

The University of Southern Mississippi  
**The Aquila Digital Community**

---

Dissertations

---

Spring 5-2016

**Teachers' Perceptions of the Mississippi Statewide Teacher  
Appraisal Rubric (M-STAR) Evaluation**

Steven Douglas Hampton  
*University of Southern Mississippi*

Follow this and additional works at: <https://aquila.usm.edu/dissertations>



Part of the [Educational Leadership Commons](#), and the [Elementary and Middle and Secondary Education Administration Commons](#)

---

**Recommended Citation**

Hampton, Steven Douglas, "Teachers' Perceptions of the Mississippi Statewide Teacher Appraisal Rubric (M-STAR) Evaluation" (2016). *Dissertations*. 315.  
<https://aquila.usm.edu/dissertations/315>

This Dissertation is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Dissertations by an authorized administrator of The Aquila Digital Community. For more information, please contact [Joshua.Cromwell@usm.edu](mailto:Joshua.Cromwell@usm.edu).

TEACHERS' PERCEPTIONS OF THE MISSISSIPPI STATEWIDE TEACHER  
APPRAISAL RUBRIC (M-STAR) EVALUATION

by

Steven Douglas Hampton

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

Approved:

---

Dr. David Lee, Committee Chair  
Associate Professor, Educational Leadership and School Counseling

---

Dr. Kyna Shelley, Committee Member  
Professor, Educational Studies and Research

---

Dr. Richard Mohn, Committee Member  
Associate Professor, Educational Studies and Research

---

Dr. James Fox, Committee Member  
Assistant Professor, Educational Leadership and School Counseling

---

Dr. Stanley Benigno, Committee Member  
Adjunct Professor, Educational Leadership and School Counseling

---

Dr. Karen S. Coats  
Dean of the Graduate School

May 2016

COPYRIGHT BY

Steven Douglas Hampton

2016

*Published by the Graduate School*



THE UNIVERSITY OF  
**SOUTHERN**  
**MISSISSIPPI.**

## ABSTRACT

### TEACHERS' PERCEPTIONS OF THE MISSISSIPPI STATEWIDE TEACHER APPRAISAL RUBRIC (M-STAR) EVALUATION

by Steven Douglas Hampton

May 2016

The focus of this study was to measure teachers' perception of the validity, reliability, feedback given from the Mississippi Statewide Teacher Appraisal Rubric (M-STAR) evaluation system, and their overall perception of the M-STAR evaluation system. This was a quantitative study that investigated whether or not a statistically significant difference existed between; teachers' years of teaching experience, teachers' teaching in a tested or non-tested subject, total number of M-STAR observation, and teachers' amount of M-STAR training or professional development time; and teachers' perception of the M-STAR's validity, perception of M-STAR's reliability, perception of the feedback given by M-STAR, and the teachers' overall perception of the M-STAR teacher evaluation system. A 40 statement survey instrument was developed to obtain quantitative data related to teacher perceptions of the M-STAR teacher evaluation system. A five-point scale ranging from Strongly Agree to Strongly Disagree with an option of no opinion or not enough information to respond was used. To test the hypotheses of this study, a multivariate analysis of variance (MANOVA) was utilized to determine statistical significance. A better understanding of the perceptions held by teachers concerning the M-STAR teacher evaluation system in hopes that this better understanding will inform the use of the current system and the design of future systems.

## ACKNOWLEDGMENTS

The researcher would like to thank the committee chair, Dr. David Lee, for his guidance and encouragement throughout the whole graduate education process starting with the Masters interview in 2004 to the culmination of the dissertation process. The consistency that Dr. Lee provided over the years was invaluable to me.

I would also want to say a special thanks to Dr. J.T. Johnson for helping me through my pilot study and also helping make sense of the results of my study.

I would like to thank Dr. Richard Mohn, Dr. Knya Shelley, Dr. Stanley Benigno, and Dr. James Fox for agreeing to serve on my committee so late in the process.

My final thank you is to the many professionals, too many to list, I have worked with throughout my career in education. I have been truly blessed to be in the presences of great educators who love teaching children.

## DEDICATION

As with anything I have accomplished or will accomplish in my life, this dissertation was made possible through the support of my wife Stacey, my children Lindsey, Ryan, and Julianna, my mother Billie Sue Fick, my father Larry Hampton, my best friend Terry Ingram, and my siblings Sandy, Shawn, Scott, and Suzanne.

To Stacey, none of this would have been possible without your love, patience, sacrifice, and unwavering support. I thank God for having you by my side through this journey of life.

To Lindsey, Ryan, and Julianna, thank you for your understanding and patience with me while I was going through classes, doing research, and writing.

To my mother, your example of what it means to be a lifelong learner had meant more to me than you will ever know. Having you go through the PhD process with me was very special to me. Thank you for your support.

To my late father, thank you for teaching me that all learning does not happen while you are sitting in a desk or reading a book.

To Terry, I am eternally grateful for your friendship. Your advice, consoling, counseling, and your help with everything has meant the world to me. I cannot express in words how much your friendship meant to me.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGMENTS .....	iii
DEDICATION .....	iv
LIST OF TABLES .....	viii
LIST OF ABBREVIATIONS.....	ix
CHAPTER I -INTRODUCTION .....	1
Statement of the Problem.....	3
Research Questions .....	4
Research Hypotheses .....	5
Definitions of Terms .....	5
Delimitations.....	8
Assumptions.....	8
Justification.....	9
CHAPTER II - REVIEW OF LITERATURE .....	10
Conceptual Foundations.....	16
Theoretical Framework.....	19
Purpose of Teacher Evaluation .....	20
Problems with Teacher Evaluation .....	25
Validity of Teacher Evaluation .....	27

Reliability of Teacher Evaluation .....	29
Problems with Teacher Evaluation Feedback.....	31
Effective Teacher Evaluation Feedback .....	32
Specific and Goal Oriented Feedback.....	34
Attainable and Actionable Feedback .....	35
Timely Feedback.....	36
Consistent and Credible Feedback.....	37
Mississippi Statewide Teacher Appraisal Rubric (M-STAR) .....	37
Teacher Perception of Teacher Evaluation .....	46
CHAPTER III - METHODOLOGY .....	48
Research Design.....	48
Participants.....	49
Instrumentation .....	49
Procedures.....	52
Data Analysis .....	53
CHAPTER IV - RESULTS .....	54
Descriptive Information of the Sample.....	54
Statistical Results .....	63
CHAPTER V - DISCUSSION.....	73
Summary of Procedures .....	73



Conclusions.....	74
Discussion.....	78
Limitations .....	80
Recommendations for Policy.....	81
Recommendations for Future Research.....	83
APPENDIX A - LETTER REQUESTING PERMISSION FROM SUPERINTENDENTS .....	85
APPENDIX B - CONSENT LETTER .....	86
APPENDIX C - SURVEY INSTRUMENT.....	87
APPENDIX D – IRB Approval Letter.....	95
REFERENCES .....	96

## LIST OF TABLES

Table 1	Frequencies and percentages of demographic variables.....	55
Table 2	Descriptive Statistics for Teachers' Perception of M-STAR Validity .....	58
Table 3	Descriptive Statistics for teachers' perception of M-STAR reliability.....	59
Table 4	Descriptive Statistics for teachers' perception of M-STAR feedback.....	60
Table 5	Descriptive Statistics for teachers' overall perception of M-STAR .....	62
Table 6	Teaching experience descriptive statistics .....	64
Table 7	Tested subject area descriptive statistics .....	66
Table 8	Total number of observations descriptive statistics.....	68
Table 9	Professional development descriptive statistics.....	71

## LIST OF ABBREVIATIONS

<i>ARRA</i>	American Recovery and Reinvestment Act
<i>ESEA</i>	Elementary and Secondary Education Act
<i>FFT</i>	Framework for Teaching
<i>M-STAR</i>	Mississippi Statewide Teacher Assessment Rubric
<i>MANOVA</i>	Multivariate Analysis of Variance
<i>MDE</i>	Mississippi Department of Education
<i>NCLB</i>	No Child Left Behind
<i>PD</i>	Professional Development
<i>RTT</i>	Race to the Top
<i>SIG</i>	School Improvement Grant
<i>STEC</i>	Statewide Teacher Evaluation Council
<i>TIF</i>	Teacher Incentive Fund

## CHAPTER I -INTRODUCTION

In recent years there have been several federal initiatives that have forced an overhauling of teacher evaluations in public schools across the United States (Popham, 2013). The \$4.35 billion Race to the Top (RTT) federal grant program enticed states to meet reform guidelines that included the implementation of strenuous teacher evaluation systems (Popham, 2013). These teacher evaluation systems included performance based standards for both administrators and teachers, and measures of student learning (Marzano & Toth, 2013). Similarly, the American Recovery and Reinvestment Act (ARRA) of 2009 sought to improve the effectiveness of teachers by establishing and implementing fair and reliable teacher evaluations that informed schools on the effectiveness of or the need for professional development (U. S. Department of Education, 2010). The Elementary and Secondary Education Act (ESEA) Flexibility Program offered a waiver from the increased requirements and sanctions of No Child Left Behind (NCLB) to states that met essentially the same teacher evaluation requirements as RTT (Popham, 2103). Popham (2013) pointed out that both the RTT and the ESEA Flexibility Program initiatives required: a) teacher evaluations be used for continuous classroom instructional improvement; b) at least three different performance levels be included; c) teachers' levels be determined by multiple evidenced-based sources with student growth carrying the most weight; d) regular evaluations of administrators and teachers be conducted; e) feedback be useful, timely and clear; and f) be used to inform evaluators on teacher retention.

Researchers and practitioners agree that the process of evaluating teachers is most often unproductive in spite of convincing evidence showing its importance. The teacher

evaluation process has the potential to increase every teacher's effectiveness, but the use of it seldom does (Duke & Stiggins, 1990). Teacher evaluations have been ineffective mainly because they have primarily been constructed using checklists based on the assumptions of what effective teaching should look like (Peterson, 2000). Recent reports and initiatives have spotlighted reasons that teacher evaluations have failed or have been ineffective (Toch & Rothman, 2008; Weisberg, Sexton, Mulhern, & Keeling, 2009). In the New Teacher Project study, Weisberg et al. (2009) extensively documented how teacher evaluation systems "fail to distinguish great teaching from good, good from fair, and fair from poor" (p. 3). Effective teachers are the greatest factor in the improvement of student achievement and yet teaching effectiveness has not been "measured, recorded, or used to inform decision-making in any meaningful way" (Weisberg et al. 2009, p. 3). Marzano (2012) also pointed to the failure of teacher evaluation systems to accurately measure teacher quality due to their inability to discriminate between ineffective and effective teachers and their inadequacies in developing high-level teacher skills. Orey (2007) concluded that when administrators fail to take action with less effective teachers, they themselves are viewed as ineffective because they appear to be tolerating or accepting unsatisfactory work. In situations like this, effective teachers are more likely to leave to find a school that will appreciate and recognize their contributions (Colvin, 2001).

In this current environment of school accountability there is an apparent need for the evaluation of teachers, but what is not as apparent is the method by which teachers should be evaluated. Current systems include summative administrator observations using checklists and rating scales; formative methods of collaboration between teacher

and administrator; professional development oriented portfolios; teacher self-assessment; and the use of multiple data sources (Danielson, 2001; Danielson & McGreal, 2000; Peterson, 2000; Stronge, 1997).

The focus of teacher evaluation is beginning to shift from being viewed as a tool used to terminate teachers to a tool used to improve teachers' pedagogical skills. Even still the person who should have the most influence in whether the process of evaluating a teacher is valid and reliable, the teacher, has historically not been consulted. This study determined if teachers perceived the evaluation system being used to determine their effectiveness is valid, reliable, and gives sufficient feedback to improve their instructional practices.

#### Statement of the Problem

A teacher evaluation that identifies both effective and ineffective teaching practices is essential to improving instruction. Equally important is the perception of the teacher that is being evaluated. If a teacher perceives the system being used to evaluate them as invalid, unreliable, and lacks adequate feedback, the teacher may not view their evaluations as a means to improve their instructional practices. Perceived validity of a teacher evaluation instrument rests in teachers' confidence that the scores accurately reflect the quality of instruction being measured. Reliability teacher evaluation systems depend on the proficiency and consistency of its raters. Danielson (2010) emphasized the importance of raters learning how to calibrate their ratings in conjunction with the ratings of others. Finally research indicates that effective feedback must be specific and goal oriented, attainable, actionable, timely, consistent, and credible (Brookhart, 2012; Hattie, 2009; Wiggins, 2012; Wiliam, 2012). Weisburg et al. (2009) stated:

By failing to produce meaningful information about instructional effectiveness, teacher evaluation systems severely limit the ability of schools and school systems to consider performance when answering critical questions or making strategic decisions about their teacher workforce. (p. 24)

In the absence of quality feedback, teachers are not able to reflect on their instructional practices which can result in a decrease in their desire to improve (Aseltine, Faryniarz, & Rigazio-DiGilio, 2006; Frase, 1992). Effective feedback should be based on observable evidence, affirm positive characteristics of teaching, and mold teachers into self-directed learners by promoting reflection (Danielson & McGreal, 2000).

#### Research Questions

This study sought to address the following research questions:

1. Do teachers perceive the M-STAR teacher evaluation system as a valid measure of their teaching effectiveness?
2. Do teachers perceive the Mississippi Statewide Teacher Appraisal Rubric (M-STAR) teacher evaluation system as reliable?
3. Do teachers perceive the feedback given from evaluators using M-STAR effective enough to influence teaching practices?
4. What is the overall perception that teachers have of the M-STAR teacher evaluation system?

## Research Hypotheses

H1 There is a statistically significant difference in teachers' perception of the validity, reliability, quality of feedback received, and the overall M-STAR teacher evaluation system based on years of teaching experience.

H2 There is a statistically significant difference in teachers' perception of the validity, reliability, quality of feedback received, and the overall M-STAR teacher evaluation system between teachers of subjects that are included in the Mississippi accountability system and teachers of subjects that are not.

H3 There is a statistical significant difference in the total number of M-STAR evaluations a teacher receives and their perception of the validity, reliability, quality of feedback received, and the overall M-STAR teacher evaluation system.

H4 There is a statistically significant difference in teachers' perception of the validity, reliability, quality of feedback received, and the overall M-STAR teacher evaluation system between teachers receiving one to two hours, two to four hours, four to six hours, six or more hours, or no training or professional development on the M-STAR teacher evaluation system.

## Definitions of Terms

This study consisted of the following terms:

*Domain:* “is a broad category of skills, knowledge, dispositions, and related elements in an educator performance framework” (MDE, 2014, p. 31). For the purpose of this study, domains are over arching descriptions defined by standards and indicators.

*Evidence:* “a factual reporting of events that are not biased or clouded with personal opinion” (MDE, 2014, p. 31). For the purpose of this study, “evidence may



include teacher and student behavior as well as teaching artifacts” (MDE, 2014, p. 31).

*Feedback:* “insight from the evaluator on a teacher’s performance that is grounded in the five domains and the twenty standards of M-STAR” (MDE, 2014, p. 31). For the purpose of this study, “the components of feedback are areas of strength, areas for growth, and the next steps identified for a teacher to make improvements” (MDE, 2014, p. 31).

*Formal classroom observation:* “is a period of time during which a trained evaluator visits a classroom and uses a rubric to measure observable classroom processes, including specific teaching practices, aspects of instruction, and interactions between teachers and students” (MDE, 2014, p. 31).

*Formative observation:* is ongoing observations throughout the school year, which includes the formal (fall) observation and all other informal observations (MDE, 2014).

*Inter-rater reliability:* For the purpose of this study, the degree to which measurements of the same observable event by different observers will yield the same results or the consistency of results (Graham, Milanowski, & Miller, 2012)

*Measures:* For the purpose of this study, “types of instruments or tools used to assess the performance and outcomes of educator practice (e.g., student growth scores, observations, student surveys, analysis of classroom artifacts, and student learning objectives)” (MDE, 2014, p. 31).

*Mississippi Statewide Teacher Appraisal Rubric (M-STAR):* the teacher evaluation system used by the State of Mississippi to make determination about a teacher’s instructional practices (MDE, 2012).

*Post-observation conference:* is a meeting between a teacher and an observer that takes place after a formal classroom observation, allowing the observer to ask clarifying questions about what was observed during the lesson and any outcomes after the lesson (e.g., assessment results and samples of student work). The teacher should also receive feedback and next steps during this conference (MDE, 2014, p. 32).

*Pre-observation conference:* a meeting between a teacher and an evaluator that takes place prior to a formal classroom observation, in order to provide the observer with background information about the lesson, the students, and any other details that may help the observer understand the context of the classroom. Additionally, it is an opportunity for the teacher to ask clarifying questions about the formal observation process (MDE, 2014, p. 32).

*Summative assessment:* “is an often high-stakes assessment administered primarily at the end of a specific period of time (e.g., a school year) to provide a judgment on an educator’s performance” (MDE, 2014, p. 32).

*Summative observation:* “a second (optional) formal observation, which in combination with all other formative observations provides data to determine a teacher’s summative rating” (MDE, 2014, p. 32).

*Teacher Evaluation:* an expert estimation of the quality, quantity, and other characteristics of teaching practices based upon common standards and indicators of teacher quality (Glickman, Gordon, & Ross-Gordon, 2001).

