using SPSS software. The first six questions solicited demographic data. The first question categorized administrators school level as elementary, middle, or high school while the second asked for their title. The final demographic questions were included to establish their experience in education, as an administrator, and with M-STAR. The dependent variable was the perception of appropriateness for domains of M-

STAR. The analysis of variance (ANOVA) was the most effective method for analyzing items that measure a dependent variable with several groups for statistical significance. The researcher manually entered the data from the questionnaires using SPSS and accounted for any missing data. Finally, the results were analyzed considering the research questions.

CHAPTER IV – RESULTS

Since the Mississippi Statewide Teacher Appraisal Rubric (M-STAR) is used for all grade levels, the purpose of this study was to determine differences or similarities in the perceptions of administrators regarding the appropriateness of each rubric domain for their school's grade levels. The study also assessed the administrators' level of confidence in the rubric based on their beliefs regarding M-STAR alignment with ISLLC standards as well as the administrators' beliefs regarding teacher efficacy using M-STAR at all levels based on their confidence in teachers' ability to attain a perfect score in each domain. These dependent variables: perception of appropriateness for each domain, confidence in M-STAR alignment to ISLLC standards, and perception of teacher efficacy using M-STAR for each domain were measured based on the administrators' grade level assignment. This chapter presents the resultant data from a questionnaire that was distributed using the United States Post Office for delivery.

Pilot Study

After conducting a pilot study with 22 respondents, all questionnaire sections were shown to be reliable with a Cronbach Alpha score greater than .70 except for the Assessment Domain. This domain contained the fewest number of items, only three, and had a Cronbach Alpha of .564. Although the reliability for this section was low, the researcher retained the items for the study in hopes that a larger sample size would increase the reliability.

Sample Results

Descriptive Information of the Sample

An additional 761 questionnaires were distributed to all school districts in Mississippi. Of the 761questionnaires distributed, 122 documents were returned for analysis. The first section of the questionnaire collected demographic data from the 122 administrators who responded. The data included: the grade range that aligned with the administrator's current assignment (grade level); the administrator's current position as principal, assistant principal or other (position); whether or not the administrator's school currently used the M-STAR instrument; the total years of administrator experience the participant had (administrator experience); the total years of teaching experience the administrator had prior to becoming an administrator (teaching experience); total number of years the administrator had used the M-STAR instrument. All schools reported using the M-STAR instrument to evaluate teachers; the remaining demographic data is reported in Table 1.

Table 1

Frequencies and percentages of demographic variables

Variable	Frequency	Percentage
Grade		
Elementary	53	43.8
Middle	29	24.0
High School	39	32.2
Administrative Position		
Principal	100	82.6
Assistant Principal	18	14.9
Other	3	2.5
Administrator Experience		
0 - 5 Years	52	43.0
6 – 10 Years	37	30.6
11 and more	32	26.4
Teaching Experience		
0 – 5 Years	24	19.8
6 – 10 Years	42	34.7
11 and more	55	45.5
M-STAR Experience		
0 – 1 Years	6	5.0
2 – 3 Years	83	68.6
4 Years	32	26.4

The second section of the questionnaire included statements regarding M-STAR's alignment with ISLLC Standards. These statements were posed using a Likert scale ranging from Strongly Disagree to Strongly Agree with Strongly Disagree represented as a 1, Disagree represented as a 2, Neutral represented as a 3, Agree represented as a 4, and Strongly Agree represented as a 5. Section 2 contained eight statements that were designed to measure administrators' beliefs regarding how well the M-STAR document reflects the professional standards of administrators as articulated by the Council of Chief State School Officers and commonly referred to as ISLLC Standards. Since administrative best practices are based on these standards, results in this area help to establish an underlying sense of validity for the M-STAR instrument prior to assessing individual domains by grade level. Table 2 lists the means and standard deviations for the statements in this section with the lowest mean, 2.96, corresponding to the ability of M-STAR to promote the faculty response to diverse community needs. The highest mean, 3.91, was produced when administrators reflected on M-STAR's ability to foster staff professional growth.

Table 2

Descriptive Statistics on M-STAR alignment with ISLLC standards.

Statement	N	Mean	Std. Dev.
1. M-STAR is an effective management too that promotes an efficient school.	121	3.54	0.97
Elementary School Level	53	3.55	1.02
Middle School Level	29	3.59	0.97
High School Level	39	3.49	0.87

2. M-STAR promotes collaboration with faculty.	121	3.45	1.03
Elementary School Level	53	3.47	
Middle School Level	29	3.21	1.02
High School Level	39	3.62	1.16
			0.89
3. M-STAR promotes responding to diverse community needs.	121	2.96	0.97
Elementary School Level	53	2.79	0.98
Middle School Level	29	2.97	1.03
High School Level	39	3.18	0.84
4. M-STAR is fair and promotes acting with integrity and ethics.	121	3.67	0.90
Elementary School Level	53	3.66	
Middle School Level	29	3.83	0.97
High School Level	39	3.56	0.75
			0.87
5. M-STAR reflects an achievable standard of excellence for a teacher of any subject area.	121	3.54	1.10
Elementary School Level			
Middle School Level	53	3.55	1.09
High School Level	29	3.69	1.05
	39	3.41	1.10
6. M-STAR reflects our school's vision of learning.	121	3.69	0.91
Elementary School Level	53	3.74	0.91
Middle School Level	29	3.83	0.91
High School Level	39	3.51	0.87
7. M-STAR reflects a school conducive to student learning.	121	3.82	0.89
Elementary School Level	53	3.83	0.93
Middle School Level	29	3.93	0.87
High School Level	39	3.72	0.85
8. M-STAR reflects a school conducive to staff profession growth.	121	3.86	0.88
Elementary School Level	53	3.91	0.87
Middle School Level	29	3.76	0.90

High School Level	39	3.87	0.85

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

Sections 3 – 7 of the questionnaire addressed each domain of the M-STAR instrument: Planning, Assessment, Instruction, Learning Environment, and Professional Responsibilities. Each section began with the same question regarding the efficacy of this domain for teachers at the grade levels of the reporting administrator. The remaining questions in each section varied in number because the M-STAR instrument itself varies in length by domain. However, all statements were posed using a Likert scale ranging from Strongly Disagree to Strongly Agree with Strongly Disagree represented as a 1, Disagree represented as a 2, Neutral represented as a 3, Agree represented as a 4, and Strongly Agree represented as a 5. The results of the identical first question are reported separately from the statistics regarding specific components of each domain. Appendix D shows the means and standard deviations for the five questions regarding the efficacy of each M-STAR domain. Middle school principals consistently reported a higher likelihood than elementary or high school principals that their teachers would achieve a perfect score in each domain with means ranging from 2.76 to 3.34. However, nearly all means for all grade levels were approaching three, which places them between disagree and neutral on the likelihood of teachers attaining a perfect score in any particular domain.

Section 3 of the questionnaire contained seven statements designed to measure each administrator's beliefs regarding the appropriateness of the M-STAR Planning Domain for his or her school's grade levels. The statements were selected using the M-STAR standards from this domain and further narrowed by identifying those items that

may vary according to children's developmental levels. Appendix E indicates the means and standard deviations of responses in this section with a high mean of 4.35 for including documentation of standards alignment in lesson plans and a low mean of 3.04 for collaborating with students to create lesson plans.

Section 4 of the questionnaire contained three statements designed to measure each administrator's perception of the appropriateness of the M-STAR Assessment Domain for his or her school's grade levels. Like the Planning Domain, the statements were selected using the M-STAR standards from this domain and further narrowed by identifying those items, which may vary according to children's developmental levels. Appendix F indicates that the means and standard deviations of responses in this section. Using summative assessments to verify learning garnered the most consensus with a mean of 4.17, but all items in the domain produced means higher than 4.0.

Section 5 of the questionnaire contained the most items with eleven statements designed to measure each administrator's perception of the appropriateness of the M-STAR Instruction Domain for his or her school's grade levels. Like the other domains, the statements were selected using the M-STAR standards from this domain and further narrowed by identifying those items, which may vary according to children's developmental levels. Surprisingly, every mean in this domain, except for engaging in cooperative learning at the high school level, earned means over 4.0. Administrators at all levels agreed that the instructional standards were appropriate for their grade levels according to this measurement. Appendix G indicates the means and standard deviations of responses in this section.

Section 6 of the questionnaire contained six statements designed to measure each administrator's perception of the appropriateness of the M-STAR Learning Environment Domain for his or her school's grade levels. The statements were selected using the M-STAR standards from this domain and further narrowed by identifying those items, which may vary according to children's developmental levels. The results show the lowest mean as 3.86 for middle school principals responding to whether teachers fostered student collaboration centered on celebrating diversity. The highest mean was recorded from elementary administrators indicating the importance of establishing a nurturing relationship. Appendix H indicates the means and standard deviations of responses in this section.

