Spring 2-2017

# The Relationship between Teacher Evaluation Ratings and Student Achievement in a Rural, Midwest School District

Margaret A. Mathus
*Lindenwood University*

Follow this and additional works at: https://digitalcommons.lindenwood.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons

The Relationship between Teacher Evaluation Ratings and Student Achievement in a

Rural, Midwest School District

Margaret A. Mathus

A Dissertation submitted to the Education Faculty of Lindenwood University

in partial fulfillment of the requirements for the

degree of

Doctor of Education

School of Education

The Relationship between Teacher Evaluation Ratings and Student Achievement in a

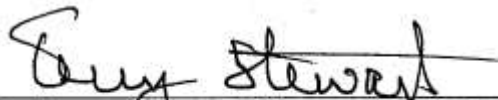Rural, Midwest School District

by

Margaret A. Mathus

This dissertation has been approved in partial fulfillment of the requirements for the

degree of

Doctor of Education

at Lindenwood University by the School of Education

Dr. Terry Stewart, Dissertation Chair      Date   2/17/2017

Dr. Lynda Leavitt, Committee Member      Date   2/17/17

Dr. Kevin Winslow, Committee Member      Date   2/17/17

Declaration of Originality

I do hereby declare and attest to the fact that this is an original study based solely upon my own scholarly work here at Lindenwood University and that I have not submitted it for any other college or university course or degree here or elsewhere.

Full Legal Name: Margaret A. Mathus

Signature: _Margaret A. Mathus_    Date: _2/17/2017_

**Abstract**

While many factors have been identified as influencing student academic performance, previous studies consistently determined effective teaching as the most significant factor, within the control of educators, leading to improved student achievement. Nonetheless, educational experts, statisticians, and policy-makers alike acknowledged the complexity of isolating the contributions of individual teachers on their students' achievement. Converging with these changing beliefs about teaching and learning, the landscape of education faced an additional challenge — marked by an increased demand for schools and individual teachers to be held accountable for the academic growth of his/her students. Local districts have been empowered to create and implement teacher evaluation systems, with the caveat they maintain student achievement data as one measure of teacher effectiveness.

While there has been research conducted investigating a relationship between performance-based teacher evaluation systems and student achievement, studies have been limited to the most common large-scale models. This study was unique because the research focused on a specific teacher evaluation system, created by and for, a rural Missouri school district during its first two years of implementation. The purpose of this mixed-methods research study was two-fold: (1) to investigate the relationship between teachers' annual evaluation ratings (as measured by the researched district's teacher evaluation tool) and their students' academic performance (as measured by the MAP and i-Ready assessments), and (2) to analyze teacher and administrator perceptions of the impact of the new teacher evaluation system on improving student achievement and the teachers' instructional performance.

This study's analysis took both math and reading achievement scores into account, considering two different standardized assessments, the state-mandated Missouri Assessment Program (MAP) and a locally-administered i-Ready Benchmark Assessment. The student achievement data yielded an increase in student achievement over the two years of the study.  However, the results of the study did not establish a correlation between the two variables: teacher quality and student achievement.  More sensitive evaluation methods are needed to isolate the variable of teacher evaluation ratings on student achievement.

**Table of Contents**

**Tables**

**Chapter One: Introduction**

Highly publicized criticism on the U.S. public education system and the ineffectiveness of the teacher evaluation process previous to this research led to increased attention on teacher and school accountability (Center for Public Education [CPE], 2012; Guggenheim et al., 2011; Kane, Taylor, Tyler, & Wooten, 2011a; Klein, 2011; U.S. National Commission on Excellence in Education, 1983; Weber, 2010; Weisberg, Sexton, Mulhern, & Keeling, 2009; Zhao, 2009). A recurring theme among the literature was that the greatest impact on student achievement and learning was teacher effectiveness (Danielson, 2001; Hattie, 2012; Grant, Hindman, & Stronge, 2013; Schmoker, 2011; Stronge, 2013; Stronge, Ward, & Grant, 2011; Tucker & Stronge, 2005). Due to the growing concern over the United States' students' ability to compete internationally, a shift began in the educational arena toward establishing rigorous accountability standards for teachers and schools (Galey, 2015; National Council on Teacher Quality [NCTQ], 2013). The influence of these reform efforts on educator accountability was as far-reaching as teacher-preparation programs and accountability measures in higher education institutions (Eaton, 2010).

A review of the literature yielded a lack of consensus on a uniform definition of teacher effectiveness. For the purpose of this study, teacher effectiveness was defined as "the impact that classroom factors, such as teaching methods, teacher expectations, classroom organization, and use of classroom resources, have on students' performance" (Campbell, Kyriakides, Muijs, & Robinson, 2012, p. 3). A teacher's impact on student learning, and the measurement thereof through evaluations, became a controversial and high profile topic in education reform. Many teachers criticized the growing trend

toward measurement which tied student achievement scores with teacher evaluation ratings (Baker et al., 2010).  A review of literature revealed the criticism stemmed primarily from the following assumptions: (1) student learning was influenced by the teacher alone, (2) tests used for assessment were an accurate measure of student learning, and (3) measuring teacher effectiveness based on student assessment data led to improved student achievement (Baker et al., 2010).  Opponents of test-based accountability also cited several unintended consequences resulting from the over-reliance of student assessment data in teacher evaluation: teaching to the test, teachers avoiding the more difficult students, increased classroom time spent on test-preparation at the expense of time spent on quality instruction, and ignoring non-tested curriculum/subjects (Linn, 2011; Schul, 2011).

This heightened awareness resulted in policymakers and the public seeking increased accountability from educators by implementing stricter teacher evaluation systems that included student achievement and growth in evaluation ratings (Center for American Progress & The Council of Chief State School Officers [CCSSO], 2014; National Governors Association, 2015).  The emphasis on evaluating teacher effectiveness quickly evolved into seeking a direct relationship between a teacher's instructional performance and student outcomes, as measured by student performance on standardized assessments (National Comprehensive Center for Teacher Quality, 2012). Educational researcher, Zhao (2009) wrote the following,

> The defining characteristics of education reform efforts in the United States
>
> during the early years of the 21st century: (1) excellence equals good test scores

in math and reading, and (2) standards- and test-based accountability is the tool to

achieve such excellence. (p. 2).

A significant volume of studies established the then-current processes for

evaluating teachers were based on minimal observations by the principal, and were

ineffective for identifying successful teachers or for measuring a teacher's impact on

student learning and achievement (Darling-Hammond, Amrein-Beardsley, Haertel, &

Rothstein, 2013; Kane et al, 2011a; Marzano, 2012, The New Teacher Project, 2010;

Toch & Rothman, 2008).   For the purpose of this study the term evaluator refers to

building administrators, who were the district staff responsible for monitoring and

evaluating teaching staff.   Youngs (2013) wrote, "In recent years concerns with

shortcomings in traditional teacher evaluation systems have led district, state, and federal

policymakers to focus on a number of new approaches to evaluating teachers" (p. 2).

Recent federal initiatives, Race to the Top (RttT) grants (2009), and No Child Left

Behind (NCLB) Act (2001) waivers provided incentives for states/districts to implement

rigorous teacher evaluation programs in which student growth played a significant factor

in rating teacher effectiveness and for informing personnel decisions (McNeil, 2013;

Popham, 2013; USDOE, 2001, 2009).

The most recent education policy at the time of this writing, Every Student

Succeeds Act (ESSA), signed into effect in 2015, updated prior federal educational laws

that originated in 1965 with the passing of the Elementary and Secondary Education Act

(The Education Trust, 2016).  This act signaled a significant change in the tone from

previous federal education policy by placing prohibitions on federal influence or

supervision of state implementation of academic standards or accountability systems

(National Conference of State Legislators [NCSL], 2015).

While Every Student Succeeds Act (ESSA, 2015) may have removed some of the

pressures on school districts to implement federal one-size-fits-all solutions for school

improvement and reduced the emphasis on some of the burdensome [standardized]

assessments, the ruling still held to the expectation of higher academic standards in every

state and maintained the directive of using student data as accountability measures at both

the state and local levels (NCTQ, 2015).  ESSA also called for a termination of waivers

under Section 9401 of NCLB by August 1, 2016 (NCSL, 2015), and required the

transition to then-new state plans to "begin in the 2016-2017 school year, with full

implementation occurring in the 2017-2018 school year" (NCSL, 2015, p. 1).

Additionally, ESSA described the provision for state grants to encourage state and local

audits of teacher evaluation and assessment systems (NCSL, 2015).

A brief review of the history of teacher evaluations, and an overview of the

context in which the national movement toward reformed teacher evaluation transpired

are provided in Chapter One.  Second, the problem statement and theoretical research that

guided this study are explained in reference to the research questions.  Finally, the key

terms were defined, the underlying assumptions presented, and the limitations associated

with this research discussed.

**Background of Teacher Evaluation Reform**

Researchers established that evaluation had long been a part of the educational

system.  According to Marzano, Frontier, & Livingston (2011), as early as the 18th

century, teacher supervision began to focus on the improvement of instruction.  However,

the fundamental ideals and purposes related to the why and how of evaluations changed

dramatically through the years.  Danielson and McGreal (2000) explained how education

was "built around a conception of practice based on current and emerging findings; as

those findings suggest new approaches, pedagogical practices must also move forward"

(p. 3).

Research into educational reforms of the 20th century was largely traced to a

seminal text entitled, *A Nation at Risk*:  *The Imperative for Educational Reform.*  This

report*,* published in 1983 by President Regan's U.S. National Commission on Excellence

in Education (1983), brought to public attention the finding of the United States'

staggering decline on national and international tests.  The authors reported this decline

was the result of a failing education system, and warned an eminent economic crisis

threatened the nation's security.  "The educational foundations of our society are

presently being eroded by a rising tide of mediocrity that threatens our very future as a

Nation and a people" (A Nation at Risk, 1983, para. 1).  The commission went on to

name several specific limitations of the American education system, and made a number

of recommendations for educational reform, based on the research findings.  Among the

suggestions was a call for more rigorous standards, a better trained teaching force, and

more time spent in school (A Nation at Risk, 1983).  Educational researchers concurred,

this report served as a landmark for catapulting education onto the national agenda and

became the catalyst for standards-based, testing-focused reform at the federal, state, and

local levels for the next three decades (Ansary, 2007).

At the state level, Missouri Senate Bill 291 (Missouri Senate, 2009) directed

school districts to adopt then-new teaching standards, which featured particular

components.  The Missouri Department of Elementary and Secondary Education

(MODESE) subsequently created a Model Educator Evaluation System (MMEES),

aligned to those standards (Missouri Department of Elementary and Secondary Education

[MODESE], 2011).  The primary goal of the then-new system, adopted by the State

Board of Education in June of 2012, was to "promote growth in effective practice that

ultimately increases student performance" (MODESE, 2013, p. 4).  Several states passed

legislation requiring student achievement/growth data accounting for a substantial weight

of teacher evaluation ratings (NCTQ, 2015).  Regardless of the reason, "In most states,

teacher performance [would] now be judged for its impact on student learning alongside

traditional measures" (Hull, 2013b, para. 2).

The state of Missouri was one of a number of states that provided local school

districts discretion in designing teacher evaluation systems (Pennington, 2014).  Missouri

also left the decision to each district of how much weight to apply to student growth, and

which assessment would be used to determine student growth (MODESE, 2011).

**Statement of the Problem**

In response to state requirements for increased standards and accountability

measures, many school districts implemented revised educator evaluation systems

aligned to the then-new state guidelines for teaching standards.  No standardized form

was required, allowing districts to develop their own — or find appropriate forms used or

created by other institutions or vendors (CCSSO, 2016).  However, research suggested

prior to taking on the task of creating teacher evaluations that it was essential that specific

performance elements be measured and clearly communicated (National Academy of

Engineering, Steering Committee for Evaluating Instructional Scholarship in Engineering [NAESC], 2009).

The researched district began the process of overhauling teacher evaluation systems in 2013. The new district-created evaluation system aligned with the seven principles of evaluation required by Missouri's Teacher Evaluation Model (MODESE, 2013). A district committee was formed, which consisted of teacher representatives from each grade level and content area, building principals, special education teachers, union representatives, and district administrators. It took the team two years of collaboration to accomplish the following missions — clarifying the district's over-arching goals for the then-new teacher evaluation system, establishing common definitions for teacher effectiveness, developing and adopting common standards for professional practice, creating a standard process for goal setting, reflection, and evaluation of teaching staff with varying levels of experience and expertise, developing a rubric for implementing fair evaluations using multiple rating categories to differentiate between effectiveness levels, and establishing methods and acceptable assessments for all staff to demonstrate and measure individual student growth. The evaluation system was implemented in the district in two phases. The first phase was a pilot year, in which student assessment data was not accounted for in a teacher's final evaluation rating. The second phase of implementation was initiated the following school year and included student assessment data as a contributing factor in final teacher evaluation ratings.

For the purpose of this research, effective teaching was defined as "a teacher whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth" (Race to the Top [RttT], 2010, p. 12). Effective teaching had

many components.  When determining how most accurately, fairly, and reliably to

evaluate teaching with any measurement tool it was important to look at the

psychometrics (NAESC, 2009).  The National Academies Press (2009) wrote the

following:

> It becomes clear that the evaluation of teaching cannot be accomplished by using
>
> a single measurement tool or by basing it on the judgment of one administrator or
>
> peer committee who have made a few classroom visits.  A more accurate and
>
> more valid assessment of teaching performance of necessity involves gathering
>
> information on all five dimensions of teaching performance. (as cited in NAESC,
>
> 2009, p. 23)

Instead, the report suggested, "The key to an effective evaluation of teaching is putting

the parts of this mosaic together in a way that accurately reflects the instructor's overall

teaching competence" (NAESC, 2009, p. 24).

A substantial body of research explored the best method for determining how to

accurately evaluate a teacher's effectiveness as a reflection of his/her impact on student

achievement.  Several researchers suggested measuring a teacher's performance using

statistical models; experts did not agree.  The most widely-used models were Value-

Added Measures and Student Growth Percentiles (SGPs) (Hull, 2013a).