*Teacher evaluation system:* a system that: (1) is used for continual improvement of instruction; (2) meaningfully differentiates performance using at least three performance levels; (3) uses multiple valid measures in determining performance levels,

including data on student growth as a significant factor, and other measures of professional practice; (4) evaluates teachers on a regular basis; (5) provides clear, timely, and useful feedback, including feedback that identifies needs and guides professional development; and (6) is used to inform personnel decisions (Danielson, 2010).

#### Delimitations

The sample for this study was limited to certified kindergarten through twelfth grade public school teachers within school districts located in South Mississippi. The survey instrument used a five-point Likert-scale. Therefore, this study was completely quantitative, and respondents did not have the opportunity to elaborate on their responses. The results were solely statistics based, limiting the reasons for the respondents' answers; therefore this study only focuses on the teachers' perceptions and does not consider any additional subjective data.

#### Assumptions

The following assumptions were present in this study:

1. All teachers in this study have been exposed to or have participated in the M-STAR teacher evaluation and therefore have some knowledge of the system.
2. All teachers participating in this study have been trained on the implementation of the teacher evaluation system.
3. Participating teachers have the information and resources to properly implement and evaluate the effectiveness of the M-STAR evaluation system.
4. Participants' responses to the survey were honest, had integrity, and their willingness to respond had an impact on the responses.

## Justification

Teacher evaluations have been performed in schools across the United States for many years, but these evaluations historically have not been viewed as a tool to develop a teacher's professional practice (Marzano, Frontier, & Livingston, 2011). There is a clear need to ensure effective teaching practices in order to increase student performance. Researchers as far back as Levin (1979) posited that schools needed to re-examine the purpose and practice of teacher evaluation in order to: incorporate multiple sources of data; rely less on principal ratings; and involve teachers in developing evaluation policies to increase their commitment to the use of the evaluation results. Research by Machell (1995) and Marshall (2005) identified four attributes of teacher evaluation systems proven to facilitate teacher growth. These attributes were: clear, relevant, and meaningful performance feedback using multiple data sources; goal setting by teachers; mutual trust between teacher and evaluator; and professional development based on the teacher evaluation. Marx (2007) stressed that an effective educational leader plays a positive role in the evaluation process by collaborating with teachers, facilitating reflection on instructional practices, and providing meaningful feedback. A school environment that ensures teacher evaluation systems are conducive and supportive of ongoing professional growth is one that supports teacher evaluation, focuses on instruction and student learning, encourages robust collaboration among teachers, and the use of reflective practices by teachers (Marx, 2007). This study is designed to gauge the perception of teachers being evaluated using the M-STAR teacher evaluation system.

## CHAPTER II - REVIEW OF LITERATURE

There have been a number of researchers who have concluded that the most influential impact on student learning is an effective teacher (Goldhaber, 2002; Haycock, 1998; Rivkin, Hanushek, & Kain, 2005; Robinson, Lloyd, & Rowe, 2008; Sanders, Wright, & Horn, 1997). One specific study by Hanushek (1992) highlighted the importance of having a high quality teacher in the classroom. This study concluded that disparities in test performance of a student having an effective teacher as opposed to students having an ineffective teacher could be as much as one grade-level or more (Hanushek, 1992). Sanders and Horn (1998) followed up by showing that students with ineffective teachers perform as much as fifty percentile points lower on norm-referenced mathematics assessments. Teacher evaluations have been performed in school districts across the United States for a number of years; however, historically they have not been viewed as a part of developing a teacher professionally (Marzano et al., 2011). There is a clear and present need to ensure effective teaching practices in order to increase student performance.

This chapter presents a review of the literature related to the history of the teacher evaluation process; the conceptual and theoretical underpinnings of the evaluation of teachers; the purpose and problems with teacher evaluations; validity and reliability issues related to teacher evaluation systems; problems with feedback and what constitutes effect feedback; the history and make-up of the Mississippi Statewide Teacher Appraisal Rubric (M-STAR); and the perception of teachers regarding their evaluations. Each of these themes was intricately linked to build a foundation for the study of a teacher evaluation system that will improve teacher efficacy.

## History of Teacher Evaluation

Early in the 1900s, teacher evaluations were most often based on a teacher's physical attributes and moral character (Shinkfield & Stufflebeam, 1995). This meant as long as teachers maintained their appearance, had good character and was friendly; the teacher was viewed as an effective teacher (Shinkfield & Stufflebeam, 1995). Evaluations were usually performed by a local clergymen or a government official without any set procedures or protocols (Marzano et al., 2011). Lacking the training to effectively perform the evaluations, the evaluator's primary purpose for the evaluation was to terminate teachers they felt were incompetent, or teachers who would not conform to the norms of the community (Peterson, 1982).

With the onset of the Industrial Revolution, the population of the United States began to grow and see a shift from rural communities to a more urban population (Peterson, 1982). Because of this growth, urban schools began to grow to the point that larger class sizes forced the separation of grade levels resulting in schools with more than one teacher (Peterson, 1982). With this increase in the number of teachers, there began to be a need for administrators to supervise teachers (Peterson, 1982). At this time administrators acted as inspectors and had little to do with the teaching process itself (Shinkfield & Stufflebeam, 1995). The role of the administrator was one of primary decision maker where teachers were managed. This meant administrators were not expected to develop teachers professionally and did not make professional development a priority (Lambert et al., 2002).

Teacher behaviors and effective lessons didn't become the primary focus of teacher evaluations until the mid 1900s (Danielson & McGreal, 2000; Hunter, 2004;

Marzano et al., 2011). This change was brought on by the launch of Sputnik by the Soviet Union on October 4th, 1957 (Clemesten, 2000; Peterson, 1982). Since the Soviet Union was the first to launch a satellite into space, most Americans blamed the United States' educational system for the country's inability to be the leader in the space race (Peterson, 1982). Peterson (1982) contended that as a result of this perception, the public was made more aware of the problems of the United States' educational system, and the legislative branches of the United States government were pressured into remedial action. Public education in the United States began to be systemized like the factories of this time, where the use of designed instructional programs was used to present the curriculum to students (Kersten & Israel, 2005). Administrators primarily managed teachers by setting specific job related targets (Lambert et al., 2002), and tracked student growth through their progression of courses or the textbook they were using (Kersten & Israel, 2005). The response of the public educational system, through changing their focus and curriculum content showed how quickly the United States' educational system, could meet the objectives to help the country remain a superpower (Clemesten, 2000).

In the late 1960s, R.E. Stake realized that in order to fix the problems in the educational system, teacher observations needed to be formalized (Coutts, 1999). Stake (1967) viewed teacher observations as too casual and subjective. Administrators needed to pay attention to the difference between what was happening in the classroom and what was intended to happen (Stake, 1967). David Ryans' (1960) book titled *Characteristics of Teachers: Their Description, Comparison and Appraisal* identified characteristics of effective teachers as warm, understanding, friendly, responsible, businesslike, systematic, stimulating, and imaginative. This became problematic when effective teachers were

found to have characteristics that were different from the characteristics thought to be possessed by effective teachers (Peterson, 1982). Despite these problems, Ryans' findings revealed information on positive characteristics of teacher behaviors and measurable objectives on predicting teacher behavior (Clemesten, 2000; Peterson, 1982). Ryans' (1960) findings ultimately lead to many states implementing teacher rating scales based on the identified desirable teacher attributes (Shinkfield & Stufflebeam, 1995). School systems across the country used research like Ryans to create evaluation systems that rated teachers on characteristics and attributes of what effective teaching should look like (Shinkfield & Stufflebeam, 1995). Evaluation systems were created that used metrics such as behavior checklists, rating scales, time and event sampling, sign systems, and narrative records (Stronge, 1997).

In the 1970s the work of Madeline Hunter continued a summative approach that focused on teacher practices involving systems of direct instruction (Brandt, 1996). Hunter's *Seven Essential Elements of Effective Lessons* focused on teacher behavior and practices and provided the foundation for teacher evaluations in school systems during the 1970s and 1980s (Danielson & McGreal, 2000; Marzano et al., 2011). Hunter's (1994) elements were the learning objective, the anticipatory set, the lesson objective, input, check for understanding, guided practice, and independent practice.

In the later part of the 1970s, Elliot Eisner's *Educational Connoisseurship* incorporated three questions into his system of teacher evaluation: what did the evaluator see; what did the evaluator think about what they saw; and how would the evaluator express in words what they saw during the evaluation process (Coutts, 1999). Coutts (1999) also pointed out that Eisner was highly in favor of full disclosure to all the



participants in the evaluation process. Eisner (1975) defined “connoisseurship as the ability to make fine-grained discriminations among complex qualities. Criticism is the connoisseur’s disclosure of those perceptions so that others not possessing his level of connoisseurship can also enter into the work” (Eisner, 1975, p. 1). The connoisseur does this through description, interpretation, evaluation, and identifying dominant features (Eisner, 1975).

In the 1980s, efforts to reform schools resulted in an increased amount of attention on teacher evaluation and its role in improving teaching quality (Brandt, 1995; Darling-Hammond, 1990). Generally teacher evaluations relied on observations that were too few in number, did not differentiate between novice and more experienced teachers, focused on low level instructional strategies, and the lack of multidisciplinary expertise by the observers (Danielson & McGreal, 2000). It was during this period of time that states began to draft laws in the attempt to standardize teacher evaluation (Wuhs & Manatt, 1983).

In 1983 the National Commission on Excellence in Education published the report *A Nation at Risk* that resulted in another call to improve education in the United States and lead the educational system into the modern age of accountability (Danielson, 2001). Recommendations from this report included the need for highly competent teachers and teacher salaries that were professionally competitive and performance based (Clemetsen, 2000). The public discontent generated by this report forced school systems to realize that teacher evaluation was the key to the improvement of teacher competency (Shinkfield & Stufflebeam, 1995). As a result, states began to enact mechanisms that tied teacher evaluations to a teacher’s certification renewal, licensure, merit pay, and career

ladders (Ellett & Teddlie, 2003). *A Nation At Risk* brought about a shift in researchers' thinking, changing the focus from teacher behaviors only. For the first time, students and their work became a part of a teacher's evaluation (Ellet & Teddlie, 2003).

In 1996 a publication from the National Commission on Teaching and America's Future titled *What matters most: Teaching for America's future*, brought attention to legislators and educators regarding the need to apply research-based solutions to the teacher evaluation process (Darling-Hammond, 1996). Teacher evaluations were now based on achievement, not on teacher behavior (Danielson & McGreal, 2000; Marzano et al., 2011).

In order to raise the level of accountability in public K12 school systems across the United States, President George W. Bush proposed and legislators passed the *No Child Left Behind Act of 2001* (NCLB). This legislation contained mandates that regulated teacher evaluation systems by defining teacher quality, defining minimum standards for training an evaluator, and requiring data collection on teacher evaluations (Hazi & Rucinski, 2009). Parts of this legislation required teachers be *highly qualified* and required schools to provide parents with information on teachers' level of education, licensure, and area of certification upon request (Linn, 2003). The U. S. Department of Education (2004) defines a highly qualified teacher as someone who is fully certified and/or licensed by the state, holds at least a bachelor degree from a four-year institution, and demonstrates competence in each core academic subject area taught.

In recent years, federal initiatives have forced an overhauling of teacher evaluations in the United States (Popham, 2013). The \$4.35 billion Race to the Top (RTT) federal grant program enticed states to meet reform guidelines that included

strenuous teacher evaluation systems (Popham, 2013). These systems included performance-based standards for both administrators and teachers, and measures student learning (Marzano & Toth, 2013). Similarly, the American Recovery and Reinvestment Act (ARRA) of 2009 sought to improve the effectiveness of teachers by establishing and implementing fair and reliable teacher evaluations that were used to inform schools about professional development needs (U. S. Department of Education, 2010). The Elementary and Secondary Education Act (ESEA) Flexibility Program offered states that met essentially the same teacher evaluation requirements as RTT a waiver from the increased requirements and sanctions of NCLB (Popham, 2013). Popham (2013) pointed out that both initiatives required that teacher evaluations:

- be used for continuous classroom instructional improvement;
- must include at least three different performance levels be included;
- determine teachers' levels via multiple evidenced-based sources with student growth carrying the most weight;
- conducted regularly;
- provide useful, timely and clear feedback;
- be used to inform evaluators on teacher retention.

#### Conceptual Foundations

Continual advancement of research in the field of teacher effectiveness, combined with effective classroom practices that change according to the research, have resulted in an environment where teacher evaluation systems need to progress accordingly to reflect these changes (Danielson & McGreal, 2000). The teacher evaluation process uses observations to make judgments on the quality of instruction being provided by a teacher

and address any need for professional growth (Danielson & McGreal, 2000; Marshall, 2005). Teacher evaluation systems should establish a clear common vision with well defined and research proven practices that promote high-quality instruction and differentiated levels of performance (Danielson & McGreal, 2000).

The Mississippi Statewide Teacher Appraisal Rubric (M-STAR) follows a system based on the research of Charlotte Danielson (MDE, 2012). Danielson's (2007) *Enhancing Professional Practice: A Framework For Teaching*, designed a framework of twenty-two components clustered into four different domains. Teachers are evaluated on these components by observations with detailed rubrics that provide them with a rationale for the evaluator's actions through artifacts, collecting evidence through scripting, and conferences (Danielson, 2007). The domains of planning and preparation, classroom environment, instruction, and professional responsibilities in Danielson's Framework for Teaching (FFT) were based on research by Madeline Hunter (1994), Lee Shulman (1987), and the Interstate New Teacher Assessment and Support Consortium (INTASC) (1992). Danielson (2007) contended that Madeline Hunter was among the first researchers to make a convincing argument that there are particular instructional practices that were more effective than others. Danielson (2007) relied on Hunter's research on process-product and cognitive science when construction the FFT. Hunter derived a behavioral teaching system that focused on the delivery of instruction. In Hunter's system, the administrator was primarily responsible for teacher performance, and rewards were used to modify teacher behavior (Catano & Stronge, 2006). Evaluation of teacher performance was gauged by checklists that administrators used to determine areas for professional development (Kersten & Israel, 2005). However, the use of checklists for

evaluation failed to measure all the complexities associated with effective instruction (Danielson & McGreal, 2000). Also, Danielson (2007) contended that her FFT was based in Shulman's (1987) research on content knowledge, pedagogical knowledge, and pedagogical content knowledge, and the standards outlined in the FFT were derived from the Interstate New Teacher Assessment and Support Consortium (INTSAC, 1992).

Parts of Danielson's system of teacher evaluation have been implemented in school districts across the country since its publication (Pritchett, Sparks, & Taylor-Johnson, 2010). Donaldson (2009) posits that in many states, Danielson's system is one of the most common teacher evaluation system being used in school districts. This system of teacher evaluation contains three necessary elements for effective evaluations, which help to ensure that teacher evaluations are valid, reliable and provided feedback that improves instruction. First, the system requires a clear, shared definition of effective instruction through eleven evaluative criteria. Next, the system requires evaluators use techniques and procedures that ensure teachers are being measured accurately and consistently in regards to a shared definition. Finally, Danielson's system allows for trained evaluators to use their judgments to offer recommendations on appropriate professional development to each teacher (Danielson & McGreal, 2000). The teacher evaluation process is used for summative decision making, but should also be used formatively in order to improve teacher effectiveness and student learning (Danielson & McGreal, 2000). Evaluators have to be able to recognize distinguished instructional practices, and be able to give effective feedback to teachers. Danielson expressed that school districts "can design evaluation systems in which educators can not only achieve

the dual purposes of accountability and professional development, but can merge them" (Danielson & McGreal, 2000, p. 10).

### Theoretical Framework

Danielson's framework is theoretically underpinned by the constructivist research of Dewey (1910), Piaget (1952), and Vygotsky (1978). In the constructivist pedagogy, the part the teacher plays in student learning is very significant in the current age of accountability (Kaplan & Owings, 2001). The evaluation of a teacher involves judgments on the effectiveness of the teacher through a series of observations (Danielson & McGreal, 2000; Marshall, 2005). Observations, in general should not be confused with documenting what is obvious but should be an active process of discovering something (Dewey, 1910).

Education comes from establishing learning communities that collaborate, exchange ideas, and actively learn (Dewey, 1916). The process of evaluating teachers, as seen through Dewey's constructivist views, should allow teachers to be active participants in discovering knowledge on effective teaching practices, make meaning of that knowledge, and reflect on effective teaching practices all within a culture that supports these cognitive processes (Danielson & McGreal, 2000; Dewey, 1916). This allows for learning communities that share similar beliefs, ways of thinking, knowledge, and goals formed by the exchange of information, teaching and learning, accountability, creativity, and reflection (Dewey, 1916). This can be difficult to achieve because schools and their teachers often operate in isolation (Dewey, 1916). This isolation or lack of community inhibits productivity and the accomplishing of goals in schools (Dewey, 1916). Dewey (1916) maintained that a failure to communicate and share learning would

lead to barriers to growth, which in turn would result in the stagnation of these communities.