Section 7 of the questionnaire contained eight statements designed to measure each administrator's perception of the appropriateness of the M-STAR Professional Responsibilities Domain for his or her school's grade levels. As with the other domains, the statements were selected using the M-STAR standards from this domain and further narrowed by identifying those items, which may vary according to children's developmental levels. The lowest mean was found for teachers assuming a leadership role at district functions while the highest mean was 4.32 for teachers seeking out professional development opportunities. Appendix I indicates the means and standard deviations of responses in this section.

Statistical Results

This study was a quantitative investigation into whether a statistically significant difference existed for the independent variable of grade level and the dependent variable of administrators' beliefs regarding appropriateness of the M-STAR standards in each of

five domains: Planning, Assessment, Instruction, Learning Environment, and Professional Responsibilities. This study used data collected from questionnaires that were mailed to every standard school in Mississippi. An analysis of variance was used to determine if a statistically significant difference existed in the dependent variables.

A scale was developed to measure each domain of the M-STAR tool. Domain 1: Planning included seven items and produced a Cronbach Alpha of .853. Domain 2: Assessment included only three items. This scale had low reliability in the pilot study with a Cronbach Alpha of .564, which was much improved in the full study with a Cronbach Alpha of .799. Domain 3: Instruction contained eleven items and had a Cronbach Alpha of .957. Domain 4: Learning Environments covered seven items and produced a Cronbach alpha of .840. Finally, Domain 5: Professional Responsibilities incorporated eight items and demonstrated a Cronbach Alpha of .902. Thus, all the domains tested fell within acceptable measures of reliability.

A one-way between subjects ANOVA was conducted to compare the impact of grade level on perceived level of appropriateness in M-STAR Domain 1: Planning, Domain 2: Assessment, Domain 3: Instruction, Domain 4: Learning Environment, and Domain 5: Professional Responsibilities. None of the domains analyses resulted in a significant difference. There was not a significant difference for grade level on Domain 1: Planning for the three conditions [$F(2, 118) = 1.47$, $p = .235$]. There was not a significant difference for grade level on Domain 2: Assessment for the three conditions [$F(2, 118) = 523$, $p = .594$]. There was not a significant difference for grade level on Domain 3: Instruction for the three conditions [$F(2, 118) = 1.00$, $p = .370$]. There was not a significant difference for grade level on Domain 4: Learning Environment for the

three conditions [F(2, 118) = .948, p = .390]. There was not a significant difference for grade level on Domain 5: Professional Responsibilities for the three conditions [F(2, 118) = .832, p = .438]. Table 9 reports the means and standard deviations for the dependent variables.

Table 3

Descriptive Statistics for M-STAR Domain appropriateness for grade level

	Mean	Std. Dev.
M-STAR Domain 1: Planning		
Elementary School Level	3.93	1.01
Middle School Level	3.80	1.04
High School Level	3.68	1.06
M-STAR Domain 2: Assessment		
Elementary School Level	4.16	0.73
Middle School Level	4.01	0.99
High School Level	4.05	0.70
M-STAR Domain 3: Instruction		
Elementary School Level	4.31	0.65
Middle School Level	4.24	0.84
High School Level	4.12	0.84
M-STAR Domain 4: Learning Environment		
Elementary School Level	4.25	0.77
Middle School Level	4.07	0.91
High School Level	4.10	0.87
M-STAR Domain 5: Professional Responsibilities		
Elementary School Level	4.19	0.73
Middle School Level	4.07	0.87
High School Level	4.08	0.73

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

In addition to assessing the degree to which elementary, middle and high school administrators perceived M-STAR to be an appropriate measure of teacher quality for their particular grade levels and assessing if the grade level administrators differed in their perceptions, this study also investigated the degree to which administrators perceived M-STAR to be aligned with their own ISLLC standards. Table 4 reports the mean and standards deviations for this question. Administrators showed consensus on this topic and generally agreed that the M-STAR instrument aligned to their own professional standards.

Table 4

Descriptive Statistics for M-STAR alignment to ISLLC standards

	Mean	Std. Dev.
ISLLC Alignment		
Elementary School	3.56	1.03
Middle School	3.60	1.01
High School	3.55	0.92

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

Finally, the questionnaire posed the same statement at the beginning of each domain, "For my grade levels, a perfect score in this domain is likely for a teacher in any subject area." Responses to this statement indicate the level of teacher efficacy using the M-STAR instrument. The standard deviations in this section were all over 1.14, which indicates a wider range of responses to this question compared to most others. Table 5 reports the mean and standards deviations for this question.

Table 5

Descriptive Statistics for teacher efficacy using the M-STAR tool

	Mean	Std. Dev.
Teacher Efficacy Using M-STAR		
Elementary School	2.73	1.20
Middle School	3.03	1.14
High School	2.76	1.18

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

When considering the data as a whole, some interesting patterns emerge. First, 57% administrators who responded to the questionnaire had more than five years of administrative experience, and approximately 70% had over five years of teaching experience prior to becoming an administrator. Only 5% of responding administrators were first-year M-STAR evaluators. Finally, elementary school principals responded in the greatest number followed by high school then middle school. This response rate reflects the ratio of the total number of elementary, middle, and high schools. In Mississippi, elementary schools make up 48% of the total number of schools, middle schools represent 21% of all schools, and high schools encompass 31% of all Mississippi schools.

The administrators' responses on items regarding M-STAR's alignment to ISLLC standards establish a pattern of consistency. The mean scores of individual items for elementary, middle, and high school administrators' beliefs regarding M-STAR alignment with ISLLC standards differed by less than two tenths when asked about M-STAR's effectiveness as a management tool that promotes collaboration as a faculty and is conducive to staff professional growth. All grade levels of administrators indicated

their lowest scores regarding M-STAR's alignment with the standard regarding the rubric's promotion of school faculty responding to diverse community needs. This item was the only one in this section with mean scores of less than three with both elementary and middle school administrators rating this item less than three.

Unlike the rare mean scores of less than three in other areas, the mean scores of items that assessed the likelihood for teachers to be awarded a perfect score on any domain were most often less than three. This response indicates that teachers are unlikely to be awarded a perfect score in any domain. Middle school principals gave higher scores on every question in this section than their elementary or middle school counterparts. Generally, elementary school principals recorded the lowest scores apart from one item.

When analyzing the mean scores for the individual items in each M-STAR domain some interesting patterns emerge. Within the Planning Domain, elementary administrators generally gave higher scores while the high school principals gave the lowest scores. This pattern continues is other domains as well. In two of the three items in the Assessment domain, elementary school principals gave the highest scores, and high school principals gave the lowest scores. The Instruction domain had eleven items, the most in any one domain. From the eleven items, elementary school principals indicated the highest mean scores on nine items while high school principals indicated the lowest mean scores on nine items as well. In the Learning Environment domain, elementary administrators again awarded the highest scores, but the lowest were distributed among middle and high school principals evenly. In the final domain, Professional

Responsibilities, elementary school principals gave the highest scores in all eight items while middle school principals awarded the lowest scores in seven of the eight items.

Finally, the overall mean scores for the M-STAR domains, alignment to ISLLC standards, and the teacher efficacy section continue to support a similar pattern. Elementary school administrators gave the overall highest mean score in each domain and the lowest overall mean score in the likelihood that a teacher would receive a perfect score in any single domain. The high school and middle school administrators each had some of the lowest mean scores on M-STAR's domains and teacher efficacy.

CHAPTER V – DISCUSSION

The purpose of this study was to examine the beliefs of administrators at the elementary, middle, and high school levels regarding the appropriateness of the M-STAR instrument for teachers of all grade levels. The research questions were developed to discover any differences in their perceptions based on grade level. Additionally, the researcher hoped to discover administrators' perceptions of M-STAR's alignment with their own professional standards, which are enumerated by the Council of Chief State School Officers as ISLLC Standards. Finally, this research sought to gauge administrators' beliefs regarding teacher efficacy using the M-STAR instrument. By discovering areas where statistical differences existed, leaders at the Mississippi Department of Education could identify any needed differentiation in the evaluation criteria for teachers according to level of school placement, as well as any areas of divergence with the ISLLC standards based on the administrators' perceived levels of appropriateness for these components. On the other hand, a lack of statistically significant differences could support the continued use of the M-STAR instrument at all levels because administrators perceive it to be as appropriate for teachers at their grade level. This chapter provides a summary of procedures used, a discussion of the findings, conclusions, implications, and recommendations for future research.

Summary of Procedures

After receiving permission from the University of Southern Mississippi Institutional Review Board, and conducting a pilot study, a questionnaire was distributed by mail to a sample population of public school administrators in all counties of Mississippi. One questionnaire was mailed to each standard elementary school, middle

school, and high school in the state. As discussed in Chapter III, Attendance Centers were excluded from the study. Of the 761questionnaires distributed to Mississippi principals, 121 (15.9%) forms were returned by participants who volunteered their responses between June 15, 2016, through August 15, 2016. The questionnaire (Appendix B) posed descriptive data questions in order to measure how the responses varied according to grade level placement. Cronbach's alpha was calculated for items in each section to measure the reliability of the items used to analyze the data. Further, the data from the questionnaire were analyzed to determine if administrators differed by their grade level assignments in their perceived level of appropriateness of the M-STAR domains using a one-way analysis of variance. Finally, their perception of M-STAR's alignment with ISLLC standards and teacher efficacy using M-STAR were analyzed using descriptive statistics.