Value-Added Models (VAMs) were based upon the belief that if student learning

was the ultimate objective of teachers, then the evaluation of teachers' effectiveness

should be centered on student learning and a key component of teacher evaluation

systems (Ritter & Shuls, 2012).  According to Benedict, Thomas, Kimerling, and Leko

(2013), VAMs attempted to measure individual teachers' influence on student learning over a year by measuring the students' learning gains on standardized tests.

Other researchers suggested the use of SGPs as the best measure for evaluating the effectiveness of teachers (Betebenner, 2009; 2011).  SGPs were used to explain a student's growth "by examining his/her current achievement relative to [the academic growth of] his/her academic peers—those students beginning at the same place" (Betebenner, 2011, p. 3).  SGPs were then "aggregated using either the median or the mean to report *average* growth achieved at the teacher level" (Diaz-Bilello & Briggs, 2014).  Unlike VAMs, SGP "calculations do not try to adjust for differences in student characteristics" (RAND Education, 2012b, para. 4).

Both models were employed to measure student gains on standardized assessments, despite the documented research establishing the instability of both teacher and school-level averages of student assessment score gains over time (Baker et al., 2010; Darling-Hammond et al., 2013; McCaffrey, Sass, Lockwood, & Mihaly, 2009). Regardless of the statistical model used, most researchers concurred the use of student achievement scores should be only one of multiple measures used to rate teachers, alongside more traditional methods like principal evaluations (Baker et al., 2010).

Psychometrics were much more difficult to establish with a locally developed system versus the large-scale, pre-packaged programs available on the market.  Braun, Chudowsky, and Koenig (2010) highlighted how educational researchers considered alternatives to measure student achievement to assist the states with the requirements of NCLB.  Subsequent research also revealed educators' improved practice was substantial when evaluation systems provided specific feedback, professional development, and self-

reflection (Darling-Hammond, 2012; Kane et al., 2011a).  Therefore, other approaches to evaluating teachers' effectiveness were crafted that were less reliant upon student assessment scores.  These methods used "systematic observation protocols with well-developed, research-based criteria to examine teaching" (Baker et al., 2010).  The recommended teacher evaluation systems typically incorporated a variety of evaluation procedures, including observations of classroom practice, student surveys, and artifacts, such as lesson plans, assignments, and samples of student work.  Baker et al. (2010) claimed these approaches were found to "improve teachers' practice while identifying differences in teachers' effectiveness" (p. 4).

As Missouri moved closer to implementing student growth measures to gauge teacher quality and effectiveness, the challenge many districts faced was how to integrate the research related to teachers' effects on student learning and achievement with other relevant research (e.g. school effects, leadership, organizational change) to inform, rather than distort, the effectiveness of teachers.  Raudenbush (2015) also proposed educational leaders "investigate the implications for policy and practice" (p. 138).

By the academic year 2016-2017, MODESE required teacher evaluations/ratings be based, at least in part, on some form of student growth measure.  In 2011, Missouri was awarded a flexibility waiver based on stipulations of Elementary and Secondary Education Act [ESEA], also known as the NCLB Act (2001).   With Missouri's waiver approval serving as the catalyst for change, the researched district worked rapidly to find a tool to demonstrate student growth at the teacher, school, and district level.  The district opted to create its own evaluation tool, guided by the Missouri Model Evaluation Tool that incorporated assessments determined by the teacher and building administrator as the

measure used for demonstrating the growth of a teacher's students.  The researched

district chose this model due to its availability, its potential for modification to

incorporate the district's specific goals and needs, and because it was a fraction of the

cost of other pre-packaged programs available on the market.

Since it was the responsibility of each individual school district to determine or

create the evaluation system for its staff, there were then-currently no studies on this

teacher evaluation model, as it is in its initial implementation phase.  While research on

identifying teacher effectiveness was relatively new, at the time of this study, the then-

current literature suggested states and local education leaders learn from each other by

examining the different approaches taken to refine and improve their own systems (CPE,

2013).

In light of the then-current educational changes and increased accountability on

teachers for their own students' achievement, this study sought to describe the influence

of a specific Missouri school district's then-new teacher evaluation model on the

district's staff and student performance.

**Rationale for the Study**

This study developed from observations during the researcher's role as a teacher.

The teacher evaluation system formerly used in the district was based upon minimal

principal observations during a school year.  The evaluation system did not require

teacher reflection, provide specific feedback on teacher effectiveness, encourage

professional growth, or lead to the dismissal or coaching of ineffective teaching staff.

This research focused on the potential relationship between the implementation of the

then-new teacher evaluation system and student achievement for measuring teacher

effectiveness.  The research explored differences over the first two years of implementation of the teacher evaluation system, specifically the growth in teacher effectiveness and growth in student achievement.

This study investigated the relationship between teachers' professional growth, measured by the district's teacher evaluation system including (PGPs and student academic performance at the elementary level.  This study also explored teacher and administrator perceptions of the district's teacher evaluation system and the students' academic progress as measured by Missouri Assessment Program (MAP) and i-Ready assessments.

This study added to the already existing body of knowledge by documenting the results of this newly implemented evaluation system in six Midwest elementary schools. This study built on the growing body of research on measuring the effectiveness of educator evaluation systems to improve the quality of teaching and to improve students' academic achievement and learning.  "We have an enormous opportunity to translate this research into classroom practice using a robust framework that illustrates the major components of effective teaching and how they interact within the classroom to positively impact student learning" (Schooling, Toth, & Marzano, 2010, p. 4).  This study was unique, due to the comparison of student achievement and teacher professional growth to implementation of the teacher evaluation system, through its first two years of implementation.

**Research Question**

The research question that guided the work of this dissertation was:

What are the teachers' and administrators' perceptions of the researched district's new teacher evaluation process (based on the MO Teacher Evaluation Model) as a method for improving professional practice (as measured by the Teacher Evaluation Rating tool) and influencing student achievement (as measured by the MAP and i-Ready assessments)?

This question addressed whether teachers and administrators buy-in and perceived value/effectiveness of the evaluation process and tools on their practice/ professional growth and on student outcomes.

**Research Question and Hypotheses**

This mixed methods study was guided by the following question and hypotheses:

$Q_1$: How do teachers and evaluating administrators perceive the district's teacher evaluation as having a potentially positive effect on improved teacher practice and professional growth?

$H_1$: Student achievement as measured on the MAP and i-Ready assessments will improve after implementation of the new teacher evaluation system.

$H_2$: There is a positive correlation between teacher performance ratings and student achievement on the MAP and i-Ready assessments.

**Limitations of the Study**

The following limitations were identified in this study:

This study utilized a relatively small sample.  A convenience sample was selected for this research study, due to accessibility.  The population was limited to the district's faculty who taught the MAP-tested subject areas of English language arts and mathematics in grades three, four, and five, students in grades three, four, and five

assessed by the MAP test, and administrators (evaluators) of teachers in the MAP tested

grades, three, four, and five.  Teacher participants were limited to elementary general

education teachers of  communication arts and mathematics content areas, because these

were the only two areas assessed on the state standardized assessments (Missouri

Assessment Program [MAP] and the revised Standards Based Assessments).   Of all of

the elementary teachers working in the researched district's six elementary schools, only

63 teachers taught in the MAP-tested subjects of mathematics and English language arts

and were included in the study.

Like any observational instrument, the researched district's evaluation tool was

susceptible to observer bias and error, even though training was implemented for all staff

on the new teacher evaluation system.   To minimize data collector bias, summative

evaluations were completed by the same administrative evaluator.  Additionally, the

qualitative survey tool used posed a limitation because participation was voluntary.  The

validity of the study was in part limited to the reliability of the teacher and administrator

survey questions and the interviewer's ability to evaluate survey responses.

The frequent change in student assessments could have an impact on the

reliability of the data comparing teacher ratings and student achievement data.  The state

assessments had undergone a drastic change.  The year 2015 marked the first year of the

Smarter Balanced assessment, after many years of the MAP assessment.  This study was

limited to only one year of data on the Smarter Balanced test, which varied greatly from

the former assessment.  MODESE's new testing vendor, Data Recognition Corporation

(DRC), was responsible for administering, scoring, and reporting the data from the 2016

MAP assessments (MODESE, 2016).

This study was situated within the unique environment of the researched district and within its specific teacher evaluation context.  The program experiences of participants may or may not have been similar to other faculty development program participants at other districts or in other taught areas within the district.  The results of this research may not be generalizable to other programs, institutions, or to the field of faculty development.  However, this study may have the potential to suggest recommendations for best practices in the field of educator evaluation.

These limitations were not all encompassing, but represented the types of life experiences that could have affected the professional careers of faculty members.  This mixed-methods study served dual focuses:  First, to examine the perceptions and personal experiences of the participants and a possible relationship between professional growth and the district's implementation of a new teacher evaluation tool through qualitative data analysis.  Second, the study focused on finding a relationship between teacher ratings using the district evaluation tool and student achievement on the MAP test through statistical data analysis.

**Definition of Terms**

**Accountability systems.**  According to the Education Trust (2016), "accountability systems are the set of policies and practices that a state uses to measure how schools are performing for students, reward those that are serving all of their students well, and prompt involvement in those that are not" (p. 2).

**Attainment-based assessments.**  According to Daley and Kim (2010) attainment-based assessment "focuses on the student's academic attainment or status as of the end of the year," (p. 13) and stands in contrast to value-added assessment.

**Effective teacher.** For the purpose of this research, effective teaching was defined as "a teacher whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth" (RttT, 2010, p. 12).

**Evaluator.** For the purpose of this study the term evaluator refers to building administrators, who were the district staff responsible for monitoring and evaluating teaching staff.

**Growth assessment models.** The "improvement realized by students rather than against a fixed target, acknowledging that all children start out at the same academic level at the beginning of the school year and thus won't all hit the same target" (Jorgenson, 2012, p. 13).

**Local Education Agency.** A public authority established within a state as the administrative agency for its public elementary schools or secondary schools (ESEA, 2012).

**Portfolio-based evaluation.** Documented evidence of teaching from a variety of sources to demonstrate both student learning and teacher growth. According to Nelson (2012), portfolio-based evaluations "are becoming a new option to make the evaluation process more time efficient, productive, and a process which is meaningful, comprehensive, and accurate" (p. 11).

**Practice standards.** Observable behaviors and actions required to meet performance standards, which are measurable and can be used as guides to establish individual performance goals, professional development plans, and evaluation conferences within a system of continuous improvement focused on expert performance (CCSSO, 2008, p. 20).

**State education agencies.** The governing agency (U.S. Department of Education [USDOE]) within each state responsible for providing information, resources, and technical assistance on educational matters to schools and its residents (as cited in The Aspen Institute, 2015).

**Student growth.**  According to the NCLB Act (2001), a change in academic achievement across two or more points in time.

**Student growth percentiles.**  According to Betebenner (2011) "describes how (a) typical student's growth is by examining his/her current achievement relative to his/her academic peers—those students beginning at the same place" and . . . "who have, in the past, 'have walked the same achievement path" (p.  3).

**Student learning objectives.**  According to the Reform Support Network (2013) student learning objectives are content and grade/course specific measurable learning goals that may be used as "one measure to assess teachers' contributions to student growth in educator evaluation systems" and are particularly useful for measuring student learning in non-tested subjects and grades (p. 1).  Additionally, Lacireno-Paquet, et al. (2014), explained student-learning objectives are one way to measure a teacher's impact on student growth as an **"**alternative to the more generally used value-added modeling with standardized test scores" (p. 1).

**Teacher evaluation ratings.** A numerical rating that determines a teacher's growth in practice on an identified area of focus between the baseline at the beginning of the year and score to the follow-up, end of the year score (MODESE, 2013).

**Value-added modeling.**  Similar to the Growth Model, according to Benedict, et al. (2013) "an attempt to evaluate teachers' contributions on student learning in a given

year by measuring students' learning gains on standardized tests" (p. 66). "This involves matching each student's test scores to his or her own previous scores, measuring the student's academic growth as the change in attainment from the beginning of the year to the end of year" (Daley & Kim, 2010, p. 13). "An effort to estimate the value that each individual teacher (or school) adds to the learning of his or her students during a given time period" (Ritter & Shuls, 2012, p. 36).

**Summary**

With increased accountability through linking an educator influence on a students' academic growth, educational systems rushed to implement system overhauls and rigorous teacher performance evaluations expected to result in positive academic student outcomes as powerful indicators of educator growth and effectiveness. Several factors affected the success of teacher evaluation programs, including, teacher buy-in and fidelity of implementation.

Chapter One provided a brief review of the history of teacher evaluations and discussed the context in which the national movement toward reform of teacher evaluation transpired. Next, Chapter One explained the problem statement and theoretical research that guided this study in reference to the research questions. Finally, the key terms were defined, the underlying assumptions were presented, and the limitations associated with this research were discussed. Chapter Two is the review of literature, which includes: the history and evolution of teacher evaluation in the United States, problems with traditional teacher evaluation methods, new purposes and methods for monitoring and determining effective teachers, and implementation of new evaluation systems employing multiple measures of student and educator growth. Chapter Three

presents the research methodology and design.  Limitations of the study were also

presented.

Chapter Four contains the data and findings.  A *t*-test for difference in dependent

percentage means was calculated to compare student mean scores by teacher from year-

to-year, in order to detect whether a significant difference existed, indication growth.

Additionally, a Pearson Product Moment Correlation (PPMCC) test was calculated for

comparison of Teacher Evaluation Ratings and student achievement scores on the MAP

and i-Ready assessments, to determine the strength of the potential relationship between

the two variables.

Chapter Five presents an analysis of the statistical calculation findings.  The

hypotheses of the study were reviewed and conclusions were discussed.  The findings and

conclusions from the qualitative survey data were also presented.  Additionally,

recommendations for future research and suggestions for implications for practice are

included.