Works from educational theorist such as Jean Piaget (1952) and Lev Vygotsky (1978) continued the constructivist perspective. Even though their works were varied, they both believed that the construction of new knowledge happens by people interacting with one another and their environment (Driscoll, 2005). Piaget (1952) believed that learning is a result of a person's cognitive effort to construct their own personal knowledge. Vygotsky (1978) emphasized the role that culture and social context plays in the learning process. His Zone of Proximal Development theory described a form of scaffolding by the teacher in supporting the student when assistance is needed and allowing the capable student to accomplish the task alone.

#### Purpose of Teacher Evaluation

Duke and Stiggins (1990) put forward that it was important to define the purpose of a teacher evaluation system. Many researchers have identified the purpose of teacher evaluation. For example, Doyle (1983) stated the reasons for teacher evaluation is to diagnose and help improve teacher instruction, to support administration about individual faculty members, to help students choose courses and plan programs, and to provide standards for research on teaching. Manning (1988) stated the purpose of teacher evaluation is to make tenure decisions, determine pay increases, assure accountability, remove incompetent teachers, enhance administrative authority, and determine promotions. Natriello (1990) viewed teacher evaluation as having three main purposes: controlling or influencing the performance of a person within specific positions; controlling a person's movement into and out of a position; and validating the

organizational control system. Haefele (1993) viewed teacher evaluation systems as a means to assist schools in the termination of incompetent teachers, provide individualized constructive feedback, provide recognition and reinforcement, give direction on professional growth, provide evidence that will endure professional and judicial examination, and to unify the collective efforts of the teacher and administration in educating students. Finally, the Joint Committee on Standards for Educational Evaluation identified entry to training, certification or licensing, definition of a role, reviewing performance, informing staff development, merit awards, decisions on tenure, decisions on promotion, and decision on termination as the purposes of teacher evaluation (Wheeler & Scriven, 1997).

According to Stronge and Tucker (2003) the meaning of life is derived from experiencing personal growth and being committed to a cause that is larger than one's own self-interest. If either personal growth or commitment to a cause is emphasized to the point that a person excludes the other the result is a person who cares too little about the welfare of society or doesn't have the knowledge to contribute to it (Stronge & Tucker, 2003). This describes the dilemma between professional growth or accountability facing school districts concerning teacher evaluation (Stronge & Tucker, 2003). Stronge and Tucker argued that because teaching matters, teacher evaluation should matter and that any reform in education cannot succeed without capable, high quality teachers. Identifying capable, high-quality teachers cannot happen without a high quality teacher evaluation system (Stronge & Tucker, 2003).

Shinkfield and Stufflebeam (1995) believed that when trying to examine the true purpose of teacher evaluation there is a major problem in deciding whether the outcomes



conform to organizational standards, or to develop professional requirements of teachers based on their interactions with students. Darling-Hammond, Wise, and Pease (1983) stated that teacher evaluation systems designed for accountability purposes should be capable of producing objective, standardized, and externally defensible data on a teacher's performance. Conversely, evaluating systems that are designed for teacher growth should produce information that informs a teacher on areas of needed improvement along with guidance to improve (Darling-Hammond et al., 1983).

Accountability and professional growth have been the two most cited reasons for teacher evaluation (Danielson & McGreal, 2000; Peterson, 2000). Often associated with a summative approach to teacher evaluation, accountability reflects the need to determine the competence of teachers so that evaluators can be assured that instructional practices are safe and effective (McGaghie, 1991). As part of the formative evaluation process, teachers anticipate honest and constructive feedback that is aligned to professional growth (Range, Young, & Hvidston, 2013). These two purposes of teacher evaluation are generally believed to be mutually exclusive, but in order for the evaluations to be beneficial, teachers must create a logical link between the two (Danielson & McGreal, 2000; Peterson, 2000; Stronge & Tucker, 2003). Glickman, Gordon & Ross-Gordon (2001) contend that when attempting to combine summative and formative purposes of teacher evaluations the emphasis is primarily placed on the summative purpose. Combining the two purposes should not be the goal, but the goal should be to clearly specify the purpose of each and allow both of them to operate together so they accomplish the function they were designed to accomplish (Glickman et al., 2001). Recognition that the two purposes are not competing is necessary to the improvement of

the delivery of educational services (Stronge & Tucker, 2003). Tucker and Stronge (2005) suggest that the evaluation of teachers should not only document the level of performance in order to hold teachers accountable for their instruction, but also help them improve their instruction. In order to serve the educational needs of the teacher, school and community at large, a teacher evaluation system that includes accountability and professional growth components are necessary (Stronge, 2006).

Effective evaluation systems have specific elements that have the potential to help schools continually improve and increase the quality of its teachers' instruction (Machell, 1995). Machell (1995) and Marshall (2005) identified characteristics of teacher evaluation systems that have proven to make teacher growth possible. These attributes were found to be clear, relevant, and meaningful performance feedback through multiple data sources; goal setting by teachers; mutual trust between teacher and evaluator; and professional development based on the teacher evaluation (Machell, 1995; Marshall, 2005). Researchers as far back as Levin (1979) believed that schools needed to reexamine teacher evaluation purposes and practices to incorporate multiple sources of data, rely less on principal ratings, and involve teachers in developing evaluation policies to increase teachers' commitment to the use of the evaluation results.

It is important to have credibility in a teacher evaluation system as it helps to consistently define good instruction. Doing so heightens the value of conversations of teachers that develop from classroom observations (Danielson, 2010). Only recently have evaluators attempted to marry the summative quality assurance with the formative professional growth in order to enhance evaluator skills by using cognitive coaching along with clinical supervision (Danielson, 2010). Danielson (2010) continued by stating

when requiring improvement in instructional practices and professional growth are embedded in the design of the evaluation system, teaching as a profession is better off.

Marx (2007) expressed that an effective educational leader plays a positive role in the evaluation process by collaborating with teachers, facilitating reflection on instructional practices by the teacher, and providing meaningful feedback. Marx (2007) also felt that it is important to gauge the school's culture and climate so as to ensure that the teacher evaluation system is conducive and supportive of ongoing professional growth. A school culture that includes a supportive teacher evaluation environment focuses on instruction and learning for all students, robust collaboration among teachers, and the use of reflective practices by teachers (Marx, 2007). This view of teacher evaluation ties the improvement of the teacher to the improvement of the school.

In order for teacher evaluation to have a positive impact on student learning, researchers have concluded that the evaluation process must meet three different criteria. The teacher evaluation process must be capable of removing poor performing teachers that fail to produce favorable student learning outcomes (Heneman, Milanowski, & Kimball, 2007; Koppich & Showalter, 2005; Odden & Wallace, 2008). The teacher evaluation process should also produce meaningful feedback that teachers can use to improve their instructional practices therefore improving student learning (Heneman et al., 2007; Odden 2004; Sanders et al., 1998). The teacher evaluation process should foster a results-oriented school culture that supports a wider set of policies that ensure the quality of teaching and learning within a school (Ellett & Teddlie, 2003; Odden, 2004).

## Problems with Teacher Evaluation

Recent reports and initiatives have spotlighted reasons that teacher evaluations have failed or have been ineffective (Toch & Rothman, 2008; Weisberg et al., 2009). Marzano (2012) pointed to teacher evaluation systems as failing to accurately measure the quality of teachers due to their inability to discriminate between ineffective and effective teachers, and their inadequacies in developing teacher skills to a high level. Administrators need to be reminded that the evaluation process should be used to help teachers improve their skills. Danielson (2010) stated that the reliability of teacher evaluations is compromised by the lack of consistency among raters. In order for a teacher evaluation system to be viewed as credible, higher proficiency levels from evaluators must be able to accurately judge teachers using a reliable, valid evaluation tool frequently, and provide feedback that is meaningful and produces productive conversations that improves their instruction (Danielson, 2010). Creating a valid and reliable teacher evaluation system starts with clearly defining teacher effectiveness because it will have an impact on how the effectiveness will be measured (Burling, 2012).

Multiple studies have shown that teacher evaluations have not emphasized improving instructional practices and have failed to provide teachers with adequate feedback (Frase & Streshley, 1994; Stiggins & Bridgeford, 1985). Weisberg et al. (2009) surveyed over fifteen thousand teachers in twelve school districts and found almost 75% of the teachers surveyed had not received any specific feedback on how to improve their instructional practices. The same study found that school districts seldom enacted formal dismissal proceedings on teachers (Weisberg et al., 2009). In fact during the five year

period of this study, half of the school districts participating did not have a single non-probationary teacher terminated for performing poorly (Weisberg et al., 2009).

Researchers have pointed to the lack of effectiveness of evaluating administrators for a reason why teacher evaluation has failed to improve student achievement. The validity of a teacher's evaluation may be affected by the skill level of the evaluator (Tucker & Stronge, 2005). For example, Donaldson (2009) communicates that because evaluators observe teachers in subjects in which they are not familiar, they may have difficulty assessing the teacher's performance accurately. In addition to the lack of subject familiarity, evaluating administrators are often unable to give feedback that may help improve a teacher's instruction because they are not in classrooms (Pritchett et al., 2010). Given the limited time administrators have to observe teachers, combined with very real limitation in subject areas, it is no wonder they have a false sense of the actual quality of the instruction they are evaluating (Pritchett et al., 2010). In general, teachers lack confidence in the ability of evaluations to improve their instruction because their evaluations are often brief and rushed, given the plethora of other administrative duties that the evaluator has to perform (Garth-Young, 2007). Also, some teachers perceive administrators as not using the evaluation process fairly or in such a manner as to terminate teachers the administrator does not like (Garth-Young, 2007). This perception leads to a lack of trust between the administrator and teachers, ultimately reducing the effectiveness of the evaluation process (Garth-Young, 2007).

Administrators must juggle the limited time they are able to spend on teacher evaluations with poorly designed evaluation systems, systems that do not provide meaningful feedback (Kersten & Isreal, 2005). Schools must foster an environment of

professional learning where teachers are responsible for continuous professional growth, but connecting teacher evaluation with professional development does not occur without work (Danielson & McGreal, 2000). School systems should seek opportunities to incorporate professional development with teacher evaluation procedures to promote, monitor, and determine teacher growth (Danielson & McGreal, 2000). Finally teacher evaluation systems have been criticized for promoting a strong focus on the teacher actions and behaviors, and not looking at student learning. The fundamental flaw in most teacher evaluations is the assumption that good teaching practices automatically translates into student learning and achievement (Tucker & Stronge, 2005). If the goal of teacher evaluation is to determine if students are learning, measuring that learning directly and not extrapolating it from the limited scope of the observations being performed is far more effective (Tucker & Stronge, 2005).

#### Validity of Teacher Evaluation

The Standards for Educational and Psychological Testing defines the term validity as referring “to the degree to which evidence and theory support the interpretations of the test” (American Educational Research Association, 1999, p. 9). Evidence of validity is used to determine if an assessment measures what it intended to measure in the way it was intended to be used (Crocker & Algina, 1986; Cronbach, 1971; Messick, 1981). Validity should not be confused with being a property of an assessment, but should be thought of as a property of the results (Messick, 1995).

Evidence of validity can be obtained through accumulating information that surrounds the assessment (Crocker & Algina, 1986; Kane, 2006). This can be done by inspecting the content of the assessment, the internal structure of the measure, and

relationship of the scores to other variables (American Educational Research Association, 1999). Collecting evidence such as this can determine if the assessment is measuring what it is intended to measure.

The validity of an instrument can be determined in several different ways. For example, when researchers correlate scores from a particular assessment to the performance area in which they are assessing, they are looking at criterion validity (Herlihy et al., 2013). In dealing with teacher quality, criterion validity is problematic due to the fact that outcomes of quality teaching (students' success in college or a career) are obtained many years after a teacher's evaluation cycle (Herlihy et al., 2013). On the other hand, construct validity measures specific constructs that do or do not theoretically correlate to quality teaching and then determines if the theoretical predictions are accurate (Herlihy et al., 2013). Face validity is where experts in a particular field, in this case teaching, agree that an instrument represents the domain of quality teaching (Herlihy et al., 2013). The final way to investigate validity is consequential validity. Intended and unintended consequences as a result of decisions or the actions taken based on high stakes tests should be examined for consequential validity. Consequences that artificially lead to the inflation of scores should raise concerns with the validity of the assessment (Koretz , 2008).

In order to validate teacher evaluations, the instrument used to collect data should be developed using a clear definition of what good instruction practices should look like (Danielson, 2008). Defining good instructional practices using an evaluation instrument may produce weak results by itself; however assessing whether the constructs are logical or not can help determine the validity of the instrument (Danielson, 2008).

Kimball (2002) points to three main elements in determining the validity of teacher evaluation instruments: content, construct and consequential validity. The traditional standard of validity is content validity, which is established by involving field related experts in the development of the evaluation and as well as confirming that the assessed standards appropriately measure the teachers' performance. Construct validity refers to the extent to which conclusions made from the application of the evaluation instrument accurately reflect what is being measured. Finally, consequential validity refers to the results of the decisions made from the evaluation (Kimball, 2002). Consequential validity focuses on aligning the evaluation process with desired outcomes - does the final decision represent the intended results. From example, does the evaluation improve teachers while maintaining performance accountability (Kimball, 2002)?

#### Reliability of Teacher Evaluation

At the center of any form of measurement is the reliability of the score. Most often in education the method of measuring teacher performance is the classroom observation. Graham, Milanowski, and Miller (2012) defines inter-rater reliability "as the measurement of the consistency between evaluators in the ordering or relative standing of performance ratings, regardless of the absolute value of each evaluator's rating" (p. 5). Danielson (2007) referred to reliability of a teacher evaluation system as being primarily related to training of evaluators. As early as Frick and Semmel (1978), researchers have stated that adequately training observers is critical for most criterion-based measures. Danielson (2011) explained that evaluators must be provided with training so that they can learn how to calibrate their judgments along with the judgments



of their colleagues. Training evaluators on the foundations and definitions of the standards used in the evaluation system is essential (Frick & Semmel, 1978). Inter-rater reliability should be assessed by using video recordings of classroom instruction rated by trainers and trainees during the evaluator's initial training and also as a part of ongoing periodic training (Frick & Semmel, 1978). In fact, Cangelosi (1991) contended that if evaluators are not trained adequately on properly designed evaluation instruments, teachers will continue to receive evaluations that misrepresent their abilities and produce unreliable scores.

The problem with inter-rater reliability is that when evaluators have similar ratings for two or more sets of evaluations the scores could be reliable but have little to no agreement (Tinsley & Weiss, 2000). Inter-rater agreement is defined by Graham et al. (2012) as "the degree to which two or more evaluators using the same rating scale give the same rating to identical observable situation" (p.5). Inter-rater agreement differs from inter-rater reliability in that it measures how consistent evaluation scores are and not how similar they are (Graham et al., 2012). Thus, two observers can assign scores that are similar and over time the scores would be considered reliable, but there could be little to no agreement between the two evaluators (Tinsley & Weiss, 2000). According to Graham et al. (2012):

Another way to think about the distinction is that inter-rater agreement is based on a "criterion- referenced" interpretation of the rating scale: there is some level or standard of performance that counts as good or poor. Inter-rater reliability, on the other hand, is based on a norm-referenced view: the order of the ratings with

respect to the mean or median defines good or poor rather than the rating itself (p. 6).

There are reasons why it is important to make a distinction between inter-rater reliability and inter-rater agreement. First, when decisions are being made about pay, promotions, or contract renewals that are based on evaluation scores, inter-rater agreement is more essential since decisions are based on scores with specific boundaries (Graham et al., 2012). Next, having inter-rater agreement provides feedback to teachers about their performance from sources that are considered to be more credible (Graham et al., 2012). Finally, inter-rater agreement produces a more accurate picture of what teachers' strengths and weaknesses are and this accuracy can better inform decisions about professional development (Graham et al., 2012).

#### Problems with Teacher Evaluation Feedback

Danielson (2010) explained that traditional teacher evaluation systems produce non-specific evaluative comments and provide little guidance to focus on improvement. Researchers also found when teachers receive feedback from their evaluators infrequently, it is not enough to impact their performance (Mielke & Frontier, 2012). Teachers have expressed frustration and have even felt cheated after years of evaluations that show no area of improvement even when self-assessments show areas they need to improve upon (Mielke & Frontier, 2012). Wiener and Lundy (2013) are pessimistic that current changes in teacher evaluation will solve the basic problem. Most principals don't provide teachers with detailed feedback on their performance, and teachers have become accustomed to a perfunctory process that rarely includes constructive criticism (Wiener & Lundy, 2013). Commonly, evaluator feedback follows what is referred to as the praise

sandwich: begin with a compliment, point out an area of concern, and then have a positive ending (Tugend, 2013). The power in feedback lies in its capability to play many roles and work at various levels of learning (Butler & Winne, 1995). Tugend (2013) points out that positive feedback is not always good and negative feedback is not always bad. The purpose of feedback is to help a person improve and not to make them feel better (Tugend, 2013).

Administrators are performing formal teacher evaluations that do not promote the exchange of important feedback between the two parties involved (Schmoker, 2006). These evaluations usually produce a rating for each observed indicator but often lack any meaningful feedback (Schmoker, 2006). Even administrators who are trying to change from being a managerial style leader to an instructional leader have yet to make the connection with the purpose of giving continual meaningful feedback to their teachers (Elmore & Fuhrman, 2001).