Summary of Data

Demographic data from the responding administrators regarding their positions, their schools' grade levels, administrative experience, teaching experience, and M-STAR experience were analyzed in order to gain insight into the participants. Of the 121 respondents, 82.6% were school principals, 14.9% were assistant principals and 2.5% were some other administrator like a lead teacher. Regardless of their position, all of these administrators were tasked with evaluating teachers using the M-STAR instrument. Overwhelmingly, the majority of participants were executive level administrators. More elementary administrators participated than any other group with 43.8%. Middle school administrators participated at a rate of 24%, while 32.2% of respondents were high school principals. These participation rates indicate that both elementary and secondary

voices were heard in this study. With respect to administrative experience, most administrators, 43%, had less than five years of administrative experience while 30.6% reported between six and ten years of experience. Slightly more than a quarter of respondents, 26.6%, had eleven or more years of administrative experience. Conversely, almost half of respondents, 45.5%, reported eleven or more years of previous teaching experience. Administrators with between six and ten years of teaching experience included 34.7% of respondents, and only 19.8% of participants had five or fewer years of teaching experience. These results would indicate an experienced group of administrators who understand the dynamics of classroom teaching through their long-ranging instructional experience. Further, when compared to the total average educational experience of Mississippi administrators, this study's sample is representative of the state as a whole (Mississippi Department of Education, 2015).

At the time of this study, M-STAR had been available for four years. Of the administrators responding, 26.4% had worked with the M-STAR instrument since its inception. The bulk of administrators, 68.6%, reported between two and three years working with M-STAR. Only 5% of participants reported they were first-year M-STAR evaluators. From these demographics, it is evident that the participants were very familiar with the M-STAR domains and standards because of their number of years of experience.

The first research question posed in this study was: "To what degree do elementary school administrators believe the overall M-STAR evaluation tool, its various domains, and its alignment to ISLLC standards are appropriate for teachers at the elementary level?". Elementary school administrators showed confidence in all domains

of the M-STAR instrument with means ranging from 3.93 to 4.31. They seemed to reach the greatest consensus with Domain 3: Instruction with the highest mean of 4.31 and a standard deviation of 0.65. The descriptive data indicated agreement that M-STAR is aligned with ISLLC standards. The mean response for all items in this section was 3.56, and most participants agreed that Domain 3 was appropriate for their grade level. However, with a standard deviation of 1.03, the ranges of data varied from disagree to strongly agree. This rate indicates a level of variance in their answers that would warrant further study.

The second research question posed in this study was: "To what degree do middle school administrators believe the overall M-STAR evaluation tool, its various domains, and its alignment to ISLLC standards are appropriate for teachers at the middle school level?". Middle school administrators showed slightly less confidence in the appropriateness of M-STAR domains with means ranging from 3.80 to 4.24. They also seemed to reach the greatest consensus with Domain 3: Instruction with the highest mean of 4.24, but a slightly larger standard deviation of 0.84. The descriptive data indicated middle school administrators were also in agreement that M-STAR is aligned with ISLLC standards. The mean response for all items in this section was 3.60, which would fall into the agree category. However, with a standard deviation of 1.01, the ranges of data varied from neutral to strongly agree. Their answers did not vary by grade level, but this standard deviation warrants further study.

The third research question posed in this study was: "To what degree do high school administrators believe the overall M-STAR evaluation tool, its various domains, and its alignment to ISLLC standards are appropriate for teachers at the high school

level?". High school administrators showed the least confidence in the appropriateness M-STAR domains with means ranging from 3.68 to 4.12. However, like the elementary and middle school administrators, they also seemed to demonstrate the greatest confidence in Domain 3: Instruction with the highest mean of 4.12, and a standard deviation of 0.84. The descriptive data indicated high school administrators were also in agreement that M-STAR is aligned with ISLLC standards. The mean response for all items in this section was 3.55. However, with a standard deviation of 0.92, the ranges of data varied from neutral to agree.

The fourth and final research question asked: "How do perceived levels of appropriateness between elementary, middle, and high school administrators differ?". The responses were analyzed using a one-way analysis of variance to find if a statistically significant difference existed between the independent variable of administrator grade level placement and the dependent variable appropriateness of M-STAR Domains for grade level. The statistical analysis revealed that there was no statistically significant difference between elementary, middle, and high school administrators' level of confidence in M-STAR domains. The significance value for Domain 1: Planning was .235, which indicated this was the domain where administrators differed most, but it was not statistically significant. Domain 2: Assessment was the area in which administrators differed least.

Discussion

In 2013, Mississippi adopted the teacher evaluation system referred to as M-STAR. Although the goal of this system was to differentiate the effectiveness of teachers, the rubric it relied on to assess educators was identical for all grade levels

(Mississippi Department of Education, 2014a). For example, Advanced Placement Statistics teachers were evaluated using the same rubric as drama teachers. Elementary physical education teachers were evaluated with the same rubric as middle school social studies teachers. This study was designed to assess the differing perspectives of elementary, middle, and high school administrators regarding the effectiveness of the M-STAR rubric for all teachers at their grade levels. Although there were no statistically significant differences between administrators at various levels of schooling, testing for those differences provides evidence supporting the effectiveness of the M-STAR instrument at all levels of education.

Research on teacher evaluation revealed that historically administrators have viewed teachers as equally effective and interchangeable (Levin, Mulhern & Schunck, 2005; Weisberg et al., 2009). This attitude is much of what spurred national evaluation reforms. This study posed the same statement at the beginning of each domain: "For my grade level, a perfect score in this domain is likely for a teacher in any subject area." The results in this section varied widely. The standard deviations were among the largest in the entire study, but the means were the lowest in the study. Most responses fell in the neutral category, but when accounting for standard deviations, they ranged from disagree to agree. If the goal of education reform was to differentiate among teachers, a wide variety of responses would seem to support this goal rather than all teachers being rated good or excellent.

The items regarding the alignment of M-STAR with ISLLC standards provides an important cross-reference between evaluation reform measures and the professional standards of school administrators. Although there are now ten updated standards for

educational leaders (Council of Chief State School Officers, 2015), there were six ISLLC standards (Council of Chief State School Officers, 2008) at the time this study was designed. ISLLC standards guided professional administrators regarding best practices. The standards on the questionnaire included implementing a shared vision, implementing instructional program, promoting student learning and staff professional growth, providing a safe, efficient, and effective environment, respecting diversity and collaborating with the community, acting ethically, and responding to larger cultural issues (Council of Chief State School Officers, 2008). The alignment of administrators' professional standards with the tools they use in their practice is a crucial point of cross-reference. The means and standard deviations associated with administrator responses in this section indicate that they varied between neutral and strongly agree on all items. These results indicate that administrators perceive the teacher evaluation rubric to be conducive to meeting their own professional standards. M-STAR works in harmony with their role as instructional leaders as defined by their own professional standards.

Limitations

This study was limited to the population from which the sample was taken, Mississippi school administrators. It relied on a self-reporting instrument and was limited to those who volunteered to participate. Mayer (1999) found that sometimes teachers recorded responses that they felt were acceptable to their superiors or colleagues. Likewise, administrators may have felt a similar pressure to validate the evaluation reform put forth by Mississippi. In addition, the process of answering surveys may involve complex thought processes that are not captured by the instrument (Desimone and LeFloch, 2004). Another threat to the validity of the research instrument is that

questionnaire items may have contained more than one construct. Finally, the length of the questionnaire was a limitation. It was a two-sided document with 48 items and five descriptive questions. Five questionnaires were discarded because the back was not completed, and another ten questionnaires were discarded because the descriptive questions were not completed. This number of incomplete questionnaires represents 12% of the returned sample from the survey. Further, the responses on the back seemed to take on a homogeneity not seen on the front of the document.

Although M-STAR has undergone some revisions since this study, it is still based on the framework by Charlotte Danielson, which is used nationally in many schools. The results of this study cannot be generalized to a larger population because the sample participants were from schools in Mississippi only. Participants' previous experiences as educators, evaluators, and those being evaluated could not be controlled. These biases could have skewed the results. Finally, respondents were not given the option to explain their responses or make comments. Some studies suggest that cognitive interviews in addition to questionnaires would yield a fuller picture of the desired data (Desimone & LeFloch, 2004).