## Chapter Two: The Literature Review

**Overview**

The purpose of this study was to research the impact of teacher evaluation on student achievement, through an examination of a potential relationship between teacher's evaluation ratings and the academic achievement of their students.  Research confirmed student achievement was influenced and shaped by many factors, several of which were thoroughly researched (Berliner, 2013; Hattie, 2009; Hattie & Anderman, 2013; Organization for Economic Cooperation and Development [OECD], 2009).  This researcher recognized that these other factors existed; however, they were not the focus of this study.  Nonetheless, this chapter briefly covered some of the factors that influence student achievement, and thereafter focused exclusively on the impact of teacher evaluations on student achievement.

The literature review includes a brief history of teacher evaluations in the United States, the then-current research on the problems with teacher evaluations, and then-new methods for improving teacher evaluation processes.  Attention was also given to some of the most recent education system reforms, at the time of this writing, to address the issue of teacher evaluations and student achievement in the United States.

**Factors That May Contribute to Student Achievement**

Many experts agreed student learning was the best indicator of the quality of teaching (Darling-Hammond, 1997, 2000; Hattie, 2003; Kane & Staiger, 2008; National Commission on Teaching and America's Future, 1996; Rothstein, 2011).  Over several decades preceding this writing, educational researchers identified and studied a number of factors affecting student achievement and the increasing achievement gap in America

(OECD, 2009).  "We know from decades of educational research that many things matter for student achievement aside from the individual teacher a student has at a moment in time for a given subject area" (Darling-Hammond, 2015, p. 133).  Some of the most significant studies examined in-school and out-of-school factors found to have an impact on student learning and academic achievement.  Factors included school factors, such as class sizes, curriculum choices, amount of instructional time, and availability of resources;, outside-of-school factors, such as a student's prior educational experiences; cultural inequities related to race, class, and gender, and differences in learning or loss of learning over summer; home factors, such as parental support and involvement, socio-economic status, food and housing security, and history of physical and mental abuse; and individual student factors, such as health and attendance. (Coleman et al., 1966; Darling-Hammond, 2015; Reardon, 2011; Baker et al., 2010).

The most comprehensive meta-analysis of research related to student achievement was conducted by Hattie (2009), who synthesized years of the research related to established influences on student achievement.  Hattie (2009) structured his findings around six key influences on student learning — the child, the home, the school, the curricula, the teacher, and the approaches to teaching.  The study concentrated on instructional methods employed by teachers and ranked the 138 influences known at that time, according to their effect size.  He employed the use of a hinge point — a 4.0 mean effect size — to gauge each one's relative effectiveness (Hattie, 2009).  Since the original publication of his findings in 2009, Hattie continued to update his list with the findings of other studies and meta-analyses.  What was most noteworthy about Hattie's (2009) findings, and that of other researchers who sought insight into what mattered most in

teaching, was how factors with the lowest effect size (e.g. class size, longer school days, performance pay, school structure changes) typically dominated the dialogue related to education among educators, policy makers, and other stakeholder groups.  Hattie (2015) called this phenomenon *The Politics of Distraction*.

In summation, prominent educational reformers conceded many factors were present and influenced student learning.  Subsequent studies concentrated on evaluating the impact of the teacher on student learning.  As Wright, Horn, and Sanders (1997) explained, "The heart of this line of inquiry is the core belief that *teachers make a difference*" (p. 57).  The underlying assumption identified in the literature was that what teachers do matter and since teachers were the only element completely within the control of educational leaders, they, thus, deserve of utmost attention.  Since this study focuses on the identification of effective teachers, the remainder of this literature review discusses the research on identification of effective teachers, how this research framed the subsequent development of evaluation systems, and the impact teacher evaluations on student performance.

**History of Teacher Evaluation**

Historians established that "teaching has existed long before teacher evaluation" (Labaree, 2008, p. 291).  As typical for the period, no special preparation was needed, just a familiarity with the subject matter taught (Labaree, 2008).  Hence, the supervision of teachers and of teaching was far from a new concept.  Educational researchers agreed that teacher supervision in the United States dated back to the mid-1800s and was referred to as the Common Schools Movement (Marzano et al., 2011).  Common Schools, introduced by reformer Horace Mann, were a precursor to then-current day

public schools (as cited in Spring, 2001).  These early public, or free, schools were grounded in the belief that all children should have access to publicly-supported schools regardless of background or social status (Spring, 2001).  Prior to that time, education was not considered a formal discipline of study or a profession.

Experts believed the establishment of Common Schools had a profound impact on both teachers and the teaching profession by placing a greater demand for teachers, and by introducing women into the field (Levin, 2001).  As typical for the time period, the supervision of teachers was under the domain of local government and the clergy, who had the power to establish criteria for instruction and make administrative personnel decisions (Marzano et al., 2011).  There was no consideration of pedagogical expertise for either teaching or the act of supervision, much less the quality of feedback or professional development of teachers (Marzano et al., 2011).

Growth of Common Schools quickly led to a dramatic increase in the number of schools and a larger teaching force.  Soon, a more complex supervision model was necessary.  Thus, the adoption and development of the Common School Model prompted the transition to a more formal and structured model of schooling, with stronger centralized administration power to serve the variety of supervisory functions then required (Labaree, 2008; Marzano et al., 2011).  Marzano et al. (2011) recognized, the onset of more teachers led to the need to provide systematic training of educators.  As a result, Normal Schools (or teacher training schools) were born (Labaree, 2008).

While they took different forms, as the name suggested, the original goal of Normal Schools was to "set the standard — the norm — for good teaching" (Labaree, 2008, p. 292).  Normal schools provided a systematic approach to the training of teachers

to ensure a quality of teaching that was not present prior to the time period (Levin, 2001).

According to Labaree (2008), the conflicting demands of providing professional teacher

training, and pressure to mass-produce teachers to fill the high demand in schools,

resulted in the Normal School Movement succumbing to "choose[ing] relevance over

rigor" (p. 293).

The remainder of the century witnessed several changing tides in public

education, teacher preparation programs, and the move toward a professionalization of

the field of education (Marzano et al., 2011; Labaree, 2008).  According to Marzano et al.

(2011), the post WWII era marked a critical shift from a more industrialized view (a

hierarchical model) of education to one focused on the importance of individual teachers

and the importance of teacher quality, known as clinical supervision.  Clinical

supervision was defined as a method of management where supervisors worked with

teachers in a "close, helping relationship" (Okafor, 2012, p. 1) to improve knowledge and

skill of practitioners that would lead to greater classroom success (Goldhammer,

Anderson & Krajewski, 1980).  Two prevailing models of teacher supervision arose

during this clinical supervision era, primarily from the work of Cogan (1973) and

Goldhammer, Anderson, and Krajewski (1980).  These models became the basis upon

which most teacher evaluations were then structured (Marzano et al., 2011).  These

models consisted of classroom observations followed by post-observation conversations

with the administrator.  The research of Marzano et al. (2011) pointed out that the

original purposes of clinical supervision were much different from that into which teacher

supervision evolved.  Clinical supervision models for teacher evaluations were designed

to evaluate the effectiveness of teachers and were intended for the purpose of improving

instruction (Danielson & McGreal, 2000; Marzano et al., 2011).

Marked research interested in drawing a correlation between teacher behaviors

and student learning "confirmed the critical role that teachers play in student learning"

(Danielson & McGreal, 2000, p. 14).  Since then, numerous studies established the

impact of teachers on students and showed that variations in student achievement were

associated with differences in the effectiveness of teachers (Rivkin, Hanushek, & Kain,

2005; Rockoff, Jacob, Kane, & Staiger, 2008; Wright, Horn, & Sanders, 1997).

Educational historians established one of the most prominent individuals, who

developed the foundational instruments for evaluation during that era was Madeline

Hunter (as cited in Marzano et al., 2011).  Hunter's (1982) research on specific teaching

strategies, called Mastery Teaching, "started a trend toward increased *instructionally*

*focused* staff development that persists to this day" (Danielson & McGreal, 2000, p. 13).

While Hunter's (1982) research centered on a structured sequence of lesson planning, her

ideas also contributed to processes of teacher evaluation and professional development

(as cited in Danielson & McGreal, 2000).

In 1996, the research of Danielson, led to the development of the Framework for

Teaching (FFT) model, which promoted "clear and meaningful conversations about

effective teaching practice" (Danielson Group, 2013, para. 1).  *Frameworks for Teaching,*

founder Danielson (2011), communicated the important requirement that a quality

evaluation systems began with a consistent definition of good teaching, and "everyone in

the system — teachers, mentors, coaches, and supervisors — must possess a shared

understanding" (p. 35).  The Danielson model became one of the most widely used model

for evaluating teaching competence in the United States and was often used as "the

reference point for any new proposals regarding supervision and evaluation" (Marzano et

al., 2011).

**Performance-based teacher assessment.**  Research suggested the turn of the

century marked another shift in emphasis from the supervision of teachers and their

behaviors to the evaluation of teaching linked with student achievement (Marzano et al.,

2011).  Before this time, the most common form of teacher evaluations relied on

infrequent observations from building principals, using an observation checklist where

evaluators looked for specific observable behaviors.  Previous research unearthed most

teacher evaluation observations were completed merely as required compliance tasks,

which were not taken seriously by either teachers or administrators (Danielson &

McGreal, 2000; Marzano et al., 2011; Ravitch, 2010; The New Teacher Project, 2010).

Researchers also agreed, while teacher evaluations could serve a valid purpose and were

necessary for evaluating teacher performance; classroom-observation-based evaluations

continued to be "at best incomplete measures of teaching that produces[ed] gains in

student achievement and attainment" (Taylor & Tyler, 2011, p. 7).  The report, "The

Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher

Effectiveness," authored by Weisberg, Sexton, Mulhern, and Keeling (2009) reached a

similar conclusion.  "The Widget Effect" confirmed the ineffectiveness of teacher

evaluation systems and cited an "institutional indifference to variations in teacher

performance" (p. 4).

Despite of these criticisms, several educational experts and practitioners believed

that well-designed evaluation systems, which incorporated robust observations, would

improve average teacher performance through the development and improvement of skill (Taylor & Tyler, 2012, The New Teacher Project, 2010).  The research of Taylor and Tyler (2011) also suggested teacher evaluation programs had sustainable results and could "improve performance even after the evaluation period ends" (p. 29).  Grissom and Youngs' (2016) examination affirmed these findings but cautioned, "Rigorous teacher evaluation systems have the potential to promote school improvement, but only if the systems are carefully designed and implemented and the data they generate are interpreted and used appropriately" (p. 2).

Subsequent revised teacher evaluation processes of the decade of this writing were based on indicators of teacher performance and connected to student outcomes (Melvin, 2011).  This method, known as performance or standards-based evaluation, represented a strategy for "both improving instruction and complying with the expectations of external stakeholders that teachers be held accountable for their performance" (Milanowski, Kimball, & White, 2004, p. 2).  According to Odden (2004), standards-based teacher evaluation systems required the following:

A set of teaching *standards* that describes in considerable detail what teachers need to know and be able to do; a set of *procedures* for collecting multiple forms of data on teacher's performance for each of the standards; a related set of *scoring rubrics* that provide guidance to assessors or evaluators on how to score the various pieces of data to various performance levels and a scheme to aggregate all micro-scores to an overall score for a teacher's instructional performance; and a way to use the performance evaluation results in a new knowledge and skills-

based salary schedule if the evaluation system is to be used to trigger fiscal

incentives. (p. 127)

These systems recognized teaching as a complex endeavor with many aspects to

consider when evaluating.  One of the requirements for state eligibility for federal Race

to the Top grants was development of "rigorous, transparent, and fair evaluation systems

that provide regular feedback on performance to teachers and principals" (USDOE, 2009,

p. 6).  In response, states across the country developed major reforms to their teacher

evaluation systems which incorporated a more comprehensive assessment of a teacher's

ability and performance, and merged multiple ratings, including data from student

achievement outcomes (Shakman, Riordan et al., 2012).

While having a strong potential for influencing teacher performance, the early

applications of Performance-Based Evaluations had several flaws (Shakman, Riordan et

al., 2012).  Among the most acknowledged design flaws were the infrequency of teacher

evaluations; evaluations often not based on specific teacher behaviors or pertained to a

teacher's direct impact on student learning; the lack of differentiated rating categories

(typically pass/fail); not providing useful feedback to teachers or impact teachers'

performance; and the results of the evaluations systems were inconsequential for staffing

decision making (Weisberg et al., 2009).   The report of Shakman, Riordan, Sanchez,

DeMeo Cook, Fournier, and Brett (2012), endorsed by the National Center for Education

Evaluation and Regional Assistance, supported these conclusions.

However, as studies brought attention to the weak state of teacher evaluations, the

literature revealed an evolution in teacher evaluation from one serving more of a quality

assurance function to a more highly standardized, educational accountability function

(Davies, 2008; Ravitch, 2002).  According to Pellegrino (2004), this increased demand

for public accountability "[could] be observed every day and in multiple forms especially

in the press and in public and political discussions about the need to improve the

educational system" (p. 5).

**Determining teacher quality.**  The evolution of teacher evaluations, from one

focused on quality assurance to one focused on accountability, soon pointed out the

necessity for more clearly defined standards of expectations for teachers.  The NCLB Act

(2001) launched this trend when it mandated a highly-qualified teacher in every

classroom.  NCLB initially defined teacher quality by a teacher's credentials or subject

knowledge and training.  Goe (2007) conducted a synthesis of research analyzing teacher

quality variables and their link to student achievement for the National Comprehensive

Center for Teacher Quality.  This analysis uncovered a lack of a common definition of

teacher quality, how to measure it, and which teacher qualities mattered most in terms of

student achievement.  Additionally, the research named four primary strands for defining

and measuring teacher quality: teacher qualifications, teacher characteristics, teacher

practices, and teacher effectiveness (Goe, 2007).  Another significant discovery identified

in Goe's (2007) research was that all of the studies used standardized test scores to

measure student learning for identifying teacher contributions.  Interestingly, since these

early stages of attention on the quality of teachers, the use of student achievement results

to determine teacher quality soon became the norm (Goe, 2007).