#### Effective Teacher Evaluation Feedback

Effective feedback should be based on observable evidence, affirm positive characteristics of teaching, and mold teachers into self-directed learners by promoting reflection (Danielson & McGreal, 2000). Typically, there has been a poor quality of feedback given to teachers following their evaluations (Frase & Streshley, 1994; Stiggins & Bridgeford, 1985, Weisburg et al. 2009). In *The Widget Effect*, Weisberg, Sexton, Mulhern, and Keeling (2009) surveyed 15,176 teachers in 12 districts and found that 75% of teachers had not received specific feedback on how they could improve their instructional practices. In fact, researchers have found that most evaluations have placed little emphasis on instructional improvement and the quality of feedback given to the

teacher has been poor (Frase & Streshley, 1994; Stiggins & Bridgeford, 1985). Most teachers want to improve their instruction, but they are often unsure as to how to do it (Johnson, 1997). Administrators can provide valuable insights and expert guidance to teachers regarding how to improve their instruction through the use of quality evaluations and specific feedback (Johnson, 1997).

In order to systematically improve the expertise of teachers, schools and districts must provide teachers with feedback (Marzano et al., 2011). Hattie and Timperley (2007) stated that “feedback is conceptualized as information provided by an agent (e.g., teacher, peer, book, parent, experience) regarding aspects of one’s performance or understanding” (p. 81). The intent of feedback should be to narrow the gap between where a person is and where that person should be (Hattie, 2012). Without feedback, efficient learning cannot take place and only minimum improvement can happen (Ericsson, Krampe, & Tesch-Römer, 1993). Researchers such as Hattie (2009), and Marzano, Pickering, and Pollock (2001) have determined that providing more feedback and giving less instruction can result in more learning. Furthering the research, Marzano and colleagues (2001), Mid-continent Research for Education and Learning (McREL) concluded that effect size of effective feedback is 0.76, which means that the average achievement difference between learners that receive effective feedback and those who do not is approximately 28 percentile points (Beesley, Apthorp, & McREL, 2010; Dean, Hubbell, Pitler, & Stone, 2012). Hattie (2009) performed over 800 meta-analyses in which he similarly found that feedback has an effect size of 0.73 on learning. Typically, feedback that is corrective in nature has the largest effect size (Marzano et al., 2001). Corrective feedback provides an explanation of what was being done correctly and what

was being done incorrectly (Marzano et al., 2001). Furthermore, Wiliam (2011) and Sutton, Douglas, and Hornsey (2012) posited that effective feedback can as much as double the rate of learning.

The importance of feedback in education and the way that it is given has been well established (Black & Wiliam, 1998). Research has shown that feedback functions to evaluate, to motivate, and to learn (Kahu, 2008). In the absence of quality feedback, teachers are not able to reflect on their instructional practices resulting in a decrease in their desire to improve (Aseltine et al., 2006; Frase, 1992). In order for feedback to be effective it must be specific and goal oriented, attainable, actionable, timely, and consistent and credible (Brookhart, 2012; Hattie 2009; Wiggins, 2012; Wiliam, 2012).

#### *Specific and Goal Oriented Feedback*

While certain leadership traits have an inconsequential effect on achievement, Witziers, Bosker, and Kruger (2003) found that goal setting has a more direct impact on student outcomes. In order for feedback to be effective, the person receiving the feedback should have a goal and then get information on their actions related to that goal (Wiggins, 2012). “Goals provide a sense of purpose and priority in an environment where a multitude of tasks can seem equally important and overwhelming. Clear goals focus attention and effort and enable individuals, groups, and organizations to use feedback to regulate their performance (Robinson et al., 2008, p. 661). In the school setting it is imperative that learners be reminded about the specific goals and criteria so they can assess where they are in reaching their goals (Wiggins, 2012). Wiliam (2012) and Brookhart (2012) stated that effective feedback should focus specifically on the goal and not on the ego of the person receiving the feedback.

Brookhart (2012) contended that supervisors can assume that the person understands what they did correctly or did not do correctly, and that the person should only be given feedback on what they did wrong. In fact, according to Chappuis (2012), feedback should offer specific information on the strengths of their efforts and also draw attention to problems the participant should address in relation to the goal. Before setting a goal it is important to understand what prior skills each person possesses, so that challenges do not surpass their prior knowledge and feedback can be targeted on the desired result (Brookhart, 2012; Hattie, 2012). Giving too much highly-technical feedback to beginning practitioners will only lead to confusion and frustration (Wiggins, 2012). Effective feedback that instructs the recipient should be related to the goal and should bridge the gap between what the recipient comprehends and what was intended to be comprehended (Sadler, 1989). Brookhart (2012) explains that feedback needs to be specific to the point the person understands what should happen next, but is vague enough to prompt reflective thinking. However, Wiliam (2011) warned that feedback that is too vague can lead to the negative effects of uncertainty, decreased motivation, and also a decline in learning. For example, providing learners with written comments as feedback resulted in significantly higher performance than that of learners provided with just a numeric score (Wiliam, 2011).

#### *Attainable and Actionable Feedback*

Attainable feedback provides information about what can help a person progress toward their goal and must be accepted by the person receiving the feedback (Wiggins, 2012). A person's effort towards attaining a goal will likely increase if he feels that he is not too far from the intended goal; he can focus on the things that are within his control

(Wiliam, 2012). More work is required of the person receiving the feedback than the person giving the feedback and this leads to growth (Wiliam, 2012).

When giving feedback, the evaluator should take into consideration how much the person can reasonably act on (Chappuis, 2012, Wiggins, 2012). People have different capacities to which they can respond to feedback. A person receiving too much feedback at one time may shut down completely ensuring that no further action will take place (Chappuis, 2012). Furthermore, even if the feedback that is given is specific and actionable, it must be understood by the person it is intended to help (Wiggins, 2012).

### *Timely Feedback*

One of the worst things an evaluator can do is provide detailed feedback days, weeks or even months after the performance has been completed (Wiggins, 2012). Delayed feedback is problematic and in most situations the sooner the learner can receive the effective feedback the better (Brookhart, 2012; Chappuis, 2012; Tovani, 2012; Wiggins, 2012). Timely feedback allows the learner to think or reflect on his performance (Brookhart, 2012). If feedback is not given in a timely fashion, especially during difficult concepts, the learner is at risk of developing misconceptions (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Timely feedback is given to learners when there is still an opportunity to apply the feedback (Chappuis, 2012; Tovani, 2012). Marzano et al., (2011) stated that focusing feedback on specific classroom strategies and behaviors within a set timeframe is instrumental in developing teacher's expertise. Ironically, timely feedback can present a paradox. Feedback that is given too quickly and frequently can result in the recipient relying on the person providing the feedback to

consistently provide help, which diminishes their ability to become self sufficient (Bangert-Drowns et al., 1991).

### *Consistent and Credible Feedback*

In order for feedback to be effective it must also be consistent and credible (Wiggins, 2012). Learners need information that is stable, accurate, and trustworthy so they can adjust important aspects of their performance successfully (Wiggins, 2012). Multiple evaluators should judge performances based on highly descriptive rubrics, and exemplars to ensure the consistency of their expectation on work (Wiggins, 2012). Teachers may use feedback only when they believe it will improve their practice. Therefore, feedback is more likely to be viewed as consistent and credible when it is, a) aligned to what the teacher views as best practices, b) parts of the system providing feedback logically connect, c) the process of scoring is reliable, and d) the indicators actually help students learn better (Cantrell & Scantlebury, 2011).

### Mississippi Statewide Teacher Appraisal Rubric (M-STAR)

The Mississippi Department of Education (MDE) has a goal of providing quality instruction for all students and improving student achievement across the state. Research has proven that a student's achievement gains are significant and lasting when they receive instruction from a high quality teacher (Rivkin et al., 2005). The MDE has adopted this research and is resolute in providing effective teaching to every Mississippi student (MDE, 2012).

In June 2010, MDE used the Mississippi Teacher Center to commission the establishment of the Statewide Teacher Evaluation Council (STEC) for the purpose of recommending a framework for the development of a statewide evaluation process for



teachers (MDE, 2012). Members of the STEC included teachers, administrators, and representatives from preparation programs, teacher unions, community organizations, the superintendent's organization, and the governor's office (MDE, 2012). The STEC met to develop guiding principles for the implementation of an effective teacher evaluation program, and they created recommendations to MDE on the framework for the new teacher evaluation system (MDE, 2012). The guiding principles were derived from discussions that identified characteristics of excellence in teachers, principals, and schools, and set parameters for the council's recommendations on the evaluation framework (MDE, 2012). In order to make better recommendations about the framework of the evaluation program, the members of the STEC discussed national initiatives dealing with how to determine student growth, professional development for teachers, teachers' career ladders, and systems for performance based compensation (MDE, 2012). Information on the United States Department of Education's funding of Race to the Top (RTT), Teacher Incentive Fund (TIF) grants, School Improvement Grants (SIG), and systems for student value added data were also presented to the STEC during these meetings (MDE, 2012). Evaluation systems from within the state of Mississippi and from states receiving the highest scores from the first round of RTT were also examined during these meetings (MDE, 2012).

The following guiding principles were finalized and adopted, in order of importance, by the STEC members. An effective Mississippi teacher evaluation system should:

1. Drive growth in student achievement at the classroom, department, school, and district levels.

2. Focus on effective teaching and learning based on national and state standards that target high expectations and meet the diverse needs of every learner.
3. Use multiple rating tools to assess levels of productivity, including a) measures of teamwork and collaboration; b) student assessment data including student growth; c) school and classroom climate; d) leadership.
4. Include comprehensive training on evaluation system components that provide fair, transparent scoring mechanisms and produce inter-rater reliability
5. Promote and guide individual and collaborative professional learning and growth based on educator content knowledge and the use of research established best practices and technology.
6. Provide appropriate data to differentiate compensation in a fair and equitable manner.
7. Differentiate the evaluation process based on the educator's expertise and student assessment results.
8. Provide appropriate and timely feedback at multiple levels to detect individual and systemic strengths and weaknesses. (MDE, 2012, p. 491)

These guiding principles were consistently referenced throughout the work by the STEC to ensure that their recommendations were consistent to their foundational statements

(MDE, 2012). The guiding principles along with the work of the STEC led to the development of the statewide teacher evaluation process for the state of Mississippi.

The Mississippi Statewide Teacher Appraisal Rubric (M-STAR) process of evaluating teachers was designed to give evaluators an understanding of teachers' strengths and weaknesses by using multiple methods of evaluation (MDE, 2012). The M-STAR process manual states that should consist of:

#### Formal classroom observations

- There will be a minimum of two formal observations per school year.
- Formal observations will be announced and scheduled in advance with the teacher.
- The first formal observation should be completed during the first half of the school year; the second should be completed during the second half of the school year.
- At least one observation will be performed by an administrator.
- The second observation will be performed by either an administrator or other trained evaluator.
- All formal observations will include a pre-observation conference and a post-observation conference.

#### Pre-observation and post-observation conferences

- The pre-observation conference should happen within one to two days prior to the observation. This conference provides the opportunity for the teacher to describe the context and plans for the class session and to provide initial artifacts.

- The post-observation conference should happen as soon after the observation as possible and no later than one week after the observation. This conference provides the opportunity for the evaluator to provide feedback, discuss areas for improvement, and create a professional development plan.

#### Informal “walkthrough” observations

- There will be a minimum of five informal observations during the school year.
- Informal observations will be unannounced, and each observation will last 5 to 15 minutes.
- Informal observations will be used as a means to inform instructional leadership functions of the school administrator by providing quick checks of teacher performance and feedback on that performance.

#### A review of artifacts

- Artifacts should include existing materials only; teachers should not create artifacts solely for the purpose of the artifact review.
- Lesson plans are required for the artifact review. Teachers must submit their lesson plan to their evaluator at least 24 hours prior to the pre-observation conference.

#### Teacher self-assessment

- Teachers will use the M-STAR rubric for self-assessment.
- Teacher self-assessment will be discussed during the summative evaluation conference.

#### Student survey

- The student survey will be given once during the school year. (MDE, 2012, pp. 405-406)

In the spring of 2011 MDE and the STEC collaborated with the American Institutes for Research (AIR) to create a draft form of the teacher evaluation instrument (MDE, 2012). This draft contained five domains. Each domain is detailed below:

#### Domain I: Planning

1. Plans lessons that demonstrate knowledge of content and pedagogy.
2. Plans lessons that meet the diversity of students' backgrounds, cultures, skills, learning levels, language proficiencies, interests, and special needs.
3. Selects instructional goals that incorporate higher level learning for all students.
4. Plans units of instruction that align with Mississippi Curriculum Framework or, when applicable, the Common Core State Standards.

#### Domain II: Assessment

5. Collects and organizes data from assessments to provide feedback to students and adjusts lessons and instruction as necessary.
6. Incorporates assessments into instructional planning that demonstrates high expectations for all students.

#### Domain III: Instruction

7. Demonstrates deep knowledge of content during instruction.
8. Actively engages students in the learning process.
9. Uses questioning and discussion techniques to promote higher order thinking skills.

10. Brings multiple perspectives to the delivery of content.

11. Communicates clearly and effectively.

#### Domain IV: Learning Environment

12. Manages classroom space and resources effectively for student learning.

13. Creates and maintains a climate of safety, respect, and support for all students.

14. Maximizes time available for instruction.

15. Establishes and maintains a culture of learning to high expectations.

16. Manages student behavior to provide productive learning opportunities for all students.

#### Domain V: Professional Responsibilities

17. Engages in continuous professional development and applies new information learned in the classroom.

18. Demonstrates professionalism and high ethical standards; acts in alignment with Mississippi Code of Ethics.

19. Establishes and maintains effective communication with families.

20. Collaborates with colleagues and is an active member of a professional learning community in the school. (MDE, 2012, pp. 407-408)

These standards were recognized by the STEC as important for Mississippi's teachers and were in line with national standards and practice (MDE, 2012). Each standard contained detailed descriptor information for each performance level of distinguished, effective, emerging, and unsatisfactory based on a number of resources including the Danielson Framework, the National Board standards, and the Interstate New Teacher Assessment and Support Consortium (INTASC) standards (MDE, 2012).

Training on how to use and how to score M-STAR evaluations was provided to all of the Mississippi school district employees involved with evaluating teachers (MDE, 2012). The trainings provided evaluators with information on a) the concept of multidimensional performance, b) the opportunity to practice using the rubric and scoring with the rubric, c) initial calibration of teacher rating through an exercise, d) a discussion on common errors on rating teachers, and e) finally a recalibration of ratings to ensure inter-rater reliability (MDE, 2012). Teachers should also receive training on the M-STAR performance domains and standards before they are formally observed (MDE, 2012).

Both a group of external expert practitioners, and a group of expert practitioners from Mississippi, were asked to provide feedback on the teacher appraisal framework to ensure it captured and reflected teacher practice (MDE, 2012). To establish content validity AIR suggested using common methods of relying on the input of the subject matter experts classified by knowledge about the field or experience in a particular position (MDE, 2012). AIR recommended a range of twenty-five to fifty subject matter experts from different regions of the state to participate in giving feedback through scales and discussion on the importance of each performance standard, the relevance of the instrument to a teacher's duties, pros and cons of evaluating a teacher's behaviors, and potential issues of fairness of each instrument (MDE, 2012). To determine construct validity AIR recommended an examination of the instrument using a multi-trait multi-method approach to determine the extent to which the domains of teacher performance were measured reliably despite the person doing the rating (MDE, 2012).

A panel of subject matter experts was assembled in September 2011 by AIR to begin the validation of the new performance standards, rubric, and guidelines on the evaluation process (MDE, 2012). The M-STAR framework was made available for public comment by MDE and in November of 2011 the Mississippi State Board of Education approved the framework to be piloted in ten schools across the state (MDE, 2012). In January 2012, MDE assembled administrators and master teachers from the pilot schools for training on the purpose and use of the framework that aimed to produce inter-rater reliability (MDE, 2012).

Also, from January to May of 2012, MDE sought stakeholder feedback from over two thousand teachers, K-12 administrators, deans of colleges, and professors, conducting over twenty focus group meetings that were held to gain feedback on the M-STAR evaluation system. Meetings were also held with teachers in subjects and grades that are not tested under the state accountability system in order to gain feedback on methods that best measured student growth in their areas (MDE, 2102). In the summer of 2012, MDE selected a group of trainers to take part in training that consisted of three days of classroom instruction and two days of observing and evaluating classroom teachers to ensure inter-rater reliability among the selected trainers (MDE, 2012). After completing this training of the trainers, the members received the necessary credentials to begin providing training during the 2012-2013 school year through five regional service agencies to the school and district level evaluators (MDE, 2012). This training prepared the school and district level evaluators to be field tested during the 2013-2014 school year.