Implications

The results of this study are consistent with the much of the other literature surrounding teacher evaluation rubrics. The Mississippi Statewide Appraisal Rubric is based on Charlotte Danielson's Framework for Teaching (Mississippi Department of Education, 2014a). According to this body of work (Danielson, 2012; Danielson and McGreal, 2000; The Danielson Group 2013; Marshall. 2013; Marzano & Toth, 2013), although the context and subject matter may vary, good teaching can be characterized by

generalized standards. Danielson states that the Framework for Teaching is a valid evaluation rubric for teachers from kindergarten through twelfth grade. She goes on to claim that it can be equally effective for all subject areas from science to art (The Danielson Group, 2013). Danielson created her framework of four domains by referencing the Interstate Teacher Assessment and Support Consortium Standards (InTASC) (The Danielson Group, 2013). InTASC enumerates ten standards for all grade levels and subjects: Learner Development, Learning Differences, Learning Environment, Content Knowledge, Application of Content, Assessment, Planning for Instruction, Instructional Strategies, Professional Learning and Ethical Practice, and Leadership and Collaboration (Council of Chief State School Officers, 2015). These ten standards are evident in Danielson's domains: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities. Mississippi includes all of Danielson's domains with the addition of Assessment as a fifth domain. Although the administrators in this study seemed to agree that the standards were appropriate across grade levels, it is important to note, as discussed in Chapter II, that these standards are based on the constructivist theory of learning. Within that paradigm, they may seem appropriate, but they may not be appropriate from other learning theory perspectives. For example, in a teacher-centered approach, standards requiring student involvement in the development of lesson plans would not likely be included.

Although much of the literature supports the validity of the Danielson-type teacher evaluation rubrics, a few studies have begun to look closely at the appropriateness of teaching standards as they relate to children of various grade levels. Since the purpose of teacher evaluation is to differentiate among teacher effectiveness in order to provide

appropriate professional development, most studies assume the relationship between teacher observations and teacher effectiveness is a linear one (Lazarev & Newman, 2013). However, when teacher effectiveness is calculated using student standardized test scores as the dependent variable, some studies reveal a nonlinear relationship between the results from teacher observation rubrics and performance data for students at certain grade and developmental levels (Lazarev & Newman, 2013). Further, studies have shown that early childhood educators from PreK through third grade have at least fifteen major observation rubrics in use, including Framework for Teaching. Some of these rubrics use fewer generic teaching standards and more developmentally specific criteria, which is helping to pinpoint deficits in the children's achievement (Guernsey & Ochshorn, 2011). These new studies indicate that when given an option, more developmentally targeted teaching standards have an impact on teachers and students. In this study, administrators were not given developmentally specific alternatives to current standards, which may have resulted in their consensus around the appropriateness of current M-STAR domains. Further, this study did not include any items regarding student outcomes, which may have helped to frame the idea of grade level appropriateness.

Although administrators seemed to agree on the appropriateness of the M-STAR standards for the grade levels of their schools, there may be differences in the degree of emphasis for various grade levels. For example, although there was a high level of agreement that cooperative learning activities are appropriate at all grade levels, the amount of these activities may vary by developmental level of the learners. This example is evident even in the seating arrangements often seen in elementary versus middle or

high school classrooms. Most elementary classrooms employ permanent group seating arrangements while high schools move seats into groups occasionally as needed. Despite the effort to standardize teaching and make teacher evaluation more objective, there is still a great deal of room for subjectivity. This quality may be why administrators generally agreed upon the appropriateness of the standards for teachers of all grade levels. Administrators still have the ability to prioritize and emphasize their goals as instructional leaders. Although the generalized nature of the standards makes it possible for administrators to exercise their discretion, it may also prevent true differentiation between effective and distinguished teachers. The one-size-fits-all best practices reflected in the rubrics may help to identify ineffective teachers, but may not be specific enough to highlight truly developmentally appropriate strategies.

Recommendations for Future Research

There has been very little research on the grade level appropriateness of teacher evaluation standards since the widespread teacher evaluation reform and implementation of one-size-fits-all rubrics. The results of this study did not show significant differences among administrator beliefs regarding appropriateness for various grade levels regarding M-STAR domains. Further, the descriptive statistics indicate that administrators perceive M-STAR to be aligned with their own professional standards and seem to be evaluating teachers with a variety of scores. Although this study did not identify glaring differences between the appropriateness of teacher standards by grade level, there may be some worthwhile research to gain further clarification on these issues.

1. Although this study did attempt to sample one administrator from every school in Mississippi, the actual response rate was mush lower. Similar

studies should be conducted using a larger sample in order to obtain more results that are generalizable.

2. This study surveyed the beliefs of administrators only. Replication of this study with a sample of teachers rather than administrators would add a fuller contextual picture of grade level appropriateness.
3. This is the type of study that may benefit from a qualitative component because the topics are complex and may not be adequately captured in a survey. For this reason, conducting cognitive interviews to explore administrator perceptions may be very helpful.

Dear Potential Participant,

I would like to ask you or your designee to consider participating in a study. The purpose of this study is to gather data concerning the appraisal by administrators of the appropriateness of M-STAR standards for their specific site's grade levels. You may complete the questionnaire personally, or you may select another administrator familiar with M-STAR to provide the needed feedback. Participating in this study will afford you or your designee with the opportunity to reflect on your own views of M-STAR standards as they relate to the developmental stages of children at your school site. The study has the potential to affect educational practice and thereby be of benefit to children and society at large. Participation involves minimal anticipated risk.

The attached questionnaire covers six domains related to M-STAR as well as basic demographic information. Completion of the questionnaire should take no more than 10-15 minutes. All data collected will be anonymous. Please do not put your name or any other identifying information on the questionnaire. Any information inadvertently obtained during the course of this study will remain completely confidential. Participation in this project is completely voluntary. Please feel free to decline participation or to discontinue participation at any point without concern over penalty, prejudice, or any other negative consequence. Data will be aggregated and summary reports will be submitted by the researchers for a graduate dissertation at the University of Southern Mississippi. Upon completion of data compilation, all questionnaires will be destroyed. If you have questions concerning this research, please contact Danette Moore at (601) 569-3051 or danette.moore@eagles.usm.edu. This research is being conducted under the supervision of Ann Blankenship, PhD, J.D.

This project has been reviewed by the Institutional Review board, which ensures that research projects involving human subjects follow federal regulations. Any questions or concerns about rights as a research participant should be directed to the Chair of the Institutional Review Board, The University of Southern Mississippi, 118 College Drive #5147, Hattiesburg, MS 39406-0001, (601) 266-6820.

By completing and returning the attached questionnaire, the respondent gives permission for this anonymous and confidential data to be used for the purposes described above.

Thank you for your consideration.

Sincerely,

Danette Moore, EdS, NBCT

APPENDIX B Survey Instrument

M-STAR Survey

Given the M-STAR observation rubric is identical for all levels of education, the purpose of this questionnaire is to determine any differences among administrators' perceptions of appropriateness of M-STAR domains according to instructional level. It will also be used to determine administrators' perceptions of M-STAR's overall alignment with ISLLC standards. **Please complete both sides of the survey.** Your participation is greatly appreciated.

School grade levels: _____ Your position: _____
Does your school use the MSTAR? ___
Years you have used MSTAR: _____
Years of administrative experience: _____
Years of teaching experience: _____

Directions:
Each of the following questions are followed by a scale ranging from strongly disagree agree to strongly agree. Every question is related to your specific instructional level only. Mark the number to indicate your response.

1 = Strongly Disagree *2 = Disagree* *3 = Neutral* *4 = Agree* *5 = Strongly Agree*

1.	M-STAR is an effective management tool that promotes an efficient school.	1	2	3	4	5
2.	M-STAR promotes collaborating with faculty	1	2	3	4	5
3.	M-STAR promotes responding to diverse community needs	1	2	3	4	5
4.	M-STAR is fair and promotes acting with integrity and ethics	1	2	3	4	5
5.	M-STAR reflects an achievable standard of excellence for a teacher of any subject area	1	2	3	4	5
	M-STAR reflects:					
6.	Our school's vision of learning	1	2	3	4	5
7.	A school conducive to student learning	1	2	3	4	5
8.	A school conducive to staff professional growth	1	2	3	4	5
	Domain 1: Planning					
9.	A perfect score in this domain is likely for a teacher in any subject area.	1	2	3	4	5
	For my instructional level, all teachers should consistently include all of the following in their lesson plans:					
10.	Documentation of collaboration with specialists for differentiation	1	2	3	4	5
11.	Collaboration with students to design lessons	1	2	3	4	5
12.	Methods of data use to determine instructional goals	1	2	3	4	5
13.	Documentation of alignment with the state's standards	1	2	3	4	5
14.	A culminating performance task	1	2	3	4	5
15.	Documentation of data sources used to select instructional goals and strategies	1	2	3	4	5
16.	Documentation of cooperative learning	1	2	3	4	5
	Domain 2: Assessment					
17.	For my instructional level, a perfect score in this domain is likely for a teacher in any subject area.	1	2	3	4	5
	For my instructional level, all teachers should consistently:					
18.	Develop or select a formative assessment or pre-test for lessons	1	2	3	4	5
19.	Seek appropriate ways to use technology in assessment	1	2	3	4	5
20.	Develop/select summative assessments that verify learning of basic facts	1	2	3	4	5