As previously discussed in this chapter, researchers recognized there were many

factors, both in-school and out-of-school, which influenced student academic

achievement (Barton & Coley, 2009; Berliner, 2009; Hattie, 2009; Ladd, 2012;

Goldhaber, 2016).  However, leading educational experts and consultants suggested

schools focus on factors over which school districts had the most control — those related

to teachers and schools (Hattie, 2009; Goldhaber, 2016; Marzano, 2003; Schmoker, 2011;

Tucker & Stronge, 2005; Whitaker, 2013).  Therefore, this study focused on the literature

encompassing in-school factors and their potential relevance for policy reform —

specifically those associated with the teacher evaluation process.

One of the most important factors in student learning, repeatedly established as an

area under the influence of the school district, was the quality of the teaching (Marzano,

2007; CPE, 2005; Chetty, Friedman, & Rockoff, 2014).  The growing body of literature

maintained good teaching had a profound effect on students, led to substantial growth in

achievement, was cumulative, and had a lasting impact for future student success

(Bracey, 2004; Goodwin, 2011; Hanushek & Rivkin, 2012; Tucker & Stronge, 2005;

Schmoker, 2011).  Schmoker (2011) proposed, "It is now a well-established fact that

even three years of fairly ordinary but effective teaching can completely change the

academic trajectory of low-achieving students — vaulting them from the lowest to the

highest quartile" (p. 51).  Additionally, Goodwin (2011) suggested, "Highly effective

teachers . . . help students learn, on average, the equivalent of a year and a half of

learning in a single year, while those in the bottom fifth only impart an average of a half

year of learning" (p. 18).  Additionally, the Center for Public Education (CPE, 2012)

publicized, "Research indicates that the achievement gap widens each year between

students with most effective teachers and those with least effective teachers" (para. 1).

Another longitudinal study, which tracked 2.5 million students over five years, publicized

some of the long-term positive effects of highly effective teachers on students, including

students progressing to a four-year college and earning a higher income throughout their

lifetime (Chetty et al., 2011).  Another related study, which consisted of over 200,000

elementary and middle students in over 3,000 public schools, concluded teacher

effectiveness had a greater effect on student achievement than all other commonly

considered factors, such as class size, after school programs, or the particular school of

attendance (Rivikin, Hanushek, & Kain, 2005).  It was the belief of several leading

education experts, including Whitaker (2013), Marzano (2003), and Schmoker (2011),

the key would then to be to ensure and sustain quality teaching.  Jordan, Mendro, and

Weerashinghe (1997), authors of *The Effects of Teachers on Longitudinal Student*

*Achievement: A Report on Research in Progress*, explained:

> Research also revealed that the best predictor of a teacher's effectiveness is his or
>
> her past success in the classroom.  Most other factors pale in comparison,
>
> including a teacher's preparation route, advanced degrees, and even experience
>
> level (after the first few years).  The lesson is clear: to ensure that every child
>
> learns from the most effective teachers possible, schools must be able to gauge
>
> their teachers' performance fairly and accurately.  (Jordan, Mendro, &
>
> Weerasinghe, 1997)

In light of these findings, policy-makers shifted emphasis of school reform efforts

recent to this writing, to establishing the relationship between teacher performance

(quality instruction) and student achievement.  Hanushek (2014) specified, "The naïve

calls for 'highly qualified teachers' in the No Child Left Behind act have been replaced

by recognition that credentials and qualifications — the objects of past policies — are not

closely related to teacher effectiveness in the classroom" (p. 28).  While still maintaining

previous expectations of a highly qualified teacher in every classroom, an executive

summary by the National Council on Teacher Quality (NCTQ), stated the new constraint

for highly effective teachers (NCTQ, 2011).  Consequently, reauthorization of NCLB

replaced the notion of highly qualified teachers to a more intense focus on 'highly

effective' teachers and on how they relate to the performance of students (Hanushek,

2014; Hanushek & Rivkin, 2012).

Despite this trend, research established little agreement on a clear definition of

what constituted effective teaching.  Goe, Bell, and Little (2008) affirmed,  "Effective

teaching has been defined in many ways throughout the years, and methods for

measuring teachers have changed as definitions and beliefs about what is important to

measure have evolved" (p. 2).

**Failure of Current Teacher Evaluations**

The growing body of research highlighted a wide disparity in effectiveness of

teachers and variation among teachers' ability to improve student performance

(Hanushek, Kain, O'Brien, & Rivkin, 2005; Kane et al., 2011b; Wright et al., 1997;

Rivkin et al., 2005; Rockoff, 2004; Rothstein, 2008; Stronge, 2013).  Some experts, like

Darling-Hammond (2010), proposed the establishment of Performance Assessments for

Teachers.  She argued,

> Unlike most high-achieving nations, however, the United States has not yet
>
> developed a national system of supports and incentives to ensure that all teachers
>
> are well prepared and ready to teach all students effectively when they enter the
>
> profession.  Nor is there a set of widely available methods to support the

evaluation and ongoing development of teacher effectiveness throughout the career (Darling-Hammond, 2010, p. 2).

Hattie's (2003, 2009, 2012, 2015) extensive analysis of the research advocated that poor teachers rarely did damage, but confirmed a wide variance existed in the effectiveness of teachers, especially as it related to student achievement.  Nonetheless, he proposed, "We need to identify, esteem, and grow those who have powerful influences on student learning" (Hattie, 2003, p. 4).  To that end, the priority for educational leaders must be the ability to identify, sustain, and improve quality teachers.

A question that repeatedly surfaced during the review of literature was, 'How can you accurately, and fairly, measure the quality/effectiveness of a teacher?'  The Center for Research on Learning and Teaching (CRLT, 2016) explained the complexity of creating reliable, valid, and fair evaluation systems, especially in light of the need for teacher evaluations to serve a number of distinctive purposes.  The CRLT recommended the following guiding principles for teacher evaluation systems: teacher evaluation systems should incorporate the use of multiple sources of data of both teacher and student performance; the development of these evaluation systems should engage stakeholders at all levels and across all disciplines; they must be individualized systems of evaluation; and teacher evaluation systems should be flexible enough to accommodate diverse instructional methods (Center for Research on Learning and Teaching, 2016).  In addition to these components, CCSSO (2016) cautioned that all measures of effective teaching were balanced with professional judgment when assigning summative ratings. Furthermore, the CCSSO (2016) recommended that the integrity of teacher evaluation

and support systems required regular examination to ensure consistency of their

implementation, as well as to ensure the consistency and accuracy of the data used.

The ever-growing body of research committed to studying teacher effectiveness

and providing accurate assessments for evaluating educators (Darling-Hammond, 2010;

Goe, Bell, & Little, 2008; Hinchey, 2010; Stronge, Ward, & Grant, 2011).  One of the

most prominent teacher effectiveness research studies was the Measures of Effective

Teaching Project (MET, 2013), funded by the Bill & Melinda Gates Foundation.  The

MET study was designed with two overarching goals:  to determine whether it was

possible to identify effective teachers, and to determine the best method for measuring

the effectiveness of teachers.  The project compiled its findings and reported them over

the years in three separate research briefs, in an effort to assist districts in their endeavors

to reform their teacher evaluation systems using the research findings.  In its overall

conclusion, the final research brief supplied evidence which maintained it is in fact

possible to both identify, and predict, the contributions of individual teachers on his or

her respective students' learning.  Additionally, the research concluded evaluating

teachers fairly and accurately required the use of multiple indicators, including classroom

observations, student surveys, and student achievement gains, as the most valid and

reliable method for identifying and measuring the effectiveness of teachers (Bill &

Melinda Gates Foundation, 2013).  This comprehensive research built upon the premise

that the purpose of evaluating teacher performance was to lead to better learning and

achievement outcomes for students (Bill & Melinda Gates Foundation, 2013).  The New

Teacher Project's (2010) brief, *Teacher Evaluation 2.0*, harmonized these views and

attempted to address the question of how to design meaningful, valid, and reliable teacher

evaluation systems.  In short, the report proposed six feature design standards most vital

for transforming teacher evaluation systems, including: clear, rigorous expectations;

multiple measures; multiple ratings; and regular feedback.  Thus, the underlying

agreement of educational researchers and practitioners was that teacher evaluations – the

best lever to change teacher practice at scale – must provide teacher clear expectations,

feedback, and support (Coe, Aloisi, Higgins, & Major, 2014; Darling-Hammond, 2012).

The research of education experts, including Hattie (2003, 2009, 2012, 2015) and

Marzano (2003, 2007, 2012), among others, removed the mystery of what made teachers

effective.  Hattie (2003) concluded, "It is teachers using particular methods, teachers with

high expectations for all students, and teachers who have created positive student-teacher

relationships that are more likely to have the above average effects on student

achievement" (p. 126). Hattie's (2009) report, *Visible Learning*, identified and ranked the

most successful teaching strategies for improving student learning.   In addition, a McRel

Industries (2009) annual report, *What Matters Most*, detailed a framework indicating

"five opportunity areas where improvements can lead to dramatic gains in student

achievement" (para. 7).  Their suggested high-payoff areas included: "1) guarantee of

challenging, engaging, and intentional instruction; 2) ensure curricular pathways to

success; 3) provide whole-child student supports, 4) create high-performance school

cultures; and 5) develop data-driven, high-reliability systems" (McRel, 2009, para. 8).

**Dual Purposes of Teacher Evaluations**

In the decade previous to this writing, federal and state mandates, as well as the

public's demand for accountability, prompted districts across the country to make

concerted efforts to overhaul their teacher evaluations.  The teacher evaluation system

modifications were initiated to provide more objective and accurate indicators of teaching

effectiveness, and then to use the information to guide on-going improvement in teacher

performance and student learning (American Federation of Teachers [AFT], 2012, 2015;

Hull, 2013a; Jerald, 2012; Stanford Center for Opportunity Policy in Education, 2013).

Districts executed a number of system changes, including designing new measures of

effectiveness and establishing new protocols, rubrics and teaching standards (AFT,

2012).

As clarified by Milner (2010), the real and "on-going challenge has been to agree

on actionable, working definitions specifying what it means to be a teacher of high

quality" (p. 1). It is apparent that evaluations can serve different purposes - for the

teacher, building administrator, or district administration (Young, Range, Hvidston, &

Mette, 2015). The research suggested, in order to create more effective and relevant

teacher evaluation and feedback systems, it was first necessary to define their intended

purposes.

The review of literature found even leading researchers in the field of education

debated the purposes of teacher evaluations (Young et al., 2015). *Frameworks for*

*Teaching* founder, Danielson (2012), believed teacher evaluations should center on

ensuring teachers were competent and to support teachers' professional growth. In

addition to developing teachers, Marzano (2012) suggested that teacher evaluation should

also measure teachers. Looney (2011) harmonized these views and advocated for the

differentiation of teacher evaluations depending on individual teacher performance, the

school context, or student outcomes. Similarly, National Institution for Excellence in

Teaching researchers, Daley and Kim (2010), advised teacher evaluation be utilized for

three primary purposes: to convey teacher expectations, to assess teachers' then-current performance — or ability, and to plan appropriate professional development geared toward the development of increased professional competence.  Similarly, Jerald's (2012) analysis stated the goals of teacher evaluation as either moving it or improving it purposes.

Maslow and Kelley's (2012) research incorporated both the formative aim — to support improvement in individual teaching practice and for summative judgment regarding individual teachers.  However, the researchers also suggested they afford a method for providing systematic feedback — to inform the management of human resources in the district — on the school system as a whole (Maslow & Kelley, 2012).  Stated another way, Stronge's (2006) breakdown asserted teacher evaluation systems served two broad functions — accountability-oriented and improvement-oriented purposes.

In the end, it was important to consider teacher evaluation as an important and natural component of the learning process, for teacher as well as students (Benedict, Thomas, Kimerling, & Leko, 2013).  CCSSO (2016) also forewarned, "Teacher support and evaluation systems should strike a productive balance between support and accountability, and should be designed to be continuously improved and evolve over time" (p. 3).

**Formative evaluation to promote teacher development.**  Formative evaluation was defined as a process of teacher evaluation that gathers information regarding a teacher's performance and uses it to provide feedback for modifying and improving on-going teaching practices (Black & Wiliam, 2004).  Since research publicized teacher

evaluation systems had the potential to serve as a powerful tool to provide formative

feedback to teachers, experts suggested evaluation systems reinforce "effective teaching

and assessment practices" while identifying areas for improvement (Looney, 2011, p.

22).  Similarly, Baker et al. (2010) maintained that progress in the development of

standards-based teacher evaluation practices led to improvements in teacher effectiveness

and student achievement gains.  These revised teacher evaluation models took a

formative approach and encompassed a comprehensive model of teacher expectations,

which included "explicit standards in multiple domains for multiple levels of

performance," and included comprehensive teacher behavioral ratings (Looney, 2011, p.

22).

According to Marzano (2012), teacher evaluation systems focused on the

development of teachers had three primary characteristics; they: 1) were comprehensive

and specific, 2) included a developmental scale for teachers to identify their current level

of performance and track their development, and 3) both acknowledged and rewarded

teacher improvement (Marzano, 2012).  Conversely, the *Teacher Assessment and*

*Evaluation* whitepaper, endorsed and published by the National Education Association

(National Education Association [NEA], 2010) recommended that teachers' engagement

in formative assessments for the improvement of their practice involved neither

punishment nor reward, but instead facilitated "interaction and feedback, with a collegial

relationship between the administrator and teacher, to encourage reflection and

discussion" (NEA, 2010, p. 5).  Their view of the process of formative assessment

proposed one which was "open, exploratory . . . and focused on practitioner development

and practice" (NEA, p. 6).

Moreover, research recent to the time of this writing determined most states made considerable progress setting up the revamped teacher evaluation systems created within the previous several years (McGuinn, 2015).  Due to the still emerging implementations of then-new district teacher evaluation systems, there were a limited number of studies available for review.  A study recent to this writing, by McGuinn (2015), evaluated the progress of early-implementing districts.  McGuinn's (2015) research acknowledged most states had only just begun full implementation of the latest teacher evaluation systems, and further revealed resultant challenges, which necessitated further scrutiny and revisions to the newly-established teacher evaluation systems.