## Teacher Perception of Teacher Evaluation

Teachers generally see evaluations of their instruction as a critique of them personally. Teachers have a tendency to derive a sense of self-worth from their profession which leads them to view their evaluations as a gauge of who they are personally (Barnett, 2006). Positive teacher perceptions regarding the implementation of a teacher evaluation depends on the consistency of an administrator's approach, integration of the teacher evaluation system into the administrator's instructional leadership, and credibility of the content within the teacher evaluation (Kimball, 2002). Additionally, teacher perceptions on the standards used in defining a quality teacher influence how positively teachers evaluate the effect of their effort and performance on their evaluations (Conley, Muncey, & You, 2005). When teachers believe that the standards reflect quality teaching and the evaluation system was administered fairly, and then teacher reactions are likely to be favorable (Milanowski & Heneman, 2001). Further, Ingvarson and Chadbourne (1997) stated that when a teacher evaluation contributes to the satisfaction of a teacher's career, the culture of the school is more likely to be supportive of a teacher evaluation system. When teacher evaluation standards are understood and are relevant, the teacher evaluation contributes more to career satisfaction than just when a teacher is just satisfied with their evaluation outcome (Conley et al., 2005). The way teachers perceive evaluations as useful and fair should be taken into consideration when implementing new systems, if these new evaluation tools are to gain acceptance.

Administrators also play a role when it comes to a teacher's perception of their evaluations. The attitude of the supervising administrator toward the teacher evaluation

process has an impact on teachers' attitude toward the evaluation process and ultimately their success with the process (Davis, Pool, & Mits-Cash, 2000). Milanowski and Heneman (2001) noted that negative attitudes towards teacher evaluation could be attributed to the teachers' perception of the administrator's unwillingness collaborate in the teacher evaluation process. Administrators indirectly influence teacher performance by helping teachers feel that they can help students become successful (Ebmeier, 2003).

## CHAPTER III - METHODOLOGY

The purpose of this study was to quantify the perceptions held by South Mississippi public school teachers concerning the Mississippi Statewide Teacher Appraisal Rubric (M-STAR) teacher evaluation system. General demographic information such as grade range being taught, number of years teaching experience, teaching tested or non-tested subject, subject being taught, highest degree obtained, whether or not the teacher was National Board Certified, the teacher's total number of M-STAR evaluations, and the teacher's total time involved in M-STAR training or professional development was collected. The focus of this study was to measure teacher perception towards the validity, reliability, feedback given from the M-STAR evaluation system, and their overall perception of the M-STAR evaluation system. An increase in the understanding of the perceptions held by educators concerning the M-STAR teacher evaluation process will inform the use of the current system and the design of future systems. This chapter presents a description of the participants, instrumentation, procedure for collecting data, and an analysis of data.

### Research Design

The research design of the study included the dependent variables of teachers' perception of reliability, validity, feedback received, and the overall perception of the M-STAR teacher evaluation system. Demographic data included the independent variables of grade range taught, total years of teaching experience, subject or grade that is tested under the state accountability system, subject area currently teaching, highest degree obtained, and whether or not the teacher is national board certified. All demographic data

was collected at the time the participants answered the statements on the survey instrument.

### Participants

There were four school districts in the Southern region of the state of Mississippi that participated in this study. Participating school districts were selected by convenience sampling based on their location being an easily accessible distance in regard to the researcher's current location. The four participating school districts contained approximately 1,260 certified teachers. According to the latest state accountability rating information, two of the school districts were rated A districts, one of them was rated a B district, and the last school district was rated a C district. The participants were certified kindergarten through twelfth grade teachers employed in these four school districts.

### Instrumentation

The 40 question Likert-style survey was developed to obtain quantitative data related to teacher perceptions of the M-STAR teacher evaluation system. A five-point scale ranging from Strongly Agree to Strongly Disagree with an option of no opinion or not enough information to respond was used. This format allowed for more accurate discrimination, and permitted a distinction between degrees of agreement and disagreement. All responses marked "no opinion or not enough information to respond" were not used in any statistical calculation.

The instrument was examined by a panel of experts for both content and face validity. The panel of experts was made up of a member of the Statewide Teacher Evaluation Council (STEC) mentioned in Chapter II, a retired district level administrator, and a current teacher. The member of the STEC is a retired middle school principal who

holds a specialist degree in educational leadership and worked in school across the state of Mississippi as a consultant for the implementation on M-STAR. This person also helped create and lead the M-STAR training modules that the Mississippi Department of Education (MDE) hosted across the state. Before her retirement, the district level administrator led the implementation of the M-STAR teacher evaluation system in her district. The teacher is a high school English teacher with 10 years of experience, National Board Certified, and has a Ph.D. in secondary education.

The Mississippi Statewide Teacher Appraisal Rubric (M-STAR) teacher evaluation system survey instrument contains five sections (Appendix C). The first section consisted of statements 1 through 8 and was designed to collect demographic information from the teacher responding to the survey. This information was used as the independent variables of this study and consisted of grade range currently teaching, total years of teaching experience, if the teacher was teaching a subject or grade currently under the state accountability system, the subject the respondent is currently teaching, highest degree the respondent has, if the respondent is a National Board Certified Teacher, the total number of observations they have participated in, and the amount of time the respondent has spent in training or professional development on the M-STAR teacher evaluation system. The second section consisted of statements 9 through 16 and was designed to measure teachers' perception of the validity of the M-STAR teacher evaluation system. The third section consisted of statements 17 through 21 and was designed to measure teachers' perception of the reliability of the M-STAR teacher evaluation system. The fourth section consisted of statements 22 through 33 and was designed to measure teachers' perception of the feedback given from the M-STAR

teacher evaluation system. The fifth and final section consisted of statements 34 through 40 and was designed to measure teachers' overall perception of the M-STAR evaluation system. The responses in sections two, three, four, and five were used to derive coefficients for each section that represents the dependent variables used to measure teachers' perception of the M-STAR teacher evaluation system in relation to the independent variables in section one.

A pilot study was conducted to analyze the reliability of the instrument being used in this study. Cronbach's alpha was used to estimate the internal-consistency reliability of the instrument. Cronbach's alpha was computed for each section of the instrument producing a composite score using pair-wise deletion for missing data. A link to a pilot study along with an explanation of the intent of the study was sent out to a random sample of approximately 30 teachers. A total of 17 responses were submitted electronically and correlation techniques were used to analyze the data to determine if changes needed to be made to the instrument before the actual study. In the second section of the pilot instrument, questions 9 through 16, teachers' perception of the validity of M-STAR, one response was excluded from the calculations. A Cronbach's alpha of .861 indicated an adequate reliability for this section. In the third section of the pilot instrument, questions 17 through 21, teachers' perception of the reliability of M-STAR, five responses were excluded from the calculations. A Cronbach's alpha of .881 indicated an adequate reliability for section 3. In the fourth section of the pilot instrument, questions 22 through 33, teachers' perception of feedback from M-STAR, three responses were excluded from the calculations. A Cronbach's alpha of .929 indicated a more than adequate reliability for this section. The last section of the pilot

instrument, questions 34 through 40, teachers' overall perception of M-STAR there were four responses excluded. A Cronbach's alpha of .663 caused some concern to the reliability of this section. Upon further inspection of the individual items in this section, it was determined that if item number 37 were deleted the adjusted Cronbach's alpha for this section would be .767. The increase in the alpha level was determined to be inconsequential and the decision was made to leave item 37 in the instrument.

Similar results were found in conducting the actual study. The Cronbach's alpha in section two of the instrument rose to .918. In section three decreased slightly to .832, and section four increased to .947. The concerns for the last section of the instrument in the pilot study were dismissed due to having a larger sample size the Cronbach's alpha for this section was an acceptable .750.

#### Procedures

The Mississippi Statewide Teacher Appraisal Rubric (M-STAR) teacher evaluation survey was created in electronic form so that it could be disseminated to the participating districts. A letter was drafted requesting school superintendents permission to survey teachers in their districts who are currently being evaluated using M-STAR (Appendix A). An email containing this letter, a copy of the consent letter (Appendix B), and a copy of the survey instrument (Appendix C) was sent to five school districts in South Mississippi. Four superintendents responded granting their permission to conduct this study and one did not respond. After gaining approval from the Institutional Review Board (Appendix D), an email containing a link to the survey was sent to the participating superintendents or their designee to be forwarded in the manner the felt was

best to the teachers in their perspective districts. The survey was left open to accept responses for thirty days from the day the first email was sent.

### Data Analysis

Using SPSS statistical software, the researcher created a data file from the completed instruments. Descriptive statistics were calculated to determine the rating of each item on the survey (frequencies, means, and standard deviations). To test the hypotheses of this study, a multivariate analysis of variance (MANOVA) was utilized. An alpha level of .05 was used in all testing of the hypotheses. Next, a MANOVA test determined significant correlations between the dependent variables of the Validity Index (the perceptions of the participants to the validity of the M-STAR evaluation system), the Reliability Index (the perceptions the participants to the reliability of the M-STAR evaluation system), Feedback Index (the perceptions of the participants to the feedback from the M-STAR evaluation system), and the Perception Index (the overall perception of the participants towards the M-STAR evaluation system), and how they related to the independent variables: total years of teaching experience; subject or grade that is tested under the state accountability system; total number of M-STAR observations (formals and informal/walkthroughs) the teacher has participated in; and total time the teacher has been involved in training or professional development on the M-STAR.



## CHAPTER IV - RESULTS

This chapter presents findings from data collected through the use of an electronic survey. The goal of this study was to examine the resultant data to determine the perceptions' teachers have about the validity, the reliability, the feedback received, and their overall perception of the Mississippi Statewide Teacher Appraisal Rubric (M-STAR) evaluation system. These dependent variables were measured based on the teacher's years of teaching experience, whether or not the teacher was teaching in a subject or grade that is tested under the state accountability system, the total number of observations the teacher has been involved in, and the total amount of M-STAR professional development hours the teacher received. There were approximately 1,260 electronic surveys distributed to 4 school districts in the Southern region of Mississippi. Of these 1,260 distributed surveys, 430 were submitted before the acceptance of responses was turned off resulting in a 34% return rate. The results of examining and analyzing this data are presented in this chapter.

### Descriptive Information of the Sample

Section 1 of the survey instrument collected demographic data from the 430 respondents. The data included: the grade range that best aligned with the teacher's current teaching assignment (grade); the total years of teaching experience the teacher has (experience); whether or not the teacher teaches in a test subject (Tested); subject area being taught (subject); the highest degree obtained by the teacher (degree); whether or not the teacher is National Board Certified (national board); the teacher's total number of M-STAR observations (observations); and the total time of professional development or training the teacher has participated in (professional development). Table 1 provides the

results of the demographic data and shows that responses were relatively equivalent across all grade levels. Table 1 also reflects the majority of the respondents had more than 10 years of teaching experience along with a higher percentage teaching in a tested subject area. More English Language Arts teachers responded to the survey instrument than teachers in other subject areas. The majority of the respondents held a Master’s degree and 13% of the respondents were Nationally Board Certified. Finally, approximately 64% of the teachers in this study have participated in 6 or more observations, and the majority of the teachers in this study have received between 1 to 4 hours of professional development or training on M-STAR.

Table 1 *Frequencies and percentages of demographic variables*

Variable	Frequency	Percentage
<b>Grade</b>		
Kindergarten – 2 <sup>nd</sup> Grade	98	22.8
3 <sup>rd</sup> – 4 <sup>th</sup> Grade	82	19.1
5 <sup>th</sup> – 6 <sup>th</sup> Grade	63	14.7
7 <sup>th</sup> – 8 <sup>th</sup> Grade	56	13.0
9 <sup>th</sup> – 12 <sup>th</sup> Grade	131	30.5
<b>Experience</b>		
0 – 2 Years	38	8.8
3 – 6 Years	76	17.7
7 – 10 Years	77	17.9
11 – 15 Years	90	20.9

Table 1 (continued).

Variable	Frequency	Percentage
16 – 20 Years	56	13.0
21+ Years	93	21.6
Tested		
No	186	43.3
Yes	244	56.7
Subject		
English Language Arts	120	27.9
Math	71	16.5
Social Studies	19	4.4
Science	30	7.0
Special Ed	73	17.0
Elective	45	10.5
Other	72	16.7
Degree		
Bachelors	163	37.9
Masters	261	60.7
Specialist	4	0.9
PhD	2	0.5
National Board		
No	373	86.7

Table 1 (continued).

Variable	Frequency	Percentage
Yes	57	13.3
Observations		
None	10	2.3
1 – 2 observations	22	5.1
3 – 5 observations	123	28.6
6 – 9 observations	117	27.2
10+ observation	158	36.7
Professional Development		
None	18	4.2
1 – 2 hours	151	35.1
2 – 4 hours	140	32.6
4 – 6 hours	56	13.0
6+ hours	65	15.1

Sections 2 through 5 of the survey instrument contained statements that used a Likert scale ranging from Strongly Agree to Strongly Disagree with an option of No Opinion or Not Enough Information. For the purpose of running the statistical analysis answers of Strongly Agree was coded as 4, Agree was coded as 3, Disagree was coded as 2, and Strongly Disagree was coded as 1. Answers of No Opinion or Not Enough Information were coded as 0 and left out of the statistical analysis.

Section 2 of the survey instrument contained 8 statements that were designed to measure each teacher’s perception of the M-STAR evaluation system’s validity. Table 2 shows the means for the statements in this section ranged from statement 14 with a mean of 2.58 to statement 10 with a mean of 2.87. The largest standard deviation of .73 was found in statement 14.

Table 2 *Descriptive Statistics for Teachers’ Perception of M-STAR Validity*

Statement	N	Mean	Std. Dev.
9. The standards used in the M-STAR evaluation system are fair.	402	2.74	.67
10. Working towards improving my performance on the M-STAR evaluation standards will also help me to improve the quality of my instruction.	403	2.87	.71
11. The M-STAR evaluation instrument includes clear explanations for each performance level descriptor.	410	2.76	.69
12. The four M-STAR levels of performance: Unsatisfactory, Emerging, Effective, and Distinguished are adequate.	413	2.80	.66
13. The M-STAR descriptors focus on the key teacher behaviors that positively impact student learning.	396	2.85	.60
15. The M-STAR instrument provides teachers with objective information about their teaching.	408	2.63	.68
16. The M-STAR instrument incorporates indicators of student learning in the evaluation process.	397	2.72	.64

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

Section 3 of the survey instrument contained 5 statements that were designed to measure each teacher's perception of the M-STAR evaluation system's reliability. Statement 9 was also used in the statistical analysis of the perception of M-STAR reliability. Table 3 shows the means for the statements in this section ranged from statement 20 with a mean of 2.38 to statement 10 with a mean of 2.93. Along with having the lowest mean in this section, statement 20 also had the largest standard deviation of .81.

Table 3 *Descriptive Statistics for teachers' perception of M-STAR reliability*

Statement	N	Mean	Std. Dev.
9. The standards used in the M-STAR evaluation system are fair.	402	2.72	.67
17. I understand the meaning of each descriptor and level of performance used in the M-STAR evaluation instrument.	413	2.79	.62
18. My evaluators have been adequately trained to consistently evaluate my teaching.	376	2.93	.75
19. I am confident that evaluators at my school interpret and score teacher evaluations in a similar manner.	388	2.74	.78
20. I am confident that evaluators from other schools in the district interpret and score teacher evaluations in a manner similar to my school administrators.	301	2.38	.81
21. The scores from my evaluations have been consistent from one evaluator to another.	380	2.88	.72

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

Section 4 of the survey instrument contained 12 statements that were designed to measure each teacher's perceptions of the importance and quality of feedback given from the M-STAR evaluation system. Table 4 shows the means for the statements in this section ranged from the lowest mean of 2.58 coming from statement 33 to highest coming from statement 23 with a mean of 3.19. The largest standard deviation in this section, .75, was found on statements 28 and 30.

Table 4 *Descriptive Statistics for teachers' perception of M-STAR feedback*

Statement	N	Mean	Std. Dev.
22. The M-STAR instrument provides specific feedback that can help guide individual professional development plans.	405	2.77	.67
23. The most important purpose of teacher evaluation is to provide feedback for improving classroom instruction.	424	3.19	.65
24. The process used under the M-STAR system fosters a climate for instructional improvement.	412	2.72	.70
25. The M-STAR instrument provides teachers with objective information about their teaching.	412	2.73	.67
26. The M-STAR system enhances dialogue and mutual understanding between teachers and evaluators about effective teaching.	400	2.67	.73
27. The M-STAR system increases teacher reflection on choices of teaching strategies.	402	2.73	.66

Table 4 (continued).

Statement	N	Mean	Std. Dev.
28. I have received regular focused follow-up and instructional support based on my M-STAR evaluations.	411	2.73	.75
29. The feedback I have received from my M-STAR evaluations has been valuable.	410	2.75	.73
30. As a result of the feedback I have received from my M-STAR evaluations, I have improved my ability to design high quality lessons.	400	2.60	.75
31. As a result of the feedback I have received from my M-STAR evaluations, I have improved the quality my overall instruction.	404	2.64	.72
32. The M-STAR rubrics provide feedback that can help guide individual professional development plans for all teachers.	399	2.73	.64
33. As a result of the feedback I have received from my M-STAR evaluations, I have increased student learning.	374	2.58	.72

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

The last section of the survey instrument, section 5, contained 7 statements that were designed to measure each teacher's overall perception of the M-STAR evaluation system. Table 5 shows the means for the statements in this section ranged from the lowest mean of 2.45 in statement 34 to highest on statement 39 with a mean of 2.92. The largest standard deviation in this section, .81, was found on statement 36.