Domain III: Instruction					
21. For my instructional level, a perfect score in this domain is likely for a teacher in any subject area.	1	2	3	4	5
For my instructional level, all teachers should consistently:					
22. Assist students in developing understanding by connecting the content to other subject areas	1	2	3	4	5
23. Assist students in developing understanding by applying content to solving timely, real-world problems	1	2	3	4	5
24. Engage students in cooperative learning activities.	1	2	3	4	5
25. Seek appropriate ways to use technology in instruction.	1	2	3	4	5
26. Link content to student interests	1	2	3	4	5
27. Use questions, coaching, and feedback that elicit extensive participation and discussion	1	2	3	4	5
28. Use questions requiring higher order thinking	1	2	3	4	5
29. Use questions to help students make connections to other students' comments	1	2	3	4	5
20. Use questions to probe for further discussion	1	2	3	4	5
31. Clearly connect instruction to students' prior knowledge	1	2	3	4	5
32. Clearly connect instruction to students' daily lives, and to aspects of their community and life experiences.	1	2	3	4	5

Domain IV: Learning Environment					
33. For my instructional level, a perfect score in this domain is likely for a teacher in any subject area.	1	2	3	4	5
For my instructional level, all teachers should consistently:					
34. Organize/use space, materials, and resources to facilitate movement	1	2	3	4	5
35. Organize/use space, materials, and resources to facilitate communication	1	2	3	4	5
36. Facilitate student collaboration by encouraging students to recognize diverse experiences of students	1	2	3	4	5
37. Facilitate student collaboration by encouraging students to celebrate diverse experiences of students	1	2	3	4	5
38. Engage students in monitoring classroom rules and behaviors	1	2	3	4	5
39. Have a nurturing relationship with students	1	2	3	4	5

Domain V: Professional Responsibilities					
40. For my instructional level, a perfect score in this domain is achievable for a teacher in any subject area.	1	2	3	4	5
For my instructional level, all teachers should consistently:					
41. Assume a leadership or supporting role within the professional learning community	1	2	3	4	5
42. Assume a leadership or supporting role for school events and projects	1	2	3	4	5
43. Assume a leadership or supporting role for district events and projects	1	2	3	4	5
44. Engage families in the instructional program and class activities	1	2	3	4	5
45. Incorporate student/family feedback in instructional content/activities when appropriate and reasonable	1	2	3	4	5
46. Seek out and participate in professional development	1	2	3	4	5
47. Make a substantial contribution to the profession	1	2	3	4	5
48. Lead or collaborate with colleagues to ensure full compliance with school and district regulations	1	2	3	4	5

THE UNIVERSITY OF
SOUTHERN MISSISSIPPI

INSTITUTIONAL REVIEW BOARD
118 College Drive #5147 | Hattiesburg, MS 39406-0001
Phone: 601.266.5997 | Fax: 601.266.4377 | www.usm.edu/research/institutional.review.board

NOTICE OF COMMITTEE ACTION

The project has been reviewed by The University of Southern Mississippi Institutional Review Board in accordance with Federal Drug Administration regulations (21 CFR 26, 111), Department of Health and Human Services (45 CFR Part 46), and university guidelines to ensure adherence to the following criteria:

- The risks to subjects are minimized.
- The risks to subjects are reasonable in relation to the anticipated benefits.
- The selection of subjects is equitable.
- Informed consent is adequate and appropriately documented.
- Where appropriate, the research plan makes adequate provisions for monitoring the data collected to ensure the safety of the subjects.
- Where appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of all data.
- Appropriate additional safeguards have been included to protect vulnerable subjects.
- Any unanticipated, serious, or continuing problems encountered regarding risks to subjects must be reported immediately, but not later than 10 days following the event. This should be reported to the IRB Office via the "Adverse Effect Report Form".
- If approved, the maximum period of approval is limited to twelve months.
  Projects that exceed this period must submit an application for renewal or continuation.

PROTOCOL NUMBER: 15120102
PROJECT TITLE: Mississippi's Teacher Observation Rubric: Administrator Appraisals of Appropriateness by Grade Level
PROJECT TYPE: New Project
RESEARCHER(S): Danette Moore
COLLEGE/DIVISION: College of Education and Psychology
DEPARTMENT: Educational Leadership and School Counseling
FUNDING AGENCY/SPONSOR: N/A
IRB COMMITTEE ACTION: Expedited Review Approval
PERIOD OF APPROVAL: 01/13/2016 to 01/12/2017
Lawrence A. Hosman, Ph.D.
Institutional Review Board

APPENDIX D Teacher Efficacy

Table A1.

Descriptive Statistics for teacher efficacy in each M-STAR domain

Statement	N	Mean	Std. Dev.
9. For my grade level, a perfect score in this domain [Planning] is likely for a teacher in any subject area.	121	2.66	1.13
Elementary School Level	53	2.60	1.12
Middle School Level	29	2.76	1.07
High school Level	39	2.67	0.87
17. For my grade level, a perfect score in this domain [Assessment] is likely for a teacher in any subject area.	121	2.72	1.11
Elementary School Level	53	2.63	1.13
Middle School Level	29	2.86	1.07
High school Level	39	2.72	1.08
21. For my grade level, a perfect score in this domain [Instruction] is likely for a teacher in any subject area.	121	2.66	1.14
Elementary School Level	53	2.63	1.19
Middle School Level	29	2.86	1.07
High school Level	39	2.54	1.08
33. For my grade level, a perfect score in this domain [Learning Environment] is likely for a teacher in any subject area.	121	3.02	1.27
Elementary School Level	53	2.89	1.27
Middle School Level	29	3.31	1.21
High school Level	39	3.00	1.28

Table A1 (continued).

40. For my grade level, a perfect score in this domain [Professional Responsibilities] is likely for a teacher in any subject area.	121	2.99	1.23
Elementary School Level	53	2.87	1.26
Middle School Level	29	3.34	1.12
High School Level	39	2.90	1.22

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

APPENDIX E M-STAR Planning Domain

Table A2.

Descriptive Statistics for the M-STAR Planning Domain

Statement	N	Mean	Std. Dev.
For my grade levels, all teachers should consistently include the following in their lesson plans:			
10. Documentation of collaboration with specialists for differentiation.	121	3.26	1.09
Elementary School Level	53	3.34	1.08
Middle School Level	29	3.17	0.99
High School Level	39	3.21	1.18
11. Collaboration with students to design lessons.	121	3.04	1.09
Elementary School Level	53	2.96	1.03
Middle School Level	29	3.17	1.08
High School Level	39	3.05	1.18
12. Methods of data used to determine instructional goals.	121	4.07	0.90
Elementary School Level	53	4.25	0.77
Middle School Level	29	4.07	0.94
High School Level	39	3.82	0.96
13. Documentation of alignment with state standards.	121	4.35	0.79
Elementary School Level	53	4.53	0.63
Middle School Level	29	4.28	0.91
High School Level	39	4.15	0.83

Table A2 (continued).

14. A culminating performance task.	121	4.04	0.89
Elementary School Level	53	4.09	0.92
Middle School Level	29	4.07	0.83
High School Level	39	3.95	0.88
15. Documentation of data sources used to select instructional goals and strategies.	121	3.90	0.91
Elementary School Level	53	4.08	0.89
Middle School Level	29	3.90	0.92
High School Level	39	3.67	0.89
16. Documentation of cooperative learning.	121	4.06	0.80
Elementary School Level	53	4.23	0.63
Middle School Level	29	3.97	0.96
High School Level	39	3.90	0.98

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

APPENDIX F – Assessment Domain

Table A3.

Descriptive Statistics for the Assessment Domain

Statement For my grade levels, all teachers should consistently:	N	Mean	Std. Dev.
18. Develop or select a formative assessment or pre-test.	121	4.05	0.84
Elementary School Level	53	4.15	0.76
Middle School Level	29	4.00	1.05
High School Level	39	3.95	0.75
19. Seek appropriate ways to use technology in assessment.	121	4.04	0.78
Elementary School Level	53	4.08	0.72
Middle School Level	29	3.90	0.96
High School Level	39	4.10	0.67
20. Develop or select summative assessments that verify learning of basic facts.	121	4.17	0.76
Elementary School Level	53	4.25	0.70
Middle School Level	29	4.14	0.94
High School Level	39	4.10	0.67

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

APPENDIX G Instruction Domain

Table A4.

Descriptive Statistics for the M-STAR Instruction Domain

Statement For my grade levels, all teachers should consistently:	N	Mean	Std. Dev.
22. Assist students in developing understanding by connecting the content to the subject area.	121	4.20	0.77
Elementary School Level	53	4.32	0.69
Middle School Level	29	4.21	0.80
High School Level	39	4.03	0.80
23. Assist students in developing understanding by applying content to solve timely, real-world problems.	121	4.24	0.78
Elementary School Level	53	4.36	0.70
Middle School Level	29	4.24	0.86
High School Level	39	4.08	0.80
24. Engage students in cooperative learning activities.	121	4.18	0.80
Elementary School Level	53	4.28	0.59
Middle School Level	29	4.28	0.87
High School Level	39	3.97	0.95
25. Seek appropriate ways to use technology in instruction.	121	4.21	0.69
Elementary School Level	53	4.28	0.63
Middle School Level	29	4.21	0.85
High School Level	39	4.10	0.78

Table A4 (continued).