**Summative evaluation to guide personnel decisions.**  Teacher evaluations must take into consideration both formative and summative feedback for teachers.  According to Stronge (2006), "Teacher evaluation is, first, about documenting the quality of teacher performance; then, its focus shifts to helping teachers improve their performance as well as holding them accountability [sic] for their work" (p. 1).  The purpose of summative evaluations was defined as a tool to use overall performance data to judge the quality of teaching — in light of the district's established criteria for teacher performance, and to make personnel decisions (NEA, 1995).  Essentially, summative teacher evaluations served an accountability function (Santiago & Bevavides, 2009).

The ever-growing body of literature demonstrated how policy makers pushed more and more for districts to link teacher evaluation data with human resource decisions, such as: for hiring teachers, for continuing contracts, and making tenure status decisions (Aldeman & Chuong, 2014; Darling-Hammond, 2013; Hull, 2013a; Hull, 2013b).  In addition to determining teachers' status, research suggested district evaluation

processes also identified and removed poor performers (Hall, Diaz-Bizello, & Marion, 2015).  Darling-Hammond (2013) agreed evaluation systems (based on standards of professional practice) must also remove individuals from the profession when they do not, after receiving assistance, meet professional standards.  Of course, these systems should also include a system of due process and review.  To that end, new education evaluation systems must include several processes — ones for improving and developing underperforming teachers, as well as for removing underperforming teachers.  According to the NEA (2010),

> Teachers who fail to meet acceptable standards should be offered professional development, remediation plans, and opportunities to observe peers. They should also be given sufficient time, support, and assistance toward meeting the standards. A process to remove chronically ineffective teachers from the classroom should begin only after extensive support and intervention that guarantees due process measures.  (p. 5)

Several experts recommended basing summative evaluations on an established set of performance standards (Danielson, 2011; Hall et al., 2015; OECD, 2009).  The NEA (2010) further recommended summative performance standards be identical to standards used in the ongoing formative process (p. 3).  Danielson's (2011) FFT model, one of the most prominent performance-based evaluation systems then-currently available and carefully researched, provided a valuable instrument for districts.  Creation of the FFT model began with establishing clear descriptions for what teachers should know and be able to do, and detailed teaching practice expectations (Danielson, 2011).  The model described each level of performance with a set of rubrics, which explicitly described

observable teacher behaviors, called critical attributes, for determining between each level of performance for each teaching standard (Danielson, 2011).

The review of available literature for this current study identified a finding alarming to this researcher — the absence of research available which explored the use of teacher evaluation data in administrative decisions.  Murphy, Hallinger, and Heck (2013) also noted the dearth of research on the use of teacher evaluation data as a tool of school improvement, and on the costs associated with managerial systems, such as teacher evaluation.

According to some experts, the management of human capital may be the most important function of the school-system management operation (Darling-Hammond, 2013; Hess & Fullerton, 2010).  Hess & Fullerton (2010) further concluded, improvement of teaching required a system capable of consistently monitoring teachers and measuring their performance, and having the capacity to manage the processes of employee hiring, transfers, and termination.  One of the foremost teacher evaluation experts, Darling-Hammond (2013), also identified the lack of attention to the administrative capacity of school district systems for supporting teacher evaluations.  She reasoned,

> One serious shortcoming of teacher evaluation reforms is that they have often focused on designing instruments for observing teachers, without developing the structural elements of a sound evaluation system.  These elements should include, at a minimum: trained, skilled evaluators; supports for teachers needing assistance; governance structures that enable sound personnel decisions; and resources to sustain the system.  (Darling-Hammond, 2013, p. 115)

This lack of attention to planning beyond the evaluation design, Darling-Hammond (2013) reasoned, could result in unsustainable evaluation systems.

This leads the discussion to the two primary, and divergent, approaches related to human-capital management — one proposing professional development as the most effective lever for improving teacher effectiveness, and the other promoting the removal of underperforming teachers.   Researchers have reported no findings able to establish an empirical link between professional development and increased student achievement (Krasnoff, 2014; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007).  Additionally, some researchers concluded while studies suggested good quality professional development may result in improved teacher knowledge and instructional practice for the short-term, the studies have not established sustained change in teaching practice over time or in improved student outcomes (Garet et al., 2008).  To that end, the whitepaper, *Movin' It and Improvin' It!* (Jerald, 2012), suggested both professional development approaches could enhance teaching effectiveness, and leveraging them together would deliver greater gains for students.

**Methods for Evaluating Teacher Effectiveness**

Given the absence of a universal definition of teacher effectiveness guiding the discussion, and lack of consensus on the proposed purpose and outcomes of teacher evaluations, several methods evolved.  According to Grissom and Youngs' (2016) research, many states and local school districts implemented new teacher evaluation systems that combined teacher observation ratings, student survey feedback, and statistical estimations of a teacher's "value-added" impact on student achievement as a result of the new federal and state mandates for teacher evaluation reform (p. 1).  The

authors also believe these new metrics to "produce more comprehensive measures of

teacher performance than have ever been available before" (Grissom & Youngs, 2016, p.

1).  Clearly alleged by educational experts, "There are many potential measures of

teacher performance that a state or district could use as part of the evaluation process"

(i.e., value-added models, other growth models, and Student Learning Objectives) (Goe,

Holdheide, & Miller, 2014, p. 19).  Additional measures that had the potential to better

capture teacher practice included: observation instruments, performance rubrics,

portfolios, teacher self-assessments, and parent/student surveys (Goe et al., 2014, p. 20).

This section highlights three of the most prominent and promising methods for

evaluating teacher effectiveness, at the time of this writing, and provides a brief summary

of the research related to their benefits and challenges.  Berk (2005), John Hopkins

researcher, issued a report titled, *Survey of 12 Strategies to Measure Teaching*

*Effectiveness*, in which he summarized the research surrounding possible data sources to

use when evaluating teachers.

**Value Added and other student growth measures.** The review of literature

established that policymakers were increasingly looking at outputs — such as student

achievement and growth measured by standardized assessments — to evaluate and hold

individual teachers accountable, as a mechanism for improving school performance and

student outcomes (Cohen & Goldhaber, 2016; Corcoran, 2010, 2016; Goe et al., 2008;

Grissom & Youngs, 2016; Piro & Mullen, 2013; Harris, Ingle, & Rutledge, 2014).  Many

policymakers saw the use of value-added models (VAMs) in teacher evaluations as one

possible source for teacher accountability and educational reform (McCaffrey,

Lockwood, Koretz, & Hamilton, 2003).  VAMs were statistical models that concentrated

solely on standardized test scores in an attempt to measure a teacher's impact on his/her

students' achievement (Cocoran, 2012; RAND, 2012a).   Essentially, they were

attempting to isolate the 'value' the teacher added from other factors that affected

achievement (including a student's individual ability, and the influence of the student's

home environment, past schooling, and other factors) (RAND, 2012a).  VAMs also

analyzed the assessment data in order to determine "how a teacher's students performed

relative to comparable students in the same grade taught by other teachers in the same

state or district" (Corcoran, 2016, p. 51). VAM data was often analyzed and used to

calculate teacher effectiveness rankings, reported as percentile rankings "based on

whether students meet, exceed, or fail to reach their predicted scores on the test" (Goe &

Croft, 2009, p. 4).  Still, Cocoran (2010) also pointed out that value-added was a relative

concept, with both teachers and students graded on a curve, which "rest[ed] exclusively

on skills assessable on very narrow standardized tests" (Corcoran, 2010, p. 14).

Whereas the term value-added may have been new, the concept of measuring

teacher performance based on student test scores was a documented trend throughout the

history of formal education (Harris & Herrington, 2015, p. 71).  Federal and state level

policies, recent to this writing, — such as the NCLB Act — represented this focus by

their mandated use (and incentivizing) of value-added measures to assess student growth

(Corcoran, 2016; Harris et al., 2014; Jorgenson, 2012, Ballou & Springer, 2015).

Researchers Harris and Herrington (2015) acknowledged, "Nothing in the past compares

with the wave of value added-based teacher accountability brought on by President

Obama's Race to the Top" (p. 71).  As a result, many states and districts had a VAM

component as part of their teacher performance evaluation systems to varying degrees

(Amrein-Beardsley, Collins, Polasky, & Sloat, 2013; Doherty & Jacobs, 2013).  Harris et al. (2014) suggested that the goals "behind these initiatives are two-fold: to improve teacher quality, and to make high-stakes decisions about teachers' careers" (p. 2). Proponents of VAMs, such as Ritter and Schuls (2012) believed,

> If citizens, policy makers and educators have decided that the primary objective of schools is to foster student learning and if we have the tools to adequately measure student learning, then it naturally follows that we should be assessing teacher effectiveness based in large part on the learning gains of students in the classroom. (p. 34)

The review of literature identified five primary benefits of using VAMs for teacher accountability.  According to Little, Goe, and Bell (2009) one advantage of value-added models was they were highly objective "because they do not involve raters making subjective judgments" (p. 5).  These authors also suggested that VAMs were also beneficial for districts because they were cost efficient and nonintrusive, as "they require no classroom visits, and test score data are already collected for NCLB purposes" (p. 5). Moreover, several studies concluded that VAMs could accurately measure the significant differences in teacher effectiveness (Chetty et al., 2014; Hanushek & Rivkin, 2010; McCaffrey, Koretz, Lockwood, & Hamilton, 2004).  Furthermore, several researchers advocated the use of VAMs as a validation measure of teaching effectiveness for evaluating the impact of larger policies, programs, and interventions (Baker et al., 2010; Corcoran, 2016; Reform Support Network, 2013).

Conversely, research acknowledged challenges, or unintended consequences, to the use of VAMs in practice.  Corcoran (2010) argued the extreme difficulty of "isolating

a teacher's unique contribution" (p. 4).  Grissom, Loeb, and Doss (2016) raised concern

that evaluation systems favoring value-added measures "are likely to overlook important

contributions to the school that many teachers make" and further suggested "the

multidimensional nature of teachers' work requires a multiple-measures approach to

evaluation" (Grissom et al., 2016, p. 4).

One criticism central to the apprehension over VAMs was whether VAMs were a

valid and reliable tool for identifying teacher effectiveness.  Educational experts and

researchers called into question their ability to control for numerous other factors that

may contribute to student achievement outside the teacher (Baker et al., 2010; Corcoran,

2010; Schochet & Chiang, 2010).  Research demonstrated a variability of teachers' VAM

scores — even within schools (Little, Goe, & Bell, 2009, p. 5) and identified

methodological problems related to their use — such as missing data and non-random

assignment of students to teachers (American Statistical Association, 2014; McCaffrey et

al., 2003; Rothstein, 2008).  Additionally, the fact that VAMs focused exclusively on

standardized assessment data raised concerns that they falsely assumed, therefore, "that

student test scores are valid, reliable indicators of learning" (Little et al., 2009, p. 5).

Another prominent concern over VAMs is that they cannot be calculated for all

(or even most) teachers within a district (Goe & Croft, 2009).  Since VAMs only

provided teacher effectiveness data for teachers who had students with standardized test

scores, VAMs were not available for teachers of non-tested subjects and programs.

Additional concerns arose in situations where teams of teachers shared responsibility for

student learning.  Corcoran (2016) identified two related issues.  Some researchers

suggested VAMs were available for less than 30% of teachers within most districts (Baker et al., 2010; Grissom & Youngs, 2016).

In light of differing levels of support for the use of VAMs as a reliable and appropriate tool for determining teacher effectiveness, researchers concurred that education leaders should not use VAMs as the exclusive method for defining a teacher's effectiveness (American Educational Research Association, 2015; Baker et al., 2010; Corcoran, 2016; Haertel, 2013).  Harris and Herrington (2015) further proposed the real question was not whether VAMs were valid tools, but whether they could be useful for informing and improving teaching and learning.  Therefore, a fourth issue related to the use of VAMs in teacher evaluation programs was the extent to which the data was useful for achieving the intended purposes (Grissom & Youngs, 2016).  Consequently, Corcoran's (2016) research recommended, "limiting their role to a supporting one may be a better strategy for genuine, meaningful, and lasting reform" (p. 59).

**Classroom observations.**  The review of research verified the time-honored, traditional method of classroom observations had consistently been the most common method used in evaluating teachers (Berk, 2005; Goe & Croft, 2009; Little et al., 2009; Mathers, Olivia, & Laine, 2008).  Steinberg and Donaldson (2014) reported that despite the more recent focus on the use of student test scores in evaluating teacher performance, the majority of teachers still received evaluation ratings based largely on observations of their classroom practice.  Research showed classroom observations can serve as a useful tool for improving teacher performance in the classroom and that they are generally accepted as credible by the majority of stakeholders (Little et al., 2009).  In contrast, Weisberg et al. (2009) suggested evidence that classroom observations had not been

effective for differentiating or distinguishing between teachers (2009).  Goe and Croft (2009) contended, "Classroom observations provide a useful measure of teachers' practice but little evidence about whether students are actually learning" (p. 5).  Cohen & Goldhaber (2016) asserted, "Classroom observations have strong face validity because they assess 'process,' or teaching variables, not student outcomes, which may feel distal from teachers' work" (p. 9).  However, researchers also discouraged the use of observational methods in isolation for adequately capturing a teacher's performance or for identifying a teacher's effectiveness (Steinberg & Garrett, 2015).

Little et al. (2009) pointed out that observations varied widely in their implementation, what they were looking for, and how they evaluated teachers. Observations could be formal or informal, scheduled or unannounced.  While there was no optimal number of observations suggested in the research, "The implication is that more is better," for the purpose of formative teacher evaluation (Marzano, 2012). Darling-Hammond, Cook, Jaquith, and Hamilton (2012) also recommended that successful teacher observation systems employ multiple classroom observations ranging the entire academic year.  Additionally, the research revealed that frequent, short, unannounced classroom observations give the most accurate picture of what goes on in the classroom and therefore were the most effective way to accurately observe teachers (Marshall, 2012).