Table 5 *Descriptive Statistics for teachers' overall perception of M-STAR*

Statement	N	Mean	Std. Dev.
34. The M-STAR evaluation system recognizes each teacher's contribution to the school as a whole (e.g. relationships with coworkers, families, professional development, and document completion).	385	2.45	.80
35. The M-STAR evaluation process is helpful to my professional growth.	405	2.61	.73
36. The most important purpose of the M-STAR evaluation is to fulfill human resource requirements for continued employment.	352	2.52	.81
37. The M-STAR evaluation incorporates indicators of student learning in the evaluation process.	394	2.76	.58
38. The process used under the M-STAR evaluation system fosters a climate for instructional improvement.	403	2.75	.66
39. All teachers should be evaluated at least twice a year to provide feedback on instructional improvement.	408	2.92	.75
40. I focus my professional development efforts on activities that directly help me achieve the M-STAR evaluation standards.	397	2.57	.74

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

## Statistical Results

This study was a quantitative study that investigated whether or not a statistically significant difference existed between the independent variables of; the teachers' years of teaching experience, the teacher teaches a tested or non-tested subject, the total number of M-STAR observation, and the amount of time the teacher has been involved in M-STAR training or professional development; and the dependent variables of; the teachers' perception of the M-STAR's validity, perception of M-STAR's reliability, perception of the feedback given by M-STAR, and the teachers' overall perception of the M-STAR teacher evaluation system. This study used data collected through electronic surveys sent to certified kindergarten through twelfth grade teachers in public school districts in the Southern region of the state of Mississippi. A multivariate analysis of variance (MANOVA) statistical test was used to determine if statistically significant differences existed between each independent variable and each of the dependent variables.

H1: There is a statistically significant difference in teachers' perception of the reliability, validity, quality of feedback received, and the overall M-STAR teacher evaluation system based on years of teaching experience. Using Pillai's Trace, it was determined that there was not a significant difference in perceptions based on a teacher's years of teaching experience and their perception of the validity, their perception of the reliability, their perception of the feedback, and their overall perception of the M-STAR teacher evaluation system,  $V = .06$ ,  $F(20,1676) = 1.354$ ,  $p = .135$ . Table 6 contains the means and standard deviations on the dependent variables.

Table 6 *Teaching experience descriptive statistics*

	Mean	Std. Dev.	N
<b>Validity</b>			
0 - 2 years	2.75	.51	37
3 - 6 years	2.75	.51	75
7 - 10 years	2.78	.46	76
11 -15 years	2.81	.52	88
16 - 20 years	2.68	.60	56
21 + years	2.64	.58	93
Total Validity	2.73	.53	425
<b>Reliability</b>			
0 - 2 years	2.71	.65	37
3 - 6 years	2.78	.65	75
7 - 10 years	2.78	.56	76
11 -15 years	2.75	.48	88
16 - 20 years	2.71	.52	56
21 + years	2.77	.61	93
Total Reliability	2.76	.57	425
<b>Feedback</b>			
0 - 2 years	2.78	.59	37
3 - 6 years	2.85	.49	75

Table 6 (continued).

	Mean	Std. Dev.	N
7 - 10 years	2.76	.52	76
11 -15 years	2.79	.49	88
16 - 20 years	2.66	.56	56
21 + years	2.60	.61	93
Total Feedback	2.74	.55	425
Perception			
0 - 2 years	2.69	.51	37
3 - 6 years	2.73	.45	75
7 - 10 years	2.68	.44	76
11 -15 years	2.69	.41	88
16 - 20 years	2.62	.52	56
21 + years	2.54	.48	93
Total Perception	2.66	.47	425

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

H2: There is a statistically significant difference in teachers' perception of the reliability, validity, quality of feedback received, and the overall M-STAR teacher evaluation system between teachers of subjects that are included in the Mississippi Accountability System and teachers of subjects that are not included. Using Pillai's Trace, no significant difference was found between teachers of subjects that are included in the Mississippi Accountability System and teachers of subjects that are not included with regards to their perception of the validity, the perception of the reliability, the

perception of the feedback, and the overall perception of the M-STAR teacher evaluation system,  $V = .01$ ,  $F(4,420) = 1.130$ ,  $p = .342$ . Table 7 contains the means and standard deviations on the dependent variables.

Table 7 *Tested subject area descriptive statistics*

	Mean	Std. Dev.	N
<b>Validity</b>			
No	2.74	.54	181
Yes	2.73	.53	244
Total Validity	2.74	.53	425
<b>Reliability</b>			
No	2.80	.57	181
Yes	2.72	.58	244
Total Reliability	2.76	.57	425
<b>Feedback</b>			
No	2.73	.57	181
Yes	2.74	.53	244
Total Feedback	2.74	.55	425
<b>Perception</b>			
No	2.66	.50	181
Yes	2.66	.44	244
Total Perception	2.66	.47	425

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

H3: There is a statistical significant difference in the total number of M-STAR evaluations a teacher receives and their perception of the reliability, validity, quality of feedback received, and the overall M-STAR teacher evaluation system. Significant differences were found in teachers' perception of the validity, the perception of the reliability, the perception of the feedback, and the overall perception of the M-STAR teacher evaluation system based on a teacher's total number of M-STAR evaluations,  $V = .08$ ,  $F(16,1680) = 2.137$ ,  $p = .005$ . Given the significant finding of the overall test, the univariate main differences were examined. Significant main differences for the total number of M-STAR observations were obtained for: the teachers' perception of validity,  $F(4,420) = 3.014$ ,  $p = .018$ ; teachers' perception of reliability,  $F(4,420) = 4.649$ ,  $p = .001$ ; teachers' perception of feedback,  $F(4,420) = 4.9$ ,  $p = .001$ ; and teachers' overall perception,  $F(4,420) = 3.734$ ,  $p = .005$ . Significant pairwise differences in the means of total number of M-STAR observations were found in the teacher perception of M-STAR validity, reliability, feedback, and overall perception using Tukey HSD. In the validity section, the mean from 3 to 5 observations was significantly lower than the mean from 10 or more observations. In the reliability section, the mean from 3 to 5 observations was significantly lower than both 6 to 9 observations and 10 or more observations, but because Levene's test of equality was violated the results should be interpreted with some caution. Similar findings were obtained with the feedback section; the mean from 3 to 5 observations was significantly lower than both 6 to 9 observations and 10 or more observations. Finally in the overall M-STAR perception section, the mean from 3 to 5 observations was significantly lower than the mean from 6 to 9 observations. Table 8

contains the means and standard deviations on the dependent variables with regard to the total number of observations the teacher had participated in.

Table 8 *Total number of observations descriptive statistics*

	Mean	Std. Dev.	N
<b>Validity</b>			
None	2.89	.52	8
1 - 2 observations	2.77	.52	21
3 - 5 observations	2.60	.54	122
6 - 9 observations	2.76	.53	116
10+ observations	2.81	.52	158
Total Validity	2.74	.53	425
<b>Reliability</b>			
None	3.08	.39	8
1 - 2 observations	2.87	.52	21
3 - 5 observations	2.58	.63	122
6 - 9 observations	2.81	.48	116
10+ observations	2.82	.57	158
Total Reliability	2.76	.57	425
<b>Feedback</b>			
None	3.08	.45	8
1 - 2 observations	2.85	.56	21
3 - 5 observations	2.57	.56	122

Table 8 (continued).

	Mean	Std. Dev.	N
6 - 9 observations	2.79	.48	116
10+ observations	2.80	.56	158
Total Feedback	2.74	.55	425
Perception			
None	2.89	.28	8
1 - 2 observations	2.78	.48	21
3 - 5 observations	2.54	.43	122
6 - 9 observations	2.73	.46	116
10+ observations	2.66	.49	158
Total Perception	2.66	.47	425

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

H4: There is a statistically significant difference in teachers' perception of the reliability, validity, quality of feedback received, and the overall M-STAR teacher evaluation system between teachers receiving one to two hours, two to four hours, four to six hours, six or more hours, or no training or professional development on the M-STAR teacher evaluation system. Significant differences were found in the teachers' perception of the validity, the perception of the reliability, the perception of the feedback, and the overall perception of the M-STAR teacher evaluation system based on the amount of training or professional development a teacher receives,  $V = .13$ ,  $F(16,1680) = 3.522$ ,  $p = .000$ . Given the significant finding of the overall test, the univariate main differences were examined. Significant main effects for the amount of training or professional



development on M-STAR were obtained for: the teachers' perception of validity,  $F(4,420) = 4.653$ ,  $p = .042$ ; teachers' perception of feedback,  $F(4,420) = 5.18$ ,  $p = .047$ ; and teachers' overall perception,  $F(4,420) = 5.422$ ,  $p = .049$ . Significant pairwise differences between the means in the amount of training or professional development on M-STAR were found in the teacher perception of M-STAR validity, feedback, and overall perception using Tukey HSD. In the validity section, the mean from 1 to 2 hours of professional development was significantly lower than the mean from both 4 to 6 hours and 6 or more hours of professional development. Also in regards to the validity section, the mean from 2 to 4 hours of professional development was significantly lower than 4 to 6 hours of professional development. In the feedback section, the mean from 1 to 2 hours of professional development was significantly lower than both 4 to 6 hours and 6 or more hours of professional development. It necessary to caution the reader of the findings on the amount of professional development and how it relates to the validity, reliability, and feedback of M-STAR due to the violation of Levene's test of equality. Finally in the overall M-STAR perception section, the mean from 1 to 2 hours professional development was significantly lower than both 4 to 6 hours and 6 or more hours of professional development. Also the mean of teachers overall perception of M-STAR with 2 to 4 hours of professional development was significantly lower than teachers both 4 to 6 hours and 6 or more hours of professional development. Table 9 contains the means and standard deviations on the dependent variables with regards to the amount of training or professional development a teacher received.

Table 9 *Professional development descriptive statistics*

	Mean	Std. Dev.	N
<b>Validity</b>			
None	2.70	.57	15
1 - 2 hours	2.66	.58	151
2 - 4 hours	2.68	.48	138
4 - 6 hours	2.93	.59	56
6+ hours	2.88	.38	65
Total Validity	2.74	.53	425
<b>Reliability</b>			
None	2.71	.50	15
1 - 2 hours	2.58	.59	151
2 - 4 hours	2.75	.57	138
4 - 6 hours	2.92	.51	56
6+ hours	3.05	.43	65
Total Reliability	2.76	.57	425
<b>Feedback</b>			
None	2.74	.5	15
1 - 2 hours	2.60	.58	151
2 - 4 hours	2.74	.49	138
4 - 6 hours	2.90	.57	56
6+ hours	2.90	.48	65

Table 9 (continued).

	Mean	Std. Dev.	N
Total Feedback	2.74	.55	425
Perception			
None	2.70	.44	15
1 - 2 hours	2.57	.49	151
2 - 4 hours	2.60	.41	138
4 - 6 hours	2.83	.47	56
6+ hours	2.80	.45	65
Total Perception	2.66	.47	425

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

## CHAPTER V - DISCUSSION

The purpose of this study was to examine the perceptions teachers held about the validity, the reliability, the feedback given to them by, and their overall perception of the Mississippi Statewide Teacher Appraisal Rubric (M-STAR). This study sought to find differences in the effects of a teacher's years of teaching experience, whether a teacher taught in a tested or non-tested subject, the total number of M-STAR evaluations, and the total number of M-STAR professional development hours provided to a teacher. By finding where statistical differences exist, school leaders can identify and implement strategies that may help improve teachers' perceptions of the M-STAR evaluation system. This chapter provides a summary of the procedures used, a discussion of the findings, conclusions, recommendations for policy, practice, and recommendation for future research.

### Summary of Procedures

After obtaining permission from superintendents in participating districts and The University of Southern Mississippi Institutional Review Board approval, an electronic survey was distributed to a sample population of public school teachers from four school districts located in South Mississippi. A total of 430 electronic surveys were collected from participants who volunteered their responses between March 30, 2015 and April 30, 2015. The survey instrument (Appendix C) collected descriptive data to measure the level to which participating teachers agreed with statements in the domains of validity, reliability, feedback, and overall perception of M-STAR. Cronbach's alpha was calculated from the items in each of the domains to measure the reliability of the items used to analyze the data. Finally, the data from the survey responses was analyzed to

determine if a teacher's years of teaching experience, whether a teacher teaches in a tested or non-tested subject, total number of M-STAR observations performed on a teacher, and the total number of M-STAR professional development hours provided to a teacher made a statistical difference in the perceptions of teachers regarding the M-STAR teacher evaluation system.

### Conclusions

Demographic data from the responding teachers about their years of teaching experience, whether or not they taught in a tested or non-tested subject, how many observations were performed on them, and the amount of M-STAR professional development time the teacher participated in were analyzed to gain a better understanding. With respect to teaching years of experience, 44.4% of the respondents indicated they had 0 to 10 years, 33.9% indicated they had 11 to 20 years, and 21.6% indicated they had 21 or more years. When asked whether or not the respondent taught in a subject tested under the state accountability system, 56.7% of the responding teachers indicated they taught in a state tested subject while 43.3% of the respondents indicated they taught in a non-tested subject. In response to the number of M-STAR observations that had taken place, 63.9% of the responding teachers had been observed using M-STAR 6 or more times 33.7% had been observed using the M-STAR system between 1 and 5 times, and 2.3% of the responding teachers claimed they had not been observed using the M-STAR system. The data from responding teachers revealed that 28.1% received 4 or more hours of M-STAR professional development on M-STAR, 67.7% received 1 to 4 hours, and 4.2% did not receive any M-STAR professional development.

The first hypothesis of this study posed there is a statistically significant difference in teachers' perception of the validity, reliability, quality of feedback received, and the overall M-STAR teacher evaluation system based on years of teaching experience. Teachers were asked to identify themselves with zero to two years, three to six years, seven to ten years, eleven to fifteen years, sixteen to twenty years, or twenty-one or more years of teaching experience. The survey results were examined using a Multivariate Analysis of Variance (MANOVA) to find if a statistical significant difference existed between the independent variables and the dependent variable of teaching years of experience. The statistical analysis revealed that no statistically significant difference existed between the number years of teaching experience a teacher had and their perceptions of the validity, reliability, feedback received, and their overall perception of the M-STAR teacher evaluation system. Even though no statistical significance difference was found, the researcher felt it was worth noting that the mean scores showed that teachers at all levels of teaching experience agreed that M-STAR is valid, reliable, provides impactful feedback, and they had a favorable perception of M-STAR. The survey results indicated that the perception of M-STAR's validity was highest with teachers that responded having eleven to fifteen years of teaching experience. Respondents with seven to ten years of teaching experience yielded the highest perception rating of M-STAR's reliability. Concerning the perception of the feedback given by the M-STAR, teachers with three to six years of teaching experience agreed the most that it impacted their instruction. Finally, teachers with three to six years of teaching experience also had a higher rating when it comes to their overall perception of the M-STAR evaluation system.

The second hypothesis of this study sought to find if differences in the perception of M-STAR's reliability, validity, quality of feedback received, and the overall perception M-STAR were statistically significant in teachers teaching in tested subjects versus teachers teaching in non-tested subjects in the Mississippi statewide accountability system. The results were analyzed using MANOVA and produced data showing that there was no statistical significant difference in the perception of the tested teachers and non-tested teachers. With further examination, the researcher found both tested teachers and non-tested teachers agree that the M-STAR teacher evaluation system is valid, reliable, provides adequate feedback, and have favorable perceptions of the overall M-STAR. The two groups were virtually the same in their ratings of the validity, feedback, and overall M-STAR system. The researcher noted that the largest non-statistical difference rating between the two groups was found in their perception of reliability. Teachers in non-tested subjects perceived M-STAR as more reliable.

The third hypothesis sought to determine if a statistical significant difference existed in the total number of M-STAR evaluations a teacher receives and teachers' perception M-STAR's reliability, validity, quality of feedback received, and overall perception of M-STAR. The respondents were asked to choose how many total M-STAR observations in which they had participated. Choices for this statement included none, one or two observations, three or four observations, six to nine observations, and the final choice was ten or more M-STAR observations. The results from the MANOVA revealed that a significant difference did exist. Tests of between subjects effects found the number of M-STAR observations performed on a teacher had a statistically significant impact on teachers' perception of validity, reliability, feedback, and the

overall M-STAR evaluation system. Post hoc testing revealed that teachers who participated in ten or more M-STAR observations had a statistically higher perceived validity of M-STAR than teachers who participated in three to five M-STAR observations. In addition, teachers who participated in six or more M-STAR observations rated the reliability of M-STAR higher than teachers who only participated in three to five M-STAR observations. Similarly, teachers that participated in six or more M-STAR observations had a statistically higher perception of the feedback given from their M-STAR observations than those with three to five M-STAR observations. The last significant difference produced by post hoc testing, revealed that teachers who chose six to nine M-STAR observations had a statistically higher perception of the overall M-STAR evaluation system than teachers marking three to five M-STAR observations.