26. Link content to student interests.	121	4.21	0.76
Elementary School Level	53	4.23	0.66
Middle School Level	29	4.31	0.83
High School Level	39	4.13	0.85
27. Use questions, coaching, and feedback that elicit extensive participation and discussion.	121	4.24	0.75
Elementary School Level	53	4.28	0.66
Middle School Level	29	4.21	0.85
High School Level	39	4.21	0.79
28. Use questions requiring higher order thinking.	121	4.38	0.75
Elementary School Level	53	4.49	0.60
Middle School Level	29	4.24	0.86
High School Level	39	4.33	0.83
29. Use questions to help students make connections to other students' comments.	121	4.11	0.76
Elementary School Level	53	4.15	0.68
Middle School Level	29	4.07	0.83
High School Level	39	4.08	0.80
30. Use questions to probe for further discussion.	121	4.34	0.73
Elementary School Level	53	4.42	0.63
Middle School Level	29	4.31	0.83
High School Level	39	4.26	0.78
31. Clearly, connect instruction to students' prior knowledge.	121	4.26	0.74
Elementary School Level	53	4.40	0.65
Middle School Level	29	4.24	0.77
High School Level	39	4.08	0.80

Table A4 (Continued).

32. Clearly connect instruction to students' daily lives and to aspects of their community experiences.	121	4.19	0.82
Elementary School Level	53	4.23	0.66
Middle School Level	29	4.28	0.83
High School Level	39	4.08	0.97

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

APPENDIX H Learning Environment

Table A5.

Descriptive Statistics for the M-STAR Learning Environment Domain

Statement For my grade levels, all teachers should consistently:	N	Mean	Std. Dev.
34. Organize/use space, materials, and resources to facilitate movement.	121	4.17	0.75
Elementary School Level	53	4.26	0.65
Middle School Level	29	4.00	0.87
High School Level	39	4.18	0.75
35. Organize/use space, materials, and resources to facilitate communication.	121	4.29	0.69
Elementary School Level	53	4.38	0.48
Middle School Level	29	4.17	0.83
High School Level	39	4.26	0.78
36. Facilitate student collaboration by encouraging students *recognize* diverse experiences for students.	121	3.99	0.90
Elementary School Level	53	4.04	0.85
Middle School Level	29	3.93	0.98
High School Level	39	3.97	0.92
37. Facilitate student collaboration by encouraging students *celebrate* diverse experiences for students.	121	3.93	0.92
Elementary School Level	53	4.02	0.90
Middle School Level	29	3.86	0.97
High School Level	39	3.87	0.91

Table A5 (continued).

38. Engage students in monitoring classroom rules and behaviors.	121	4.17	0.88
Elementary School Level	53	4.17	0.91
Middle School Level	29	4.17	0.83
High School Level	39	4.15	0.89
39. Have a nurturing relationship with the students.	121	4.40	0.78
Elementary School Level	53	4.64	0.52
Middle School Level	29	4.28	0.87
High School Level	39	4.15	0.89

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

APPENDIX I Professional Responsibilities

Table A6.

Descriptive Statistics for the M-STAR Professional Responsibilities Domain

Statement	N	Mean	Std. Dev.
For my grade levels, all teachers should consistently:			
41. Assume a leadership or supporting role within the professional learning community.	121	4.02	0.74
Elementary School Level	53	4.09	0.65
Middle School Level	29	4.07	0.91
High School Level	39	3.90	0.71
42. Assume a leadership or supporting role for school events and projects.	121	4.07	0.68
Elementary School Level	53	4.13	0.55
Middle School Level	29	4.00	0.83
High School Level	39	4.05	0.71
43. Assume a leadership or supporting role for district events and projects.	121	3.84	0.77
Elementary School Level	53	3.89	0.66
Middle School Level	29	3.79	0.85
High School Level	39	3.82	0.84
44. Engage families in the instructional program and class activities.	121	4.08	0.77
Elementary School Level	53	4.11	0.74
Middle School Level	29	4.03	0.85
High School Level	39	4.08	0.73

Table A6 (continued).

45. Incorporate student/family feedback in instructional content/activities when appropriate and reasonable.	121	4.01	0.77
Elementary School Level	53	4.06	0.81
Middle School Level	29	3.97	0.81
High School Level	39	3.97	0.66
46. Seek out and participate in professional development.	121	4.32	0.75
Elementary School Level	53	4.49	0.69
Middle School Level	29	4.10	0.84
High School Level	39	4.26	0.71
47. Make a substantial contribution to the profession.	121	4.26	0.78
Elementary School Level	53	4.36	0.75
Middle School Level	29	4.07	0.91
High School Level	39	4.26	0.67
48. Lead or collaborate with colleagues to ensure full compliance with school and district regulations.	121	4.31	0.72
Elementary School Level	53	4.38	0.71
Middle School Level	29	4.17	0.83
High School Level	39	4.33	0.61

Scale: 5 = Strongly Agree; 4 = Agree; 3 = Neutral; 2 = Disagree; 1 = Strongly Disagree

REFERENCES

Alliance for Childhood. (2002). *Time for play every day*. Retrieved from http://www.allianceforchildhood.org/publications

Bandura, A. (1982). Self-efficacy mechanisms in human agency. *American Psychologist*. *37*, 122-147.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Beachum, F. (2011). Culturally relevant leadership for complex 21st-century school contexts. In F. English (Ed.), *The sage handbook of educational leadership.* (2nd ed., pp. 26-35). Thousand Oaks, CA: SAGE Publications, Inc. doi: http://dx.doi.org/10.4135/9781412980036.n3

Bill and Melinda Gates Foundation. (2009). *Learning about teaching: Initial findings from the measures of effective teaching project.* Retrieved from http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf

Bill and Melinda Gates Foundation. (2013). Ensuring fair and reliable measures of effective teaching: Culminating findings from the met project's three-year study. Policy and Practice Brief. Retrieved from http://metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Bloom, B., Englehart, M., Frost, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York, NY: David McKay.

Bocock, B. (1986). *Hegemony*. London, UK: Tavistock.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in

entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy, 1*(2), 176-216.

Breault, R. (2013). "She was great but…": Examining preservice recollections of favorite and most effective teachers. *The Professional Educator (37)1*, 1-14.

Bridges, E. (1986). *The Incompetent Teacher*. Philadelphia, PA: Flamer Press.

Brooks, J., & Kensler, L. (2011). Distributed leadership and democratic community. In F. English (Ed.), *The sage handbook of educational leadership.* (2nd ed., pp. 56-68). Thousand Oaks, CA: SAGE Publications, Inc. doi: http://dx.doi.org/10.4135/9781412980036.n5

Bruce, C., Esmonde, I., Ross, J., Dookie, L., & Beatty, R. (2010). The effects of sustained classroom-embedded teacher professional learning on teacher efficacy and related student achievement. *Teaching and Teacher Education, 26*, 1598–1608.

Bruner, J. (1960). *The process of education.* New York, NY: Vintage Books.

Bush, T. (2009). Leadership development and school improvement : Contemporary issues in leadership development. *Educational Review, 61(4)*, 375-389.

Campbell, R. (2006). *Jean Piaget's genetic epistemology: Appreciation and critique.* Retrieved from http://myweb.clemson.edu/~campber/piaget.html

Common Core State Standards Initiative. (2015). *Development process*. Retrieved from http://www.corestandards.org/about-the-standards/development-process/

Council of Chief State School Officers. (2008). *Educational leadership policy standards: ISLLC 2008 as adopted by the national policy board for educational administration.* Retrieved from

http://www.ccsso.org/Documents/2008/Educational_Leadership_Policy_Standards_2008.pdf

Council of Chief State School Officers. (2015). *Educational leadership policy standards: ISLLC 2015 as adopted by the national policy board for educational administration.* Retrieved from http://www.ccsso.org/Documents/2015/ProfessionalStandardsforEducationalLeaders2015forNPBEAFINAL.pdf

Cranston, J. (2009). Holding the reins of the professional learning community: Eight themes from the research on principals' perceptions of professional learning communities. *Canadian Journal of Educational Administration and Policy*, *20,* 1-22.

Cranston, N. (2013). School leaders leading: Professional responsibility not accountability as the key focus. *Educational Management Administration & Leadership*, *41*(2), 129-142.

Cranston, N., Ehrich, L. C., & Kimber, M. (2006). Ethical dilemmas: The "bread and butter" of educational leaders' lives. *Journal of Educational Administration*, *44*(2), 106-121.

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

Danielson, C. (2012). It's your evaluation – collaborating to improve teacher practice.