Observations can serve short-term outcomes (i.e., formative purposes to immediately impact teacher performance), as well as be used for making long-term (summative) decision-making (Little et al., 2009).  However, Darling-Hammond et al., (2012) purported that to be effective, the purpose of the observation must focus on the

timely and meaningful feedback given to the classroom teacher for change or to improve instruction.  Researchers also recommended that teacher observations used for summative purposes take into consideration multiple years of data to avoid mischaracterization of a teacher's effectiveness (Steinberg & Garrett, 2016, p. 21).

While classroom observations had many benefits, there were also a number of concerns and disadvantages related to their implementation and use.  A study by Steinberg and Garrett (2016) highlighted that a number of factors significantly influenced both a teacher's classroom performance (previously discussed) and the observation-based measures of the teacher's performance.  Some of these factors include fidelity of implementation, rater reliability, and bias concerns.  A study recent to this writing identified evaluation raters as the largest source of error when employing observations as an evaluation tool, because evaluators were always influenced by their own subjective bias (Cohen & Goldhaber, 2016; Hoe & Kane, 2013).  It is important to remember that classroom observations are a subjective measure of teaching, and by their nature will vary by evaluator (Little et al., 2009; Darling-Hammond et al., 2012).

Consistently, the review of research underscored the importance of implementing observation instruments with fidelity and integrity to obtain valid and reliable data (Eisenbach, 2014).  Cohen and Goldhaber (2016) hypothesized that research showing the concerns about the limitations and biases of value-added measures may have pushed practitioners to more highly value observation-based measures, yet observation instruments face many of the same sources of inaccuracy and bias.  Pianta and Hamre (2016) outlined several fundamental components that must be in place to facilitate high-quality observation systems in order to produce reliable measures.  The authors further

proposed the usefulness of the information gathered from teacher observations was unlikely without the time and investment in creating a highly reliable, valid, and standardized evaluation system (Pianta & Hamre, 2016).  Cohen and Goldhaber (2016) concurred and further conveyed that a "necessary first step is a carefully designed system for training and certifying observers, or raters, to use the tools consistently with the theoretical principles underlying the scales" (p. 15).  Additionally, they advocated the need for further research to better understand how observational measures "are sensitive to true changes and practice" (Cohen & Goldhaber, 2016, p. 17).

Unfortunately, as Grissom and Youngs (2016) discovered, implementation of many new observation systems has been initiated "with insufficient training for raters and too little attention to ensuring fidelity to instruments and protocols, with some educators raising concerns that they provide information no more useful than what was previously" (2016, p.2).  Goe, Holdheide, and Miller (2014), in collaboration with the Center on Great Teachers and Leaders, authored a *Practical Guide to Designing Comprehensive Teacher Evaluation Systems,* to serve as a reference for states and districts in developing teacher evaluation systems.  In this document the authors defined eight critical components of designing a comprehensive teacher evaluation system that addressed all of the above-mentioned validity and fidelity concerns.  As summarized by Little et al. (2009), "When using observations, care should be taken to select validated instruments and properly train and calibrate raters in order to obtain the most accurate results" (p. 7). To this end, Missouri's updated teacher evaluation model included the addition component that administrators undergo comprehensive training in the use of the evaluation instrument (MODESE, 2013).

**Teacher portfolios.**  Another alternative, or complimentary component to then-current teacher evaluation systems, was the teaching portfolio.  They could also be beneficial for evaluating teachers on student growth when standardized-test scores are not available.  Portfolios are a collection of materials for the purpose of providing evidence of a teacher's practice and student achievement (Little et al., 2009).  There was a wide range of materials, or artifacts, which may be included in a teacher's portfolio, including, but not limited to: lesson plans, assessments, student work samples, professional learning or coursework, and personal reflections (Berk, 2005; Little et al., 2009).   Portfolio document selection required careful consideration to appropriately represent the ongoing progress and processes that contributed to one's student achievement.  Furthermore, the artifacts selected should represent the entire learning process and document the student (or teacher's) improvement over time (Robelen, 2013).  According to Grissom and Youngs (2016), the use of teacher portfolios "were based in part on research on associations between student learning and teachers' use of formative assessment, their provision of feedback to students, and their knowledge of content-specific pedagogy" (p. 170).  These scholars also advocated that while time-consuming, the rewards of engaging in reflection on practice and compiling documentation through portfolios, was so valuable that all educators would benefit from his practice (Grissom & Youngs, 2016).  Grissom and Youngs (2016) claimed, "The use of such portfolios can help teachers determine which aspects of their planning, instruction, and assessment practices are in need of improvement" (p. 170).  Other educational experts supported the use of artifacts to complement other evaluative measures, due to their ability to provide evidence of teachers' practice that may not be readily apparent in the analysis of student achievement

data and classroom observations (Burnett, Cushing, & Bivona, 2012).  However, as endorsed by Berk (2005), the accumulation of artifacts through a teacher portfolio can be time-consuming and laborious, and may be best reserved for summative decision-making practices.

**Other measures.**  In addition to, or instead of student learning outcomes on standardized assessments, there were a number of potential sources of evidence of teaching effectiveness that were available for use in the evaluation process.  A partial list includes: feedback from students, parents, and colleagues; a teacher's self-reflection; classroom videos; and compiling a comprehensive teaching portfolio of artifacts (Berk, 2005; Goe & Croft, 2009; Little et al., 2009).  Each strategy could provide insight into a teacher's contributions to student learning, but with very different lenses (Goe & Croft, 2009).

**Applying multiple measures.**  Years of researchers have recognized that teaching is a complex and multifaceted endeavor, and acknowledges that as a result, the assessment of teacher performance requires the use of multiple measures (Darling-Hammond et al., 2013; Goe et al., 2008).  The research of Cohen and Goldhaber (2016) validated these findings and described a number of strengths and weaknesses associated with then-current methods of evaluating teacher performance.

The review of literature revealed that the majority of educational scholars suggested the combined use of a number of methods to create stronger indicators of effective teaching (Darling-Hammond, 2013; Goe & Holdheide, 2011; Hansen, Lemke, & Sorensen, 2013; Henry & Guthrie, 2016; Kane & Staiger, 2012).  Furthermore, researchers encouraged the use of a "combination of formative and summative measures

to inform both short and long term professional growth plans" (Burnett et al., 2012, p. 5).

Unfortunately, the burden of selecting and integrating these multiple measures into a

performance evaluation system with a single performance rating primarily fell on districts

(Hansen et al., 2013).  The selection of measures requires careful consideration of their

potential strengths and weaknesses, as well as consideration for their reliability, validity,

and feasibility (Darling-Hammond, 2010).  Regardless of the measures selected for use,

states and districts should consider all of the implications and ramifications of these

decisions.  The goal for districts would then be to determine a system (with the data

available) to "make the information available efficiently actionable" and to pinpoint

practices that each teacher could improve (Henry & Guthrie, 2016, p. 153).

As summarized by Goe and Holdheide (2011), multiple measures have the

potential to strengthen teacher evaluation, contribute to teachers' processional growth,

and set the stage for improved teaching and learning.

**Strength of measures.**  History and research confirmed that not all teacher

evaluation methods yield valuable or actionable information.   Moreover, each method

has strengths and weaknesses (Goe et al., 2014).

Esteemed educational historian, Rothstein, (2011) brings to light the importance

of "a balanced set of measures that are relatively unsusceptible to manipulation and

gaming" for establishment of successful teacher evaluation policies (p. 7).  Rothstein

(2011) also recommended that further research should evaluate "alternative teacher

evaluation *policies* rather than measures" (p. 7).

Even with more evidence, two factors will continue to complicate interpretation.

The first is that value-added measures are almost always bundled with other

measures, typically classroom observations.  This makes it hard to separate the

influence of value-added measures from other measures.  In addition to the mix of

measures, policies vary in how they incorporate these measures into personnel

decisions, and even *which* policies they are a part of.  Accountability varies in

both intensity and the types of decisions it can be designed to influence—tenure,

certification, compensation, promotion, and dismissal to name a few. (Harris &

Herrington, 2015, p. 74)

Harris and Herrington also warned, that since the use of VAMs were still in the

developmental phases, the dependency upon their results should be considered in

combination with other policy alternatives (Harris & Herrington, 2015, p. 74).

As a cautionary message, Grissom and Youngs (2016), proposed, "As currently

put into practice, we worry that measures employed in many school systems have

reliability and validity properties that are too questionable to be used for summative

evaluation and associated with high-stakes decisions" (p. 7).

However, for data from new teacher evaluation measures to be useful in

informing decisions concerning teacher advancement, pay, and dismissal, two

conditions must be met.  First, it is necessary to establish evidence of their

reliability and validity with regard to these particular uses.  Second, it is necessary

for teachers and administrators to feel confident that these new measures provide

accurate, stable ratings of teacher performance. (Grissom & Youngs, 2016, p.

171)

**Creating a Robust Teacher Evaluation System**

Goe et al. (2014), in collaboration with the Center on Great Teachers & Leaders, developed a *Practical Guide to Designing Comprehensive Teacher Evaluation Systems,* to serve as a tool for states and districts to use in developing new and robust teacher evaluation systems.  The authors proposed eight critical components for designing a comprehensive teacher evaluation system.  The components were:

> 1) Specifying evaluation system goals, 2) securing and sustaining stakeholder investment and cultivating a strategic communication plan, 3) selecting measures, 4) determining the structure of the evaluation system, 5) selecting and training evaluators, 6) ensuring data integrity and transparency, 7) using teacher evaluation results, and 8) evaluating the system. (p. 9)

A primary assumption of the development of then-new teacher evaluation systems, processes, and cycles was that they were clear to all stakeholders (administration, evaluators, and teaching staff) so that they understood the when, where, and how of teacher evaluations (The Education Trust, 2012).  Furthermore, Marzano, Toth, and Schooling (2012), emphasized that a common language/model of instruction should serve as the foundation of any performance evaluation system.  These sentiments were also reinforced by Darling-Hammond et al. (2012) and colleagues, who also emphasized the importance of beginning with common standards, building upon those standards to develop performance assessments, and creating the local evaluation systems and necessary support structures (Darling-Hammond, Cook, Jaquith, & Hamilton, 2010).

Of course, any system that evaluates teachers based on student learning must take into consideration what measures it uses for the accurate evaluation of teachers and how

that tool will serve in their efforts for continuous improvement (Hull, 2013a).  According to studies recent to this writing, the inclusion of student achievement measures in teacher evaluation was the most dramatic and controversial changes instituted to teacher evaluations and accountability measures (or educational reform as a whole) to date (Baker et al., 2010; Glazerman, Loeb, Goldhaber, Raudenbush, & Whitehurst, 2010; Hull, 2013a).

**Combining multiple measures into a common metric.**  As mentioned earlier in this literature review, rigorous teacher evaluation systems required the use of multiple measures and multiple data points to ensure accurate, reliable evaluation of teachers. Therefore, an additional — but necessary — undertaking of the evaluation system's design process was determining how to display the multiple measures of teacher performance and then combine them into a single score (Marzano & Toth, 2013).  For example, a district evaluation system that collected performance measures from three different sources (ex: VAM scores, observation scores, and surveys) must use these three measures to classify its teachers.  This framework was representative of many then-current evaluation systems (Hansen et al., 2013).  These authors identified three commonly used methods to combine multiple performance measures:  a numeric approach (where a teacher's summative effectiveness rating was a function of where a teacher falls in the distribution of their combined measure), a hybrid approach (where summative ratings were determined by categorizing teacher performance along each measure before combining, then rounding the overall score), and a profile approach (categorized teacher performance along each measure before combining and combined the measures) (2013, p. 4).

Several methods existed to accomplish this goal, such as using a weighting scheme to compete a composite growth score for teachers, or presenting an array of comparable scores to form an overall score for each teacher (Marzano & Toth, 2013, location 657). Measuring teacher growth is based upon the assumption that teacher growth "over time, should enhance the achievement of students" (Marzano & Toth, 2013, kindle location 1711). One simple method of determining relative teacher's growth would be to subtract a teacher's initial score on a specific element (from a domain) at the beginning of the year, to his or her score at the end of the year for that element (Marzano & Toth, 2013, location 1711).

These complicated statistical measures are important considerations for the design and implementation of the next generation of teacher evaluation systems and processes; however, they are outside the scope of this study.

**Rigorous training of evaluators**. A common theme identified within the literature was the importance of investing in high-quality training of staff, and in particular, the evaluators, as a method for establishing and maintaining fidelity of the system, and building educators' trust in the then-new processes (Fetters, 2013; Pennington, 2014; Shakman, Breslow et al., 2012). As Banks (2015) elucidated in The New Teacher Project blog, "When we look at how evaluation has fared from an implementation standpoint — what's changed on the ground — the picture isn't as rosy. Ratings inflation, inadequate training and norming, and low quality feedback re still major issues for many states" (para. 2).

In addition to clear standards and processes, a key element of high-quality teacher evaluation system involved the appropriate support structures for the practical

implications of these new systems (Darling-Hammond et al., 2012).  The authors

explained that these necessary support systems included, but were not limited to, ensuring

quality evaluator training, aligning professional learning opportunities to support the

improvement of teachers and instructional quality, mentoring for teachers needing

assistance, and providing the guidelines for making personnel decisions, and providing

the necessary resources for sustaining and monitoring the systems (Darling-Hammond et

al., 2012).   Others advocated for the use of multiple evaluators (from within or outside

the school) and suggested they provide on-going training and monitoring (fidelity checks)

to increase the reliability and objectivity of observations and evaluations (Fetters, 2013).

Regardless, consensus could be found among experts that the evaluators should be

knowledgeable about instruction and strategies, well trained in the evaluation

system/process, and how to give meaningful and actionable feedback to support

improvement and on-going learning and growth of teachers (Darling-Hammond, 2012).

Additionally, Darling-Hammond and peers recommended that "as often as possible, and

always at critical decision-making junctures (e.g., tenure or renewal), the evaluation team

should include experts in the specific teaching field (Darling-Hammond, et al., 2012, p.

iv).