The final hypothesis posed by this study asked if the amount of M-STAR professional development (PD) a teacher received made a statistically significant difference in teachers' perceptions of M-STAR's validity, the reliability, the feedback, and the overall perception of M-STAR. The MANOVA reflected that the amount of PD was statistically significant in a teacher's perception of M-STAR. Tests of between subjects effects were analyzed and statistical significant differences occurred in the respondents' perceptions of M-STAR's validity, feedback, and the M-STAR overall. No statistical significance was found between the amount of M-STAR PD and a teacher's perception of M-STAR reliability. Post hoc testing was used to determine where the statistical differences existed between teachers who received one to two hours of M-STAR PD, two to four hours of M-STAR PD, four to six hours of M-STAR PD, six or



more hours of M-STAR PD, and teachers that have not received any PD on the M-STAR. Teachers that received six or more hours of PD on M-STAR produced scores that were statistically higher in their perception of M-STAR's validity compared to teachers receiving only one to two hours of M-STAR PD. The perception of M-STAR validity was also statistically higher with teachers receiving four to six hours of M-STAR PD as opposed to teachers receiving between one and four hours of M-STAR PD. Also, teachers receiving the four to six hours of M-STAR PD perceived the feedback by their M-STAR observations at a statistically significant higher level than teachers receiving on one to two hours of M-STAR PD. Finally, the amount of M-STAR PD made a statistically significant difference in improving the overall perception of the M-STAR evaluation system when teachers received four to six hours of M-STAR PD compared to teachers only receiving one to four hours of M-STAR PD.

### Discussion

In order for teacher evaluation to have a positive impact on student learning, researchers have concluded that the evaluation process must meet three different criteria. First, the teacher evaluation process must be capable of removing poor performing teachers who fail to produce favorable student learning outcomes (Heneman et al., 2007; Koppich & Showalter, 2005; Odden & Wallace, 2008). Next, the teacher evaluation process should also produce meaningful feedback that teachers can use to improve their instructional practices therefore improving student learning (Heneman et al., 2007; Odden 2004; Sanders et al., 1998). Finally, the teacher evaluation process should foster a results-oriented school culture that supports a wider set of policies that ensure the quality of teaching and learning within a school (Ellett & Teddlie, 2003; Odden, 2004).

Stronge and Tucker (2003) argued that because teaching matters, teacher evaluation should matter and that any reform in education cannot succeed without capable, high quality teachers. Identifying capable, high-quality teachers cannot happen without a high quality teacher evaluation system (Stronge & Tucker, 2003). Tucker and Stronge (2005) suggested that the evaluation of teachers should not only document the level of performance in order to hold teachers accountable for their instruction, but also help them improve their instruction. It is important to have credibility in a teacher evaluation system in order to consistently define good instruction. Doing so heightens the value of conversations with teachers that develop from classroom observations (Danielson, 2010).

In order for a teacher evaluation system to be viewed as credible, proficiency levels from evaluators must be able to accurately judge teachers using a reliable, valid evaluation tool that provides feedback that is meaningful and produces productive conversations that improves instruction (Danielson, 2010). In general, teachers lack confidence in the ability of evaluations to improve their instruction because their evaluations are often brief and rushed, given the plethora of other administrative duties that the evaluator has to perform (Garth-Young, 2007). Also, some teachers perceive administrators as not using the evaluation process fairly or in such a manner as to terminate teachers the administrator does not like (Garth-Young, 2007). This perception leads to a lack of trust between the administrator and teachers, ultimately reducing the effectiveness of the evaluation process (Garth-Young, 2007). Administrators must juggle the limited time they are able to spend on teacher evaluations with poorly designed evaluation systems. These systems often do not provide meaningful feedback (Kersten &

Isreal, 2005). Schools must foster an environment of professional learning where teachers are responsible for continuous professional growth, but connecting teacher evaluation with professional development does not occur without work (Danielson & McGreal, 2000). School systems should seek opportunities to incorporate professional development with teacher evaluation procedures to promote, monitor, and determine teacher growth (Danielson & McGreal, 2000).

### Limitations

The generalizations from this study are limited to the population from which this sample was taken. This study relied on an instrument that was self-reporting and limited to the number of respondents who volunteered to participate. In terms of self-reported survey data for teachers, there are at least three potential threats to validity and reliability (Mayer, 1999): (a) The context and act of teaching and learning is so complex that it cannot be sufficiently distinguished by survey responses; (b) Survey items may include ambiguity or wording that skews responses; and (c) Teachers may be sensitive to particular items or concepts on the survey which in turns leads to responses that are not accurate but are considered socially desirable.

In addition, the sample of participants was from schools and districts in South Mississippi. This limitation restricts the researcher's ability to make generalizations about the findings applicable to all schools and districts in the state. Next, participants' previous work circumstances and evaluation history, whether good or bad, cannot be controlled. Participants' bias towards teacher evaluation may positively or negatively skew results. Finally, respondents were not given the option to make comments or

explain their choices. Consequently, this study focused solely on the teachers' perceptions reported in the survey and did not consider any additional objective data.

### Recommendations for Policy

When examining the results of this study it became apparent to the researcher that the four independent variables could be separated into two categories. These two categories were characterized by the researcher as being variables in which the teacher had the primary control for the outcome and variables where the outcome was primarily controlled by the school district in which the teacher worked. For example, a teacher can control when and how long they are employed in the teaching profession and therefore the teacher has primary control of their years of teaching experience. Likewise, a teacher has control over what subject areas they hold certifications in. Therefore, whether or not a teacher teaches in a tested subject area is also primarily under the teacher's control. For the purpose of further discussion, a teacher's years of teaching experience along with whether they teach in a tested or non-tested subject will be referred to simultaneously as teacher controlled variables.

A teacher's total number of M-STAR observations and the hours of M-STAR PD are primarily under the control of the school district in which the teacher is or was employed. Other than requesting an administrator to perform an M-STAR observation, a teacher has little to no control over the number of M-STAR observations performed in their classroom. The Mississippi Department of Education "suggests" (MDE, 2012, pp.405-406,) that administrators perform at least two formal observations during the school year. Ultimately, it is left up to the individual school district to set the required number of M-STAR observations.

Likewise, a teacher may find opportunities for professional development for M-STAR on their own. However, the school district decides whether or not the teacher will be allowed to take professional leave from work, pay for a substitute, pay registration fees that may be required, and also to reimburse any travel expenses the teacher may incur. It is for these reasons that it is believed by the researcher that the school district has primary control on the amount of M-STAR PD a teacher receives.

Due to statistically significant differences being found in the school controlled variables, school districts have the ability to positively affect their teacher's perceptions of the M-STAR evaluation system. Data from this research reflects that school districts wanting to improve their teacher's level of agreeance towards the validity, reliability, feedback provided, and the teacher's overall perception of their M-STAR evaluation should do the following:

- Increase the total number of M-STAR observations they require administrators to perform on a teacher to at least ten per school year.
- Increase the total amount of M-STAR professional development hours teachers receive to a minimum of six hours.

State departments of education can also use the findings of this study for guidance on recommendations or requirements they make for existing or future teacher evaluations systems. States wanting to improve existing or future perceptions teachers have about their statewide teacher evaluation system should also consider requiring both a minimum number of observations performed by school level administrators and a minimum number of professional development hours a teacher attends. Setting a minimum requirement in both of these areas should help states and school districts increase their teachers'

perception with the validity, reliability, feedback, and overall perception of their teacher evaluation system.

### Recommendations for Future Research

As a result of conducting this research additional statements emerged in areas the researcher felt would either continue or add to the findings of this study.

1. With the statistical significance of this study being found within the variables labeled as school controlled, it is recommended that continued research into whether or not similar variables that may be classified as school controlled have statistical significance. Research conducted on such similar variables will help to either support or deny the hypothesis that, to a certain extent, schools can positively affect the perceptions their teachers have about evaluation systems being used to measure their effectiveness.
2. With the results of this study showing teachers' perceive the M-STAR teacher evaluation system as validity, reliability, and have a favorable overall perception, future research could focus more on the feedback obtained by M-STAR. Are teachers receiving feedback in a manner that is timely enough to improve their instruction? Is the feedback given by the evaluator specific enough to lead to changes in the teacher's pedagogy? Is the feedback given to the teacher actionable, or in other words, can the teacher implement the feedback given by the evaluator?
3. Future research could also focus on the reliability of the M-STAR evaluation system. In Chapter III the difference between inter-rater reliability and inter-rater agreement were briefly explained. Differences

in M-STAR's inter-rater reliability and inter-rater agreement can be examined to determine how well evaluators interpret and implement the rubric associated with the evaluation system.

## APPENDIX A - LETTER REQUESTING PERMISSON FROM SUPERINTENDENTS

Date

Dear Superintendent,

I am currently enrolled as a doctoral student in the Educational Leadership department at The University of Southern Mississippi. I am in the process of completing the dissertation stage of the program. My research focuses on teacher evaluation, specifically the Mississippi Statewide Teacher Appraisal Rubric (M-STAR). The goal of my research is to determine whether or not teachers in the southern part of the state of Mississippi perceive the M-STAR evaluation system as valid and reliable. Also, to determine if teachers feel as though their evaluators provide them with enough useful feedback to improve their instruction.

I am requesting permission to elicit voluntary responses for my study from the teachers in your district. The study is designed to use a quantitative approach consisting of collecting data from a 40 question survey that your teachers will complete by either online link or with a paper pencil survey. It is my intention for this study to benefit administrators by gauging teachers' perceptions of the M-STAR process.

For your convenience, I have enclosed a form letter to be signed and returned granting permission to survey the teachers in your district. If you have questions regarding this study, please contact me directly at (601) XXX-XXXX or email me at [steven.hampton@eagles.usm.edu](mailto:steven.hampton@eagles.usm.edu). You may also contact the chairperson of my committee, Dr. David Lee at The University of Southern Mississippi, at (601) XXX-XXXX or at [david.e.lee@usm.edu](mailto:david.e.lee@usm.edu). Thank you for your time and consideration.

Sincerely,

Steven D. Hampton  
Doctoral Candidate  
The University of Southern Mississippi  
Department of Educational Leadership and School Counseling



## APPENDIX B - CONSENT LETTER

Dear Potential Participant,

My name is Steven Hampton. I am a practicing administrator in an area school district and also in the process of completing my PhD at The University of Southern Mississippi. I would like to request your help in my research study I am conducting as a part of my doctoral dissertation. In this study, I am surveying teachers to measure their perceptions of the Mississippi Statewide Teacher Appraisal Rubric better known as M-STAR. Participating in this study would afford you the opportunity to reflect and provide your opinion on the teacher evaluation system currently being used throughout our state.

The procedures for this study will be as follows: Teacher participants will receive a questionnaire entitled *The Mississippi Statewide Teacher Appraisal Rubric (M-STAR) Teacher Evaluation System Survey* one of two ways: 1) The participant may choose to hard copy provided to them or 2) use the link posted on the hard copy to complete an online version of the survey. If the participant chooses to complete the hard copy of the survey they will return it to the designated person listed on the survey itself.

If you would like to participate, please fill out the attached questionnaire. It should take about 10-15 minutes. Please do not write your name or any information on the questionnaire that could identify you so that all data collected is anonymous. You have the right to not respond to any question that makes you uncomfortable. By reading this consent letter, you agree that you understand the procedures and any risks and benefits involved in this research. You are free to refuse to participate or to withdraw your consent to participate in this research at any time without penalty or prejudice; your participation is entirely voluntary. Any information that is inadvertently obtained during the course of this study will remain completely confidential. The results will be compiled and submitted as a doctoral study. After all the surveys have been turned in at each location, they will be placed in manila envelopes and sealed until the time the data will be examined. The surveys will be shredded and the files will be erased five years after the study has been completed. There are no risks involved by participating in this study.

The University of Southern Mississippi Institutional Review Board, which ensures that research studies involving human subjects follow federal regulations, has approved the research and this consent letter. Questions regarding your rights as a participant in this study should be directed to the chair of the Institutional Review Board, The University of Southern Mississippi, 118 College Drive #5147, Hattiesburg, Mississippi 39406, (601) 266-6820. Mr. Steven D. Hampton a USM Educational Leadership doctoral student will answer any questions regarding the research itself by calling (601) 310-0943. Any new information that develops during the study will be provided to you if the information might affect your willingness to continue participation in the study.

By completing the questionnaire you are acknowledging you have read this consent letter and agree to participate in this study.

Sincerely, Steven D. Hampton

APPENDIX C - SURVEY INSTRUMENT

**Mississippi Statewide Teacher Appraisal Rubric (M-STAR)  
Teacher Evaluation System Survey**

**Section 1**

**Please mark the answer to the following questions:**

1. Please select the grade range that best aligns with the grade you are current teaching.

- Kindergarten – 2<sup>nd</sup> Grade
- 3<sup>rd</sup> – 4<sup>th</sup> Grade
- 5<sup>th</sup> – 6<sup>th</sup> Grade
- 7<sup>th</sup> – 8<sup>th</sup> Grade
- 9<sup>th</sup> – 12<sup>th</sup> Grade

2. How many total years of teacher experience you have?

- 0 - 2 years
- 3 - 6 years
- 7 - 10 years
- 11 - 15 years
- 16 - 20 years
- 21+ years

3. Do you teach in a subject and/or grade level that is tested under the state accountability system?

- yes
- no

4. What subject area best describes your current teaching assignment?

- English Language Arts
- Math
- Social Studies
- Science
- Special Education
- Elective
- Other

5. What is the highest degree you have obtained?

- Bachelors
- Masters

- Specialist
- Ph.D.

6. Are you a National Board Certified Teacher?

- yes
- no

7. The total number of M-STAR observations (formals and informal/walkthroughs) you have been participated in?

- None
- 1 - 2 observations
- 3 - 5 observations
- 6 - 9 observations
- 10+ observations

8. The total time you have been involved in training or professional development on the M-STAR teacher evaluation system?

- None
- 1 - 2 hours
- 2 - 4 hours
- 4 - 6 hours
- 6+ hours

**Section 2: M-STAR Validity**

Select one response per statement

9. The standards used in the M-STAR evaluation system are fair.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

10. Working towards improving my performance on the M-STAR evaluation standards will also help me to improve the quality of my instruction.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

11. The M-STAR evaluation instrument includes clear explanations for each performance level descriptor.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

12. The four M-STAR levels of performance: Unsatisfactory, Emerging, Effective, and Distinguished are adequate.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

13. The M-STAR descriptors focus on the key teacher behaviors that positively impact student learning.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

14. The M-STAR evaluation standards do a good job of defining good teaching.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

15. The M-STAR instrument provides teachers with objective information about their teaching.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

16. The M-STAR instrument incorporates indicators of student learning in the evaluation process.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

**Section 3: M-STAR Reliability**

Select one response per statement

16. I understand the meaning of each descriptor and level of performance used in the M-STAR evaluation instrument.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

17. My evaluators have been adequately trained to consistently evaluate my teaching.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

18. I am confident that evaluators at my school interpret and score teacher evaluations in a similar manner.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

19. I am confident that evaluators from other schools in the district interpret and score teacher evaluations in a manner similar to my school administrators.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

No Opinion or Not Enough Information to Respond

21. The scores from my evaluations have been consistent from one evaluator to another.

Strongly Agree

Agree

Disagree

Strongly Disagree

No Opinion or Not Enough Information to Respond

#### **Section 4: M-STAR Feedback**

Select one response per statement

24. The process used under the M-STAR system fosters a climate for instructional improvement.

Strongly Agree

Agree

Disagree

Strongly Disagree

No Opinion or Not Enough Information to Respond

25. The M-STAR instrument provides teachers with objective information about their teaching.

Strongly Agree

Agree

Disagree

Strongly Disagree

No Opinion or Not Enough Information to Respond

26. The M-STAR system enhances dialogue and mutual understanding between teachers and evaluators about effective teaching.

Strongly Agree

Agree

Disagree

Strongly Disagree

No Opinion or Not Enough Information to Respond

27. The M-STAR system increases teacher reflection on choices of teaching strategies.

Strongly Agree

Agree

Disagree

- Strongly Disagree
- No Opinion or Not Enough Information to Respond

28. I have received regular focused follow-up and instructional support based on my M-STAR evaluations.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

29. The feedback I have received from my M-STAR evaluations has been valuable.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

30. As a result of the feedback I have received from my M-STAR evaluations, I have improved my ability to design high quality lessons.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

31. As a result of the feedback I have received from my M-STAR evaluations, I have improved the quality my overall instruction.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

32. The M-STAR rubrics provide feedback that can help guide individual professional development plans for all teachers.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

No Opinion or Not Enough Information to Respond

33. As a result of the feedback I have received from my M-STAR evaluations, I have increased student learning.

Strongly Agree

Agree

Disagree

Strongly Disagree

No Opinion or Not Enough Information to Respond

**Section 5: M-STAR Perception**

Select one response per statement

35. The M-STAR evaluation process is helpful to my professional growth.

Strongly Agree

Agree

Disagree

Strongly Disagree

No Opinion or Not Enough Information to Respond

36. The most important purpose of the M-STAR evaluation is to fulfill human resource requirements for continued employment.