Education Digest: Essential Readings Condensed for Quick Review (77)8, 22-27. Retrieved from http://www.njea.org/news-and-publications/njea-review/october-2011/its-your-evaluation

The Danielson Group. (2013). The framework. The Danielson Group. Retrieved 11/5/2014 from http://danielsongroup.org/framework/.

Desimone, L., & LeFloch, K. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis, 26(1),* 1-22.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process.* Boston, MA: Heath.

Doherty, K., & Jacobs, S. (2013). *State of the states 2013 connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. Retrieved from National Council on Teacher Quality website: http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

DuFour, R., & Marzano, R. (2011). *Leaders of learning: How district, school, and classroom leaders improve student achievement*. Bloomington, IN: Solution Tree.

DuFour, R., & Mattos, M. (2013). How do principals really improve schools?. *Educational Leadership 70(7),* 34-40.

Eacott, S. (2009). New look leaders or a new look at leadership: re-conceptualizing the principalship for the future of Australian education. Paper presented at the Australian Association for Research in Education International Conference, Canberra.

Educational Broadcasting Corporation. (2004). *Workshop: Constructivism as a paradigm for teaching and learning. Concept to Classroom*. Retrieved 11/1/2014 from http://www.thirteen.org/edonline/concept2class/constructivism/

Estes, C. (2004). Promoting student-centered learning in experiential education. *Journal of Experiential Education, 27,* 141-160.

Firestone, W., & Shipps, D. (2007). Orchestration, coherence and the problem of conflicting accountabilities. *Managing Change in the Public Services*, 212. Malden, MA: Blackwell Publishing.

Fitzgerald, T., & Gunter, H. (2008). Contesting the orthodoxy of teacher leadership. *International journal of leadership in education, 11(4),* 331-340.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.

Gargani, J., & Strong, M. (2014). Can we identify successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education (65)5*, 389-401. doi: 0.1177/0022487114542519

Glazerman, S., Golfhaber, D., Loeb, S., Raudenbush, S., Staiger, D., & Whitehurst, G. (2011). *Passing muster: Evaluating teacher evaluation systems*. Washington, DC: The Brookings Institute. Retrieved from http://www.brookings.edu/~/media/research/files/reports/2011/4/26-evaluating-teachers/0426_evaluating_teachers.pdf

Glickman, C., Gordon, S., & Ross-Gordon, J. (2010). *Supervision and instructional leadership: A developmental approach* (8th ed.). Boston, MA: Allyn & Bacon.

Goe, L., Holdheide, L., Miller, T. (2014). *Practical guide to designing comprehensive teacher evaluation systems: A tool to assist in the development of teacher evaluation systems.* Washington, DC: American Institutes for Research.

Goldstein, J., & Noguera, P. (2006). A thoughtful approach to teacher evaluation. *Educational Leadership*, *63*(6), 31-37.

Gordon R., Kane T., & Staiger D. (2006). "Identifying effective teachers using performance on the job." *Hamilton Project Discussion Paper*. Washington, D.C.: Brookings Institution.

Gramsci, A. (1971). *Selections from the prison notebooks*. London, UK: Lawrence and Wishart.

Guernsey, L. & Ochshorn, S. (2011). *Watching teachers work: using observation tools to promote effective teaching in the early years and early grades*. Washington, DC: New America Foundation.

Guilfoyle, C. (2013). Quantifying teacher effectiveness. *Policy Priorities: An Information Brief from ASCD, 19*(1), 1-7. Retrieved from http://www.ascd.org/publications/newsletters/policy-priorities/vol19/num01/Quantifying-Teacher-Effectiveness.aspx

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

Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education, 33*(3), 329-351.

Hallinger, P. (2005). "Instructional leadership and the school principal: A passing fancy

that refuses to fade away." *Leadership and Policy in Schools 4*(3), 1-20.

Hanover Research. (2012). *Best practices for including multiple measures in teacher evaluations.* Retrieved from http://www.hanoverresearch.com/wp-content/uploads/2012/05/Best-Practices-for-Including-Multiple-Measures-in-Teacher-Evaluations-Membership.pdf

Hardy, I. (2010). Critiquing teacher professional development: Teacher learning within the field of teachers' work. *Critical Studies in Education 51,* 71–84.

Hargreaves, A., & Shirley, D. (Eds.). (2009). *The fourth way: The inspiring future for educational change*. Thousand Oaks, CA: Corwin Press.

Harris, D., Ingle, W., & Rutledge, S. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal, 51*(1), 73-112.

Hopkins, J. (2011). The enduring influence of Jean Piaget. *Association for Psychological Science Observer (24)10.* Retrieved from http://www.psychologicalscience.org/index.php/publications/observer/2011/december-11/jean-piaget.html

Hopkins, R. (1994). *Narrative schooling: Experiential learning and the transformation of American education*. New York, NY: Teacher's College Press.

Horng, E., & Loeb, S. (2010). New thinking about instructional leadership. *Phi Delta Kappan, 92*(3), 66-69.

Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher*

performance. Retrieved from http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-Evaluation-Full-Report-PDF.pdf

Ingle, K., Rutledge, S., & Bishop, J. (2011). Context matters: Principals' sensemaking of teacher hiring and on-the-job performance. *Journal of Educational Administration, 49(5),* 579-610.

Johanningmeier, E. (2010). A nation at risk and Sputnik: Compared and reconsidered. American *Educational History Journal (37)*2, 347-365.

John Dewey Project on Progressive Education. (2002). A brief overview of progressive education. *The University of Vermont: College of Education and Social Services*. Retrieved from http://www.uvm.edu/~dewey/articles/proged.html

Joint Legislative Committee on Performance Evaluation and Expenditure Review. (2014). *The common core state standards: Mississippi's adoption and implementation*. Retrieved from http://www.peer.ms.gov/reports/rpt582.pdf

Jordan, H., Mendro, R., & Weerasinghe, D. (1997). *Teacher effects on longitudinal student achievement*. Paper presented at the sixth National Evaluation Institute sponsored by CREATE, Indianapolis, IN. Retrieved from http://www.dallasisd.org/cms/lib/TX01001475/Centricity/Shared/evalacct/research/articles/Jordan-Teacher-Effects-on-Longitudinal-Student-Achievement-1997.pdf

Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (Research Paper). Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from

http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

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

Kern, A. (2009). *Human capital development theory: Implications for education comparison of influential twenty-first-century economists* Samuel Bowles and Gary S. Becker. Retrieved from http://www.personal.psu.edu/afk119/blogs/career_tech_ed/2009/12/human-capital-development-theory.html

Knowles, M., Holton, E., & Swanson, R. (2005). *The adult learner: the definitive classic in adult education and human resource development* (6th ed.). Boston, MA: Elsevier.

Koutrouba, K. (2012). A profile of the effective teacher: Greek secondary education teachers' perceptions. European *Journal of Teacher Education (35)* 3, 359-374.

Lazarev, V. & Newman, D. (2013). How non-linearity and grade-level differences complicate the validation of observation protocols. Paper presented at Society for Research on Educational Effectiveness conference. Retrieved from https://eric.ed.gov/?id=ED563108

Lefgren, L. & Jacob, B. (2006). When principals rate teachers. *Education Next, 2*(6). Retrieved from http://educationnext.org/whenprincipalsrateteachers/

Levin, J., Mulhern, J., & Schunck, J. (2005). *Unintended consequences: The case for*

reforming staffing rules in urban teacher union contracts. Brooklyn, NY: The New Teacher Project.

Lord, T. (1999). A comparison between traditional and constructivist teaching in environmental science. *Journal of Environmental Education, 30*, 22-27.

Marshall, K. (2013). *Rethinking teacher evaluation and supervision.* San Francisco, CA: Jossey-Bass.

Marzano, R. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R., & Toth, M. (2013). *Teacher evaluation that makes a difference.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mascolo, M. (2009). Beyond student-centered and teacher-centered pedagogy: Teaching and learning as guided participation. *Pedagogy and the Human Sciences 1(1)*, 3-27.

Mascolo, M., & Fischer, K. (2004). Constructivist theories. In Hopkins, B., Barre, R., Michel, G., Rochat, P. (Eds.). *Cambridge encyclopedia of child development*. Cambridge, UK: Cambridge University Press.

Mayer, D. (1999). Measuring instructional practice: Can policymakers trust survey data?. *Educational evaluation and policy analysis, 2(1)*, 29-45.