Prioritizing the training of evaluators and all staff and stakeholders would require

a substantial investment of time and resources "to maximize implementation fidelity and

ensure that your evaluation system produces accurate, consistent, and legally defensible

results and ultimately can improve teaching practice" (Fetters, 2013, p. 4).

**Evaluating the system.**   In addition to ongoing support and resources directly

applied to the turnaround effort, an attempt to implement new teacher evaluation system

required sustained efforts toward continuous improvement through a process of systematic review and fine-tuning (Darling-Hammond & Snyder, 2015).  The research of Goe et al. (2014) also recommended regular systematic reviews of the evaluation process and results as an important part of the evaluation system vetting process.  Most likely, it would require modifications to the structure, processes, or format to ensure the system's efficacy and sustainability.  Furthermore, "for an accountability approach to be truly responsible for the outcomes our children deserve and our communities require, it must support a system that is cohesive, integrative and continuously renewing" (Darling-Hammond & Snyder, 2015, p. 3).  Further recommended is that district leaders reflect on the initial implementation with a critical insight.  As Banks (2015) advised,

> District leaders should take a step back and look at the rubrics they used during their initial evaluation overhaul, to make sure they are still in line with the district's values and vision for the evaluation system.  If they're not — or if they're not as clear or concise as they could be — it's worth the hard work of making changes and providing the training and support to implement those changes.  (para. 8)

Darling-Hammond and associates (2012), recommended establishing a panel of teachers and school leaders to develop, monitor, and ensure the evaluation implementation and support processes.  These authors also suggested this panel facilitate the personnel decisions to avoid potential litigation and ensure that it operated effectively and produced valid results (Darling-Hammond et al., 2012).

Researchers also conceded that, "Designing a comprehensive teacher evaluation system in an effective and sustainable manner is a difficult process, especially with few research-based models to consider" (Goe et al., 2014, p. 50).

**Challenges to Implementing new Teacher Evaluation Systems**

As redesigned teacher evaluation systems emerged across the country, studies recent to this writing only just began to examine their effectiveness, reliability, and validity.  There were a growing number of studies related to teacher evaluations, with the majority of the studies focused on the reliability of specific instruments (Donaldson, 2012; Donaldson & Papay, 2012; Riordan, Lacireno-Paquet, Shakman, Bocala, & Chang, 2015).  However, very few studies existed that documented the implementation of these executed systems or addressed their efficacy for achieving their desired results (McGuinn, 2012, 2015; Riordan et al., 2015; Shakman, Riordan et al., 2012).  Riordan, Lacireno-Paquet, N., Shakman, K., Bocala, C., & Chang (2015) emphasized the importance of studying implementation since local context could influence the outcomes and implementation could inevitably reshape practices and policies.

Nonetheless, these preliminary investigations of early-adopting districts suggested some factors that influenced the implementation processes.  McGuinn's (2012) study addressed the challenge that many local district leaders had with implementation due to concerns of state education agencies in supporting the human capacity needs and other struggles due to rapid implementation timelines, and questions regarding the use of student growth data in evaluations.  Shakman, Breslow, Kochanek, Riordan, & Haferd (2012) explained, "For many districts, state policymaking on educator evaluation systems

has mandated significant reform to existing systems, while leaving much of the detail of how to measure educator effectiveness to the districts themselves" (p. 4).

Researchers identified several district-level implementation challenges, including limited human capacity to facilitate evaluation systems, limited support and monitoring of new programs; inadequate time and attention spent on staff and evaluator training; and insufficient stakeholder support (Riordan et al., 2015).  Shakman, Breslow et al. (2012) also addressed this matter and resolved that effective implementation of these new and more rigorous teacher evaluations must address the organizational structures and processes involved to support new approaches.  "Successful change requires attention to the organizational systems, processes, and structures to ensure they support the intended reforms" (Shakman, Breslow et al., 2012, p. 8).

Another important consideration was that districts attempting to design and implement then-new teacher evaluation systems were doing so with minimal direction and support from state agencies (McGuinn, 2015).  Riordan et al. (2015) acknowledged the resulting strains this placed on local education agencies and also highlighted additional potential concerns, such as capacity challenges, lack of stakeholder support. All state agencies (and school districts) had "a unique history and operates in different fiscal, political, statutory, and constitutional context;" and therefore, varied in their roles and level of impact on education, and on how local agencies approached teacher evaluation within the state (McGuinn, 2012, p. 37).  In addition, McGuinn (2012) highlighted the difficulty of this daunting task thrust upon states and districts.

Experts also recognized the need for collaboration among educators across districts (and states) to learn from the accumulated wisdom of experiences (Darling-

Hammond, Wilhoit, & Pittenger, 2014).  The key lessons and challenges that emerged

from the experience of some early-adopting states would be beneficial for those that

followed (Steinberg & Donaldson, 2014).  Shakman, Breslow et al. (2012) explained

> building a community of practice among districts, along with a parallel effort to
>
> develop a body of research that follows districts' work, would provide much
>
> needed support to districts in their efforts to implement more rigorous and useful
>
> evaluation systems and improve teaching and learning for all. (p.24)

**Chapter Summary**

The literature concerning teacher evaluation revealed an evolution over time,

shifting strategies and emphasis in response to social and governmental influences.  This

evolution from teacher evaluations focused on quality assurance to one focused on

accountability, resulted in evaluation systems that served two functions — promoting the

development of teachers, and serving as an educational accountability function of the

twenty-first century.

Despite the proliferation of a variety of teacher evaluation strategies and

programs, few follow-up evaluations assessed the long-term impact or results of these

new programs.  Though outside the scope of this study, recommendations for further

research would include: establishing a clear definition of effective teaching and

determining a more sensitive and reliable tool for determining teacher impact on student

learning and achievement.

In summary, there is no simple system for evaluating the quality of teaching.  This

study intended to fill that gap by exploring both the long-term impact of one of the new

generation of teacher evaluation systems, based on Missouri's Model Evaluator system

(which was based on Danielson's (2011) *Frameworks of Teaching* model) developed and implemented within one school district.

While the literature was clear on the importance of designing multiple methods of evaluation to suit various purposes, the research proposed in this study sought to accomplish two goals.  First, determine the significance of the relationship between teacher quality, as defined within this teacher evaluation system, and student performance, as defined by two assessment metrics (MAP & i-Ready).  Second, the study explored the perceptions of teachers and administrators on the ability of this evaluation system to improve educator practice

## Chapter Three: Methodology

**Introduction & Problem**

Studies recent to this writing provided evidence supporting a correlation between teacher effectiveness and student achievement.  The research of several researchers established the quality of instruction provided by the teacher as the most important factor contributing to student achievement (Marzano, 2003; Rivkin et al., 2005; Hattie, 2009). While enlightening, this highlights the challenge for school districts to meet the responsibility of providing every student with a high quality and effective teacher, with the task of evaluations typically falling on the building principal.

Compelling studies also revealed that most evaluation systems do not adequately distinguish between effective and ineffective teachers (Baker et al., 2010).  As Daley and Kim (2010) explained, "Instructional practice varies among teachers in important ways, which in turn suggests that schools need ways to evaluate and improve the instructional practice of their teachers" (p. 3).  Others claimed teachers are consistently rated at the highest levels, even though evidence exists that suggest students are not performing at high academic levels (Weisberg et al., 2009; Daley & Kim, 2010).  According to Rothman (2009), "One of the best-kept secrets in educational research, it seems, is the fact that differences in the quality of instruction from classroom to classroom within schools are greater than differences in instructional quality between schools" (para. 1). An interesting observation noted in the literature was that despite the wide variance between teachers, the majority of teachers still received high evaluation marks (Donaldson & Peske, 2010; Jacob & Lefgren, 2008; The New Teacher Project, 2010, Weisberg et al., 2009).  In contrast, other studies concluded that teacher evaluation

ratings did have a relationship with student achievement (Jacob & Lefgren, 2008; Kane & Staiger, 2012; Kane et al., 2011a, 2011b; Stronge et al., 2011).

Educational researchers and historians suggested early school reform efforts increased public attention on teacher evaluations, and highlighted their potential for improving the quality of teaching, while they traditionally failed in that area (Daley & Kim, 2010; Darling-Hammond, 1990).  Daley and Kim's (2010) research cited the potential of teacher evaluations for the dual purposes of "personal growth and accountability," while acknowledging that previous evaluation procedures had not been designed for, or resulted in, improving educator practice (p. 5).

Despite the large number of studies conducted on an educator's impact on his/her students' achievement and on the ineffectiveness of traditional teacher evaluation methods for identifying/differentiating the quality of effective teachers, very few studies investigated the teacher evaluation programs amended recently before this writing.  This study attempts to fill that void.  Additionally, no studies looked specifically at the researched district's teacher evaluation tool.

**Background on Researched District's Teacher Evaluation System**

The researched school district designated a framework for professional practice based on the then-current research for promoting improved student learning.  Missouri's MMEES, based upon Danielson's (2011) FFT model, served as the framework for the researched district's teacher evaluation tool.   The researched district's teacher evaluation model established a common language and structure for professional conversations, provided a consistent agreed-upon understanding of teaching expectations, and established clearly-defined levels of excellence for all teachers of the district, from

novice to veteran, to which they were to hold themselves.  The district's teacher

evaluation manual also specified the procedures to determine a teacher's targeted goals

on his/her PGP.

The researched district stated, "The primary purpose of the Teacher Evaluation

system/process is to promote growth in effective practice that ultimately increases student

performance" (Teacher Evaluation System PowerPoint, 2015).  The district

administration implemented the then-new evaluation instrument in response to Essential

Principle of Effective Evaluation number four, "The use of measures of student growth in

learning;" however, the district's emphasis was specifically on "promoting growth in

effective practice that ultimately increases student performance" (Missouri School

District, 2015, p. 3).

The teacher evaluation model provided greater clarity on the expectations of

teachers and the evaluation criteria, allowing teachers to know expectations and to more

accurately demonstrate their knowledge.  The evaluation tool was aligned with

contemporary research on the Seven Essential Principals for creating an effective teacher

evaluation system (MODESE, 2013).  The then-new evaluation system made explicit the

specific requirements for evaluators to observe teachers more frequently (number of

observations based on the teacher's years of service) and provide more frequent and

consistent feedback through the use of Formal Feedback forms (number of formal

feedback forms received throughout the year also depended on the teacher's years of

service).  The evaluation tool evaluated teachers on the nine Professional Teaching

Standards, based upon the research by the CCSSO's Interstate Teacher Assessment and

Support Consortium (InTASC), and the creation was guided by the National Board for

Professional Teaching Standards (MODESE, 2013).   The nine Standards included in the

evaluation system were: (1) content knowledge aligned with appropriate instruction, (2)

student learning and development, (3) implementation of the curriculum, (4) instruction

on critical thinking skills, (5) classroom management and classroom environment, (6)

effective communication, (7) student assessment and data analysis, (8) self-assessment

and improvement, and (9) professional collaboration (Missouri School District, 2015).

The district's Teacher Evaluation Manual provided guidance through this instrument,

very different from establishment of a professional growth plan, through the process.

**Purpose**

The purpose of this mixed-methods study was to examine a potential relationship

between teachers' evaluation ratings, measured by the researched district's teacher

evaluation system, and the academic achievement of their students, as measured by MAP

and i-Ready assessments.  Thus, the study explored the validity of the teacher

performance rating as a measure of teacher effectiveness.  Additionally, the study

examined teachers and the evaluating administrators' perceptions of the impact of the

district's newly implemented evaluation process on the teachers' growth in practice and

impact on their students' learning.

Chapter Three will discuss the research design and methodology used to develop

this study conducted within a small rural Missouri school district.  Chapter Three also

presents the data collection and sampling procedures employed in the study.

This study was unique, because it was the first to focuses on the researched

district's self-created Teacher Evaluation Tool, as well as its impact on student learning

and teacher improvement.  Additionally, this study was unique because it focused on the

strength of the relationship between specific professional goals and strategies and impact on student academic growth, measured by reading and mathematics i-Ready and MAP assessments.

This study built upon the existing body of knowledge on the use of educator evaluation systems for measuring and improving the quality of teaching as related to student academic achievement.  This study may add to the growing body of research on the use of performance-based evaluations that measure teacher effectiveness using multiple measures, including observations, professional growth plans, artifact portfolios, and the use of student growth data, by documenting the results of a newly implemented evaluation system throughout six elementary schools.  The information from this study could result in modification to the researched district's then-new teacher evaluation system.

**Research Question**

The research question that guided the work of this dissertation was:

What are the teachers' and administrators' perceptions of the researched district's new teacher evaluation process (based on the MO Teacher Evaluation Model) as a method for improving professional practice (as measured by the Teacher Evaluation Rating tool) and influencing student achievement (as measured by the MAP and i-Ready assessments)?

This question addressed whether teachers and administrators bought-in and perceived value/effectiveness of the evaluation process and tools on their practice/professional growth and on student outcomes.

**Research Question and Null Hypotheses**

This mixed-methods study was guided by the following question and null hypotheses:

**Q1:** How do teachers and evaluating administrators perceive the district's teacher evaluation as having a potentially positive effect on improved teacher practice and professional growth?

**H10:** Student achievement as measured on the MAP and i-Ready grade level assessments did not improve after the teacher evaluation system was implemented.

**H20:** There is not a positive correlation between teacher performance ratings and student achievement on the MAP test.

The first and second hypotheses addressed the strength of the relationship between the teacher evaluation ratings and student achievement.  As Milanowski, Kimball, & White's (2004) research explained,

> Knowing whether this relationship is consistent from year to year is important in understanding the construct validity of the evaluation scores as well allowing users of these standards-based systems to assess whether the criterion-related validity of the evaluation scores is limited to specific teachers, students, and years.  (p. 5)

**Variables**

This study focused on seeking a relationship between the variables of teacher effectiveness ratings and student achievement.  The (independent) variable of teacher effectiveness was defined by teacher follow-up ratings on the teacher evaluation system's PGP component.  Due to the high number of staff members for each administrator to

evaluate, teachers were placed on a rotating summative report system. Therefore, the teacher evaluation follow-up scores were utilized, because they were the only rating that all teachers received every year. The (dependent) variable of student achievement was defined by students' MAP scores in reading and mathematics. For this study, student achievement was defined as students performing at grade level on the state standardized MAP assessment and the district i-Ready assessment. Specifically, for this study the researcher used the combined total percentage of students within each teacher's class scoring at the Proficient and Advanced performance levels on MAP assessment, and used the combined total percentage of students within each teacher's class that scored at the On Level and Above Level on the i-Ready assessment.