Strongly Agree

Agree

Disagree

Strongly Disagree

No Opinion or Not Enough Information to Respond

37. The M-STAR evaluation incorporates indicators of student learning in the evaluation process.

Strongly Agree

Agree

Disagree

Strongly Disagree

No Opinion or Not Enough Information to Respond

38. The process used under the M-STAR evaluation system fosters a climate for instructional improvement.

Strongly Agree

Agree



- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

39. All teachers should be evaluated at least twice a year to provide feedback on instructional improvement.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

40. I focus my professional development efforts on activities that directly help me achieve the M-STAR evaluation standards.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- No Opinion or Not Enough Information to Respond

## APPENDIX D – IRB Approval Letter



**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](http://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: 15031602  
PROJECT TITLE: Teachers' Perceptions of the Mississippi Statewide Teacher Appraisal Rubric (M-STAR) Evaluation  
PROJECT TYPE: New Project  
RESEARCHER(S): Steven D. Hampton  
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: 03/18/2015 to 03/17/2016  
**Lawrence A. Hosman, Ph.D.**  
**Institutional Review Board**

## REFERENCES

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Aseltine, J. M., Faryniarz, J. O., & Rigazio-DiGilio, A. J. (2006). *Supervision for learning: A performance-based approach to teacher development and school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991, 12). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*, 61(2), 213-238. doi: 10.3102/00346543061002213
- Barnett, J. A. (2006). *Considering growth in teaching: Teacher perspectives on the Ontario performance appraisal* (Doctoral dissertation, University of Toronto, Canada). Available from ProQuest Dissertations and Theses database. (AAT No. NR21803)
- Beesley, A. D. E., Apthorp, H. S. E., & Mid-Continent Research for Education and Learning (Mcrel). (2010). *Classroom Instruction that Works, Second Edition: Research Report*.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, 5(1), 7-74.

- Brandt, R. (1995). On Restructuring Schools: A Conversation with Fred Newmann. *Educational Leadership*, 53(3), 70-73.
- Brandt, R. (1996). On a New Direction for Teacher Evaluation: A Conversation with Tom McGreal. *Educational Leadership*, 53(6), 30-33.
- Brookhart, S. M. (2012). Preventing Feedback Fizzle. *Educational Leadership*, 70(1), 24-29.
- Burling, K. (2012). Evaluating teachers and principals: Developing fair, valid and reliable systems. Retrieved From <http://educatoreffectiveness.pearsonassessments.com/wp/wp-content/uploads/2012/03/educator-effectiveness-roadmap-03-01-12.pdf>
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81.
- Cangelosi, J. S. (1991). *Evaluating classroom instruction*. New York, NY: Longman.
- Cantrell, S., & Scantlebury, J. (2011). Effective Teaching: What Is It and How Is It Measured? *Voices in Urban Education, Effective Teaching as a Civil Right* (31), 28-35. Retrieved from <http://annenberginstitute.org/sites/default/files/vue/356/VUE31.pdf#page=30>
- Catano, N., & Stronge, J. H. (2006). What are principals expected to do? Congruence between principal evaluation and performance standards. *NASSP Bulletin*, 90(3), 221-237.
- Chappuis, J. (2012). How Am I Doing? *Educational Leadership*, 70(1), 36-40.

- Clemesten, H. E. (2000). *Comparison Of Perceptions Of Elementary Teachers About Teacher Evaluation Procedures In Two Different School Districts* (Doctoral dissertation, Wayne State University, Detroit, MI). Available from ProQuest Dissertation and Theses Database. (UMI No. 9992181).
- Colvin, G. (2001). We can't all be above average. *Fortune*, 144, 3.
- Conley, S., Muncey, D. E., & You, S. (2005). Standards-Based Evaluation and Teacher Career Satisfaction: A Structural Equation Systeming Analysis. *Journal of Personnel Evaluation in Education*, 18(1), 39-65.
- Coutts, J. D. (1999). Turning the Corner on Quality Education Through Monitoring Instruction. *NASSP Bulletin*, 83(609), 110-120.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Holt, Rinehart, and Winston
- Cronbach, L. J. (1971). Test validation. *Educational Measurement*, 2, 443-507.
- Danielson, C. (2001). New trends in teacher evaluation. *Educational Leadership*, 58(5), 12-15.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. (2008). *The handbook for enhancing professional practice: Using the framework for teaching in your school*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. (2010). *Teaching methods*. Upper Saddle River, NJ: Merrill/Pearson.
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.

- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation: To enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (1990). Teacher evaluation in transition: Emerging roles and evolving methods. *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*, 17-32.
- Darling-Hammond, L. (1996). What matters most: Teaching for America's future. New York, NY: *National Commission on Teaching and America's Future*.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Davis, D. R., Pool, J. E., & Mits-Cash, M. (2000). Issues in implementing a new teacher assessment system in a large urban school district: Results of a qualitative field study. *Journal of Personnel Evaluation in Education*, 14(4), 285-306.  
doi:10.1023/A:1011139203370
- Dean, C. B., Hubbell, E. R., Pitler, H., & Stone, B. (2012). *Classroom instruction that works: Research based strategies for increased student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Definitions. (n.d.). Retrieved November 16, 2014, from <http://www.ed.gov/race-top/district-competition/definitions>
- Dewey, J. (1910). *How we think*. Boston, MA: DC Heath.
- Dewey, J (1916). *Democracy and Education: An introduction into the philosophy of education*. New York, NY: The Macmillan Company.

- Donaldson, M.L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Washington, DC: Center for American Progress. Retrieved from <https://www.americanprogress.org/issues/education/report/2009/06/25/6243/so-long-lake-wobegon/>
- Doyle, K. O. (1983). *Evaluating teaching*. Lexington, MA Lexington Books.
- Driscoll, M. P. (2005). *Psychology of learning for instruction*. Pearson; 3rd edition (August 13, 2004)  
[http://ocw.metu.edu.tr/pluginfile.php/9013/mod\\_resource/content/1/driscoll-ch10%20\(1\).pdf](http://ocw.metu.edu.tr/pluginfile.php/9013/mod_resource/content/1/driscoll-ch10%20(1).pdf)
- Duke, D. L., & Stiggins, R. J. (1990). Beyond minimum competency: Evaluation for professional development. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation* (pp. 116-132). Newbury Park, CA: Sage.
- Ebmeier, H. (2003). How supervision influences teacher efficacy and commitment: An investigation of a path system. *Journal of Curriculum & Supervision*, 18(2), 110-141.
- Eisner, E.W. (1975). *The perceptive eye: Toward the reformation of educational evaluation*. Stanford, CA: Stanford Evaluation Consortium.
- Ellett, C. D., & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17(1), 101-128.
- Elmore, R. F., & Fuhrman, S. H. (2001). Holding schools accountable: is it working?. *Phi Delta Kappan*, 67-72.

- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3), 363.
- Frase, R. S. (1992). Implementing commission-based sentencing guidelines: The lessons of the first ten years in Minnesota. *Cornell Journal of Law and Public Policy*, 2, 279.
- Frase, L. E., & Streshly, W. (1994). Lack of accuracy, feedback, and commitment in teacher evaluation. *Journal of Personnel Evaluation in Education*, 8(1), 47-57.
- Frick, T., & Semmel, M. I. (1978). Observer agreement and reliabilities of classroom observational measures. *Review of Educational Research*, 157-184.
- Garth-Young, B. (2007). *Teacher evaluation: Is it an effective practice? A survey of all junior high/middle school principals in the State of Illinois*. (Order No. 3314402, Loyola University Chicago). ProQuest Dissertations and Theses, 159. Retrieved from <http://search.proquest.com/docview/304848276?accountid=13946>. (304848276).
- Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2001). *Supervision and instructional leadership: A developmental approach* (6<sup>th</sup> ed.). Needham Heights, MA: Allyn & Bacon/Longman Publishing.
- Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, 2(1), 50-55.
- Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. *Online Submission*. Retrieved from [http://cecr.ed.gov/pdfs/Inter\\_Rater.pdf](http://cecr.ed.gov/pdfs/Inter_Rater.pdf)



- Haefele, D.L. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education*, 7, 21-31.
- Hanushek, E. A. (1992, Feb.). The trade-off between child quantity and quality. *The Journal of Political Economy*, 100(1), pp. 84-117.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hattie, J. (2012). Know Thy Impact. *Educational Leadership*, 70(1), 18-23.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Haycock, K. (1998). *Good teaching matters: How well-qualified teachers can close the gap*. Washington, DC: ERIC Clearinghouse.
- Hazi, H. M., & Rucinski, D. A. (2009). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *Education Policy Analysis Archives*, 17(5), 1-22.
- Heneman III, H. G., Milanowski, A., & Kimball, S. (2007). Teacher Performance Pay: Synthesis of Plans, Research, and Guidelines for Practice. Policy Brief RB-46. *Consortium for Policy Research in Education*.
- Herlihy, C., Karger, E., Pollard, C., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record*, 116(1), 1-28.
- Hunter, M. C. (1994). *Enhancing teaching*. New York, NY: Macmillan College Publishing Company.

- Hunter, R. (2004). *Madeline Hunter's mastery teaching: Increasing instructional effectiveness in elementary and secondary schools*. Thousand Oaks, CA: SAGE.
- Ingvarson, L., & Chadbourne, R. (1997). Reforming teachers' pay systems: The advanced skills teacher in Australia. *Journal of Personnel Evaluation in Education*, 11(1), 7-30.
- Interstate New Teacher Support and Assessment Consortium. (1992). *System standards for beginning teacher licensing and development: A resource for state dialogue*. Washington, DC: Council of Chief State School Officers.
- Johnson, B. L. (1997). An Organizational Analysis of Multiple Perspectives of Effective Teaching: Implications for Teacher Evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 69-87.
- Kahu, E. (2008). Feedback: The heart of good pedagogy. *New Zealand Annual Review of Education*, 17, 187-197.
- Kane, M. T. (2006). Validation. *Educational measurement*, 4, 17-64.
- Kaplan, L. S., & Owings, W. A. (2001). Teacher Quality and Student Achievement: Recommendations for Principals. *NASSP Bulletin*, 85(628), 64-73.
- Kersten, T. A., & Israel, M. S. (2005). Teacher Evaluation: Principals' Insights and Suggestions for Improvement. *Planning and Changing*, 36, 47-67.
- Kimball, S. M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16(4), 241-268.

- Koppich, J., & Showalter, C. (2005). *Strategic management of human capital: a cross-case analysis of five districts*. Madison, WI: Strategic Management of Human Capital.
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Lambert, L., Walker, D., Zimmerman, D., Cooper, J., Lambert, M., Gardner, M., & Szabo, M. (2002). *The constructivist leader*. New York, NY: Teachers College Press and National Staff Development Council.
- Levin, B. (1979). Teacher Evaluation - A Review of Research. *Educational Leadership*, 37(3), 240-45.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3-13.
- Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science*, 15(5), 265-268.
- Machell, J. (1995). The teacher evaluation environment: An examination of attributes related to teacher growth. *Journal of Personnel Evaluation in Education*, 9(3), 259-273.
- Manning, R. C. (1988). *The teacher evaluation handbook: Step-by-step techniques and forms for improving instruction*. Englewood Cliffs, NJ: Prentice Hall.
- Marshall, K. (2005). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan*, 86(10), 727-735.
- Marx, G. (2007). *Research brief-teacher evaluation*. Retrieved August, 8, 2009.  
<http://www.principalspartnership.com/Teacherevaluation.pdf>

- Marzano, R. J. (2012). The Two Purposes of Teacher Evaluation. *Educational Leadership, 70*(3), 14-19.
- Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J., & Toth, M. (2013). *Teacher evaluation that makes a difference: A new system for teacher growth and student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mayer, D. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), 29-45.
- McGaghie, W. C. (1991). Professional competence evaluation. *Educational Researcher, 20*(1), 3-9.
- Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher, 10*(9), 9-20.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist, 50*(9), 741.
- Mielke, P., & Frontier, T. (2012). Keeping Improvement in Mind. *Educational Leadership, 70*(2), 10-13.

- Milanowski, A. T., & Heneman III, H. G. (2001). Assessment of Teacher Reactions to a Standards-Based Teacher Evaluation System: A Pilot Study. *Journal of Personnel Evaluation in Education, 15*(3), 193-212.
- Mississippi Department of Education. (2012). *Mississippi ESEA Flexibility Request* (February 2012). Retrieved from <http://www.mde.k12.ms.us/docs/federal-programs-fy13-cfpa/attachments.pdf?sfvrsn=0>
- Mississippi Department of Education. (2014). *Mississippi Statewide Teacher Evaluation Rubric Process Manual* (July 2014). Retrieved from <http://www.mde.k12.ms.us/docs/teacher-center/revised-m-star-process-manual-july-2014.pdf?sfvrsn=2>
- Natriello, G. (1990). Intended and unintended consequences: Purposes and effects of teacher evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers the new handbook of teacher evaluation* (pp. 35-45). Newbury Park, CA: Sage.
- Odden, A. (2004). Lessons learned about standards-based teacher evaluation systems. *Peabody Journal of Education, 79*(4), 126-137.
- Odden, A., & Wallace, M. (2008). *How to create world class teacher compensation*. Saint Paul, MN: Freeload Press, Inc.
- Orey, M. (2007, April 22). Fear of firing, How the threat of litigation is making companies skittish about axing problem workers. *Business Week, (4031)*, 52.
- Partee, G. L. (2012). Using Multiple Evaluation Measures to Improve Teacher Effectiveness: State Strategies from Round 2 of No Child Left Behind Act

- Waivers. *Center for American Progress*. Retrieved from <http://files.eric.ed.gov/fulltext/ED539743.pdf>
- Peterson, C. H. (1982). *A century's growth in teacher evaluation in the United States*. New York, NY: Vantage Press.
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin Press.
- Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International Universities Press.
- Popham, W. J. (2013). On Serving Two Masters-Formative and Summative Teacher Evaluation. *Principal Leadership, 13*(7), 18-22.
- Pritchett, J., Sparks, T. J., & Taylor-Johnson, T. (2010). *An analysis of teacher quality, evaluation, professional development and tenure as it relates to student achievement*. (Doctoral dissertation, Saint Louis University, Saint Louis, MO). Available from ProQuest Dissertation and Theses Database. (UMI No. 821590306).
- Range, B. G., Young, S., & Hvidston, D. (2013). Teacher perceptions about observation conferences: What do teachers' think about their formative supervision in one US school district? *School Leadership and Management, 33*(1), 61-77.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.
- Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 635-674.

- Ryans, D. G. (1960). *Characteristics of teachers, their description, comparison, and appraisal: a research study*. Washington, DC: American Council on Education.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science, 18*(2), 119-144.
- Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247-256.
- Sanders, W.L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of personnel evaluation in education, 11*(1), 57-67.
- Schmoker, M. (2006). *Results Now: How We Can Achieve Unprecedented Improvements in Teaching and Learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Shinkfield, A. J., & Stufflebeam, D. L. (1995). *Teacher Evaluation: Guide to Effective Practice*. Norwell, MA: Kluwer Academic Publishers.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-23.
- Stake, R. E. (1967). *Toward a Technology for the Evaluation of Educational Programs*. Chicago, IL: Rand McNally & Company.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 271-286.

- Stronge, J. H. (1997). *Evaluating teaching: A guide to current thinking and best practice*. Thousand Oaks, CA: Corwin Press.
- Stronge, J. H. (2006). *Evaluating teaching: A guide to current thinking and best practice* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Corwin Press
- Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. Larchmont, NY: Eye on Education.
- Sutton, R., Douglas, K., & Hornsey, M. J. (2012). *Feedback: The communication of praise, criticism, and advice*. New York, NY: Peter Lang.
- Tinsley, H. E. A., & Weiss, D. J. (2000). Interrater reliability and agreement. In H. E. A. Tinsley & S. D. Brown (Eds.). *Handbook of applied multivariate statistics and mathematical systeming* (pp. 95–124). New York, NY: Academic Press.
- Toch, T., & Rothman, R. (2008). *Rush to Judgment: Teacher Evaluation in Public Education*. Washington, DC: Education Sector.
- Tovani, C. (2012). Feedback Is a Two-Way Street. *Educational Leadership*, 70(1), 48-51.
- Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tugend, A. (2013, April 5). You've Been Doing a Fantastic Job. Just One Thing.... *The New York Times*. Retrieved from [http://www.nytimes.com/2013/04/06/your-money/how-to-give-effective-feedback-both-positive-and-negative.html?pagewanted=all&\\_r=1&](http://www.nytimes.com/2013/04/06/your-money/how-to-give-effective-feedback-both-positive-and-negative.html?pagewanted=all&_r=1&)
- U. S. Department of Education. *New No Child Left Behind Flexibility*. (March 2004). Retrieved from <http://www2.ed.gov/nclb/methods/teachers/hqtflexibility.pdf>



- U. S. Department of Education. (2010). *A blueprint for reform: The reauthorization of the Elementary and Secondary Education Act*. U.S. Department of Education, Office of Planning, Evaluation and Policy Development.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: our national failure to acknowledge and act on teacher differences*. New York, NY: The New Teacher Project.
- Wheeler, P. H., & Scriven, M. (1997). Building the foundation: Teacher roles and responsibilities. *Evaluating teaching*, 27-58.
- Wiener, R. & Lundy, K. (2013). Want to Build a Better Teacher Evaluation. *Education Week*, 32(30), 26-27.
- Wiggins, G. (2012). Seven Keys to Effective Feedback. *Educational Leadership*, 70(1), 10-16.
- William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- William, D. (2012). Feedback: Part of a System. *Educational Leadership*, 70(1), 30-34.
- Witziers, B., Bosker, R. J., & Kruger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 398-425.
- Wuhs, S. K., & Manatt, R. P. (1983). The Pace of Mandated Teacher Evaluation Picks Up. *American School Board Journal*, 170(5), 28.