McGuinn, P. (2012). *The state of teacher evaluation reform: State education agency capacity and the implementation of new teacher-evaluation systems.* Washington, DC: Center for American Progress. Retrieved from

https://cdn.americanprogress.org/wp-content/uploads/2012/11/McGuinn_TheStateofEvaluation-1.pdf

Meister, D. (2010). Experienced secondary teachers' perceptions of engagement and effectiveness: A guide for professional development. *The Qualitative Report, 15*(4), 880-898. Retrieved from hhtp://www.nova.edu/ssss/QR/QR 15-4/meister.pdf

Mississippi Department of Education. (2010). *Recommendation of the statewide teacher evaluation council*. Retrieved from http://www.mde.k12.ms.us/docs/federal-programs/11b_evaluation_council_final_recommendations.pdf?sfvrsn=2

Mississippi Department of Education. (2012). *Mississippi principal evaluation system: Standards and indicators.* Retrieved from http://www.mde.k12.ms.us/docs/cte/standards-and-indicators-for-principals.pdf?sfvrsn=0

Mississippi Department of Education. (2014a). *Mississippi statewide teacher appraisal rubric process manual*. Retrieved from http://www.mde.k12.ms.us/docs/teacher-center/revised-m-star-process-manual-july-2014.pdf?sfvrsn=2

Mississippi Department of Education. (2014b). *MS teacher evaluation system: Implementation timeline pie charts*. Retrieved from http://www.mde.k12.ms.us/docs/teacher-center/implementation-timeline-pie-charts.pdf?sfvrsn=2

Mississippi Department of Education. (2014c). *Mississippi teacher evaluation system updates.* Retrieved from http://www.mde.k12.ms.us/docs/teacher-center/teacher-evaluation-modifications-for-2014-2015.pdf?sfvrsn=2

Mississippi Department of Education. (2015). *Public reports*. Retrieved from http://reports.mde.k12.ms.us/

Mitchell, C., & Sackney, L. (2009). *Sustainable improvement: Building learning communities that endure*. Rotterdam, Neth.: Sense.

Mitchell, C., & Sackney, L. (2015). School improvement in high-capacity schools: Educational leadership and living systems ontology. *Educational Management Administration & Leadership,* 1-16. DOI: 10.1177/1741143214564772.

Moore, D. (2014). *Leadership and teacher evaluation.* Unpublished manuscript. University of Southern Mississippi, Hattiesburg, MS.

Muguti, J., & Mawere, V. (2013). Defining an effective teacher: An examination of characteristics that school heads associate with the concept. *International Journal of Social Science and Education (4)*1, 109-121.

Mulford, B., & Edmunds, B. (2009). *Successful school principalship in Tasmania*. Launceston, Tasmania: Faculty of Education, University of Tasmania.

National Council on Teacher Quality. (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies.* Retrieved from http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf

National Council on Teacher Quality. (2014). *Teacher evaluation policy in Mississippi: Where is Mississippi in implementing teacher effectiveness policies?* Retrieved from http://www.nctq.org/dmsView/Evaluation_Timeline_Brief_Mississippi

The New Teacher Project. (2009). *Human capital reform in Cincinnati public schools: Strengthening teacher effectiveness and support.* Retrieved from http://tntp.org/assets/documents/TNTP_Cincinnati_Report_Dec09.pdf

The New Teacher Project and Student Achievement Partners. (2013). *Fixing classroom observations: How common core will change the way we look at teaching.* New York, NY: Author. Retrieved from http://tntp.org/assets/documents/TNTP_FixingClassroomObservations_2013.pdf

Nye, B., Konstantopoulus, S., & Hedges, L. (2004). "How large are teacher effects?" *Educational Evaluation and Policy Analysis, 26*(3), 237-257.

OECD. (2015). *About the OECD*. Retrieved from http://www.oecd.org/about/

OECD. (2013). *Lessons from PISA 2012 for the United States, Strong Performers and Successful Reformers in Education.* OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264207585-en

Papa, R. (2011). Standards for educational leaders: Promises, paradoxes, and pitfalls. In F. W. English (Ed.), The SAGE handbook of educational leadership: Advances in theory, research, and practice (2nd ed.) (pp. 195-209). Los Angeles, CA: SAGE.

Petty, S. (1899). 1662. Treatise of taxes and contributions. In vol. 1 of *The economic writings of sir William Petty*, edited by CH Hull. Cambridge, UK: Cambridge University Press.

Phillips, D. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher, 24*, 5-12.

Piaget, J. (1954). *The origins of intelligence in children*. (M. Cook, Trans.). New York, NY: International University Press.

Powell, A. (2007). How Sputnik changed U.S. education. *Harvard Gazette*. Retrieved from http://news.harvard.edu/gazette/story/2007/10/how-sputnik-changed-u-s-education/

Ravitch, D. (2002). A brief history of testing and accountability. *Hoover Digest: Research and Opinion on Public Policy*, (4). Retrieved from http://www.hoover.org/research/brief-history-testing-and-accountability

Rivkin, S., Hanushek, E., & Kain, J. (2005). "Teachers, schools and academic achievement." *Econometrica, 73*(2), 417–458.

Rockoff, J. (2004). "The impact of individual teachers on student achievement: Evidence from panel data." *American Economic Review, 94*(2), 247-252.

Sahin, A., & Adiguzel, T. (2014). Effective teacher qualities from international mathematics, science, and computer teachers' perspectives. *Eurasia Journal of Mathematics, Science & Technology, 10*(6), 635-646.

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

Schalock, H. (1998). Student progress in learning: Teacher responsibility, accountability, and reality. *Journal of Personnel Evaluation in Education, 12*(3), 237–246.

Schultz, T. (1961) Investment in human capital. *American Economic Review 51*, 1-17.

Schooling, P. (2011). *The Marzano causal teacher evaluation model*. New Jersey

Education Association. Retrieved from https://www.njea.org/news-and-publications/njeareview/november-2011/comparing-teacher-evaluation-models/marzano

Sebastian, J., & Allensworth, E. (2012). The influence of principal leadership on classroom instruction and student learning: A study of mediated pathways to learning. *Educational Administration Quarterly, 48*(4), 626-623. doi: 10.1177/0013161X11436273

Simpson, T. (2002). Dare I oppose constructivist theory? *The Educational Forum, 66*(4), 2002.

Shanahan, M. (2010). Cross-sectional design. In *Encyclopedia of Case Study Research.* Retrieved from http://srmo.sagepub.com/view/encyc-of-case-study-research/n99.xml

Skinner, B.F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24*(2), 86-97.

Smith, A. (1937). *The wealth of nations [1776]*. New York, NY: Random House.

State University. (2015). *Edward L. Thorndike (1874–1949) - the man and his career, a psychology for educators, education as specific habit formation.* Retrieved from http://education.stateuniversity.com/pages/2509/Thorndike-Edward-L-1874-1949.html

Steup, M. (2005). Epistemology. *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/entries/epistemology/

Strong, M. (2011). *The highly qualified teacher: What is teacher quality and how do we measure it?* New York, NY: Teachers College Press.

Towndrow, P., & Tan, K. (2009). Teacher self-evaluation and power. *Teacher Development, 13(3)*, 285-295.

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

U. S. Department of Education. (1983). *A nation at risk: The imperative for educational reform*. Retrieved from http://www2.ed.gov/pubs/NatAtRisk/title.html

U. S. Department of Education. (2008). A nation accountable: Twenty-five years after A Nation at Risk. Retrieved from http://www2.ed.gov/rschstat/research/pubs/accountable/accountable.pdf

U. S. Department of Education. (2012). *The federal role in education*. Retrieved from http://www2.ed.gov/about/overview/fed/role.html

U. S. Department of Education. (2013). *The threat of educational stagnation and complacency: Remarks of U.S. Secretary of Education Arne Duncan at the release of the 2012 Program for International Student Assessment (PISA).* Retrieved from http://www.ed.gov/news/speeches/threat-educational-stagnation-and-complacency

U.S. Department of Education. (2014). *White House report: Race to the top setting the pace for gains across the education system*. Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2014/03/25/white-house-report-race-top-setting-pace-gains-across-education-system

U. S. Department of Justice. (2013). *Title VI of the Civil Rights Act of 1964*. Retrieved from http://www.justice.gov/crt/about/cor/coord/titlevi.php

University of Vermont. (2002). *John Dewey Project on Progressive Education*. Retrieved from http://www.uvm.edu/~dewey/articles/proged.html

Vygotsky, L. (1978). *Mind in Society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Walker, R. (2008). *12 characteristics of an effective teacher*. Morrisville, NC: Lulu Publishing.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project.

Whitehurst, G., Chingos, M., & Lindquist, K. (2015). Getting classroom observations right: Lessons from four pioneering districts. *Education Next, Winter (2015),* 62-68. Retrieved from http://educationnext.org/files/ednext_XV_1_research_whitehurst.pdf

Williams, C. (2007). Research Methods, *Journal of Business and Economic Research, 5*(3), p. 65-72.

Wilson, L., Horch, M., & Wilson, H. (2002). Implications of brain research for teaching young adolescents. What research says. *Middle School Journal, 34*(1), 57-61.

Wong, H., & Wong, R. (2012). *Developing and retaining effective teachers and principals*. Retrieved from http://newteacher.com/pdf/Developing.pdf

Woolfolk, A. (2007). *Educational Psychology* (10th ed.) Boston, MA: Allyn and Bacon.

Wright, S., Horn, S., & Sanders, W. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57–67.