**Research Design**

A correlational study was used to determine the potential relationship between two variables: teacher ratings and student growth. A mixed-methods design, combining quantitative and qualitative approaches, was used to collect data for this study. However, in this study, priority was given to the quantitative data. The use of quantitative or qualitative data depended on the research question. The sources of quantitative data collection used in this study were: teacher Performance Ratings (district tool), student MAP scores (MODESE report), and student i-Ready scores (district report). The quantitative analyses were conducted to test the null hypotheses.

**$H_{01}$**: Student achievement as measured on the MAP and i-Ready grade level assessments did not improve after the teacher evaluation system was implemented.

**$H_{02}$**: There is not a positive correlation between teacher performance ratings and student achievement on the MAP test.

The sources of qualitative data used in this study included responses to teacher questionnaires (researcher created) and a small number of teacher interviews.  The questionnaires were used to gather data on the perceptions of the teachers, also represented the quantitative data of the study.  The qualitative data were gathered to answer the research question:

**Q$_1$:**  How do teachers and evaluating administrators perceive the district's teacher evaluation as having a potentially positive effect on improved teacher practice and professional growth?

**Population and Access**

The data source for this study was MODESE, which was the primary state entity responsible for the collection, analysis, and maintenance of educational data for the state of Missouri.  The setting for this research study was a rural Missouri school district.  The district had a population of 1,556 students between the third, fourth, and fifth grades, distributed among six elementary school buildings.  The distribution of students within grade levels and among the buildings varied upon the local geography, resulting in an unequal distribution of students among the six schools.

**Student data.**  The population for this study consisted of the researched district's third, fourth, and fifth grade general education students, who were assessed in the content areas of communication arts and mathematics by the MAP (State-mandated, standardized end-of-year assessments) and the i-Ready assessment (district-implemented, pre/post grade level assessments) during the 2013-2014, 2014-2015, and 2015-2016 school years.  The researched district had six elementary schools with a then-current enrollment total of

539 third grade students, 533 fourth grade students, and 484 fifth grade students.  The

total sample size for this study was 1,556 students and 45 classroom teachers.

**Teacher data.**  The population for this study also included the district's third,

fourth, and fifth general education grade teachers.  This study was limited to these

teachers, because they provided the curricular instruction for the students assessed by the

MAP and i-Ready assessments in the content areas of communication arts and

mathematics used this research.  All of the teachers comprised in this study were also a

part of the districts' then-new teacher evaluation process for both years.

**Evaluator data.**  The population for this study also included the elementary

school principals, who served as the evaluating administrators for the teachers comprised

in this study.  Each of the six elementary buildings employed one principal.  An

administrative intern served as an assistant to the principals, but divided their time

between two elementary buildings.  Due to the small sample size (nine administrators

total), the analysis of the survey data represented in this study is descriptive.

**Access.**  To begin the procedure for research, a research proposal was submitted

to the Lindenwood University Institutional Review Board (LU IRB).  Prospectus

approval was received on December 4, 2015 (see Appendix A).   The school district was

contacted to obtain permission to conduct the research and to access the district's data.

Permission was granted on August 3, 2015 (see Appendix B).  Following prospectus

approval, a research proposal was submitted to the LU IRB. LU IRB approval was

received on February 19, 2016 (see Appendix C).  Upon approval, data collection from

the targeted district began.  A third party district administrator provided the secondary

student assessment data and teacher evaluation data for the researcher, in order to protect

student and teacher privacy.  The administrator compiled a sheet (hard copy) with

randomly assigned student and teacher identifiers in lieu of names.  The researcher

received student MAP scores in communication arts and mathematics for the 2013-2014,

2014-2015 and 2015-2016 school years, and student i-Ready scores in reading and

mathematics for the 2014-2015 and 2015-2016 school years.  The researcher also

received teacher evaluation Initial and Follow-Up ratings for regular education classroom

teachers of grades three, four, and five.  The researcher used Excel to conduct the

computations and Survey Monkey to collect and manage the survey data.

**Data Collection and Analysis Procedures**

This research study involved archival student data compiled by MODESE, Data

from yearly MAP grade-level assessments and student growth data compiled by the

researched district using i-Ready beginning and end-of-the-year benchmarking

assessment tool.  This research also included secondary teacher evaluation data compiled

by the researched district resulting from instruments that were a portion of the district's

then-new teacher evaluation model.  A third-party district administrator collected and de-

identified all of the secondary data to protect the privacy of all individuals involved in

this study.  An online survey program, Survey Monkey, collected teacher and

administrator perception data to maintain subject anonymity.

First, the mean percentage of student MAP scores performing Proficient and

Advanced was determined for each classroom teacher, for each of the following school

years: 2013-2014, 2014-2015, and 2015-2015.  Next, the mean percentage of student i-

Ready scores performing On Level and Above Level was determined for each classroom

teacher, for each of the following years: 2014-2015 and 2015-2016.  The statistical

procedure used to analyze this data was a *t*-test for difference between dependent mean

percentages, for comparing these mean student percentages by teacher, from year-to-year,

in order to detect whether significant differences existed between.  A PPMCC test was

the statistical measure used to determine whether a relationship existed between the

study's two variables: Teacher Evaluation Follow-Up Ratings and Student Achievement

Levels on the MAP and i-Ready.  The Teacher Evaluation Follow-Up ratings were used

as the independent variable (x) and the student achievement measures were used as the

dependent variable (y).

The researcher applied two *t*-tests for difference in independent mean percentages

to explore the first hypothesis.  The researcher applied two PPMCC analyses to explore

the second hypothesis.

**Survey design process.**  The researcher constructed a customized survey to

gather data from the teachers and administrators in this study.  The researcher developed

the survey instrument in collaboration with the research chairperson, to address the

study's research questions.  The researcher asked colleagues and practitioners in the field

to review the survey questions to ensure that it accurately assessed what the researcher

was trying to measure.  Three teachers and three administrators reviewed the survey.

Based on these participants' feedback, the researcher made minor revisions to the survey

questions.  The same panel of professional colleagues reviewed the revised questions.

The researcher field-tested the survey tool with two teachers and two administrators to

experiment with the Survey Monkey's program (ease of use, confidentiality measures,

and how it managed and reported the data). The research committee approved the survey

instrument.  The researcher used open coding of teacher and administrator survey

responses to analyze the qualitative data.  Three other professionals also reviewed and analyzed the survey results.

**Research design**. This study employed a non-experimental correlational model to address the hypotheses and research questions, and to attempt to characterize the strength of association between the two variables — teacher rating scores and student academic achievement.  The study sought to compare two different student achievement measures by using archived student achievement score data from annual standardized end-of-grade-level assessments (MAP tests) and pre/post year benchmarking assessment data (i-Ready tests), with teacher evaluation ratings from the district teacher evaluation instrument.  This study compared the achievement of individual teachers' students from the year prior to implementation of the districts' new teacher evaluation system, and their students' achievement for the first two years of the then-new teacher evaluation system's implementation.  The study further sought to understand the views of teachers and their evaluators on the impact of the then-new teacher evaluation tool on the professional development and improvement of the teacher.  As previously mentioned, a strong correlation between these two variables could support the use of both methods in teacher evaluations.

Research commonly employed in education when seeking to find a relationship between two variables was of non-experimental design because of the practical challenges to conducting well-controlled experiments in educational settings (Fraenkel & Wallen, 2006; McMillan, Mohn, & Hammack, 2013).  Some of the challenges of educational research included the non-randomized assignment of students to classroom teachers, and other non-controllable factors, such as the variance in student

demographics, attendance rates, student motivation, and parental involvement (McMillan et al., 2013).  Due to the impossibility of controlling for these factors in the educational setting, the use of a non-experimental design was appropriate for this study.

A correlational design method was chosen, due to its ability to compare two different variables without having subjective bias. Some studies found that evaluators inflated or deflated evaluations to match previous test scores from specific teachers.  In most of the previous studies that evaluated teacher performance, student gain scores were compared with the traditional method of observation (Milanowski et al., 2004; Taylor & Tyler, 2012), or compared teacher value-added scores to observational scores (Kane & Staiger, 2008; Rockhoff & Speroni, 2010).  However, these previous studies struggled to find a correlation, due to the observed subjective bias by administrators in their observational evaluations (Cohen & Goldhaber, 2016).  Therefore, this study sought to avoid the subjective bias observed in previous studies by comparing two different objective measures of student achievement: student growth scores from the MAP and i-Ready assessments and teacher rating scores (teacher evaluation ratings), which were determined using a rubric and multiple measures of teacher improvement (student growth data).

While this study utilized a mixed-methods design, combining quantitative and qualitative approaches to collect data, the quantitative data took priority over the qualitative data.  The hypotheses and research questions guided the quantitative and qualitative data respectively.

**Dependent variable.**  The dependent variable (y) for this study was student achievement.  For the purpose of this study, the researcher defined student achievement

as students performing Proficient or Advanced on the MAP grade-level assessments and

On Level or Above Level on the i-Ready benchmark assessments during the 2014-2015

and 2015-2016 school years.

**Independent variable.** The independent variable (x) for this study was the

teacher rating scores on the district's teacher evaluation instrument.  An overall mean

score was determined for each teacher by first developing a mean score based on the

MAP and i-Ready assessment scale scores earned by the teacher's students during the

2015-2016 school year.  Next, the researcher calculated a mean score for each teacher,

based on the MAP assessment scale scores for the 2014-2015 school year.  This measure

was selected because it takes into account variation within the scores.  This measure also

takes into account the comparison of groups of students on two different assessments as

suggested by Schagen and Hodgen (2009).   While these considerations were not the

focus of this research study, the use of a calculated mean score as the independent

variable was still applicable.

A strong correlation between these two variables would support the validity of

both evaluation methods as a form of triangulation.  Triangulation was the process of

using different types of data and/or using different methods to examine the same research

question (Fraenkel & Wallen, 2006).  In this current study, the researcher employed

triangulation to study the same research questions using different methods and types of

data.  The researcher compared all collected data sets in an effort to determine a

relationship between archival student MAP assessment data in the content areas of

communication arts and mathematics, archival student i-Ready assessment data in the

content areas of communication arts and mathematics, teacher evaluation ratings

(collaboratively determined by the teacher and evaluator), and teacher and evaluator survey data of the same teacher participants' improvement in instructional practice.

The state of Missouri began the mandated standardized student assessments in grade three.  Therefore, third grade was the earliest tested data in the series of the three-year trend data considered for this study.  For this reason, student growth percentiles could not be determined for students and define the benchmark for the subsequent grade assessments.

**Instrumentation**

This section provides a description of all the methods of instrumentation employed in this current research study, which measured both teacher performance and student performance.

**MAP.**  The researched district's Human Resources director provided student achievement data from the Missouri Assessment Program (MAP) standardized assessment, which was administered during the spring of the 2014, 2015, and 2016 school years.  The MAP assessed "student progress toward mastery of the Show-Me Standards which are the educational standards in Missouri.  The Grade-Level Assessment is a yearly standards-based test that measures specific skills defined for each grade by the state of Missouri" (MODESE, 2013, para. 1).  MODESE's testing vendor was Data Recognition Corporation (DRC), who administered, scored, and reported all grade-level assessments.  The analysis included results from communication arts and mathematics. The measure of student achievement was based on the percentage of a teacher's students reaching the Proficient and Advanced performance levels, as indicated on the MAP standardized assessments.

School districts and state education departments collected and archived student standardized test data after each cycle of annual testing.  MODESE archived test data. Data for this study originated from the MODESE's core data reports published and distributed to the administration of the researched school district.  Only student MAP in communication arts and mathematics scores were required for analysis.  Therefore, this study did not include the collection of teacher or student identifiers.

For the purposes of this study, student achievement was defined as the amount of measurable growth students demonstrated on the MAP grade-level assessments during the 2013-2014, 2014-2015, and 2015-2016 school years.  According to MODESE, the Missouri MAP used scale scores that were unique to Missouri.  "The characteristic growth seen on the scale from grade to grade for the standardized test has been utilized and built upon to give the MAP its vertical scale characteristics" (MODESE, 2014, p. 72).  This vertical scale was also referred to as a student growth scale or growth percentile.  Missouri was unique because it used both a student growth and teacher value-added model, or a Student Percentile Growth (SPG) model and a Value Added Model (VAM).  Beginning in grade three, a baseline score was established, based upon the student's first MAP administration.  Missouri's 2008 application for NCLB stated, "Growth targets remain constant from the students' baseline year through the next four years (or the end of grade 8), whichever comes first" (USDOE, 2008, p. 7).  Grade level MAP tests were analyzed and found to produce consistent and reliable results (MODESE, 2014).

**i-Ready.**  The achievement results from the i-Ready grade level pre- and post-assessments provided a second indicator of student academic success in the researched

school district.  The i-Ready Diagnostic was an adaptive assessment used to provide

"valid and reliable growth metrics across a district or school environment" (Curriculum

Associates, 2014, p. 4).  The i-Ready Diagnostic was extensively researched and found to

be highly correlated with other state assessments as a valid and consistent predictor of

student proficiency rates on Common Core Assessments (Curriculum Associates, 2014).

The researched school district used the i-Ready Diagnostic computer-adaptive assessment

to measure student growth from the beginning-to-the-end of each school year for grades

K through 8.  Student growth was determined by measuring the difference in a student's

score from the pre-test administered at the beginning of the year (August/September) to

the post-test, administered at the end of the year (April/May).  Like the MAP, the i-Ready

assessment used a vertical scale for scoring to measure which skills a student gained from

one point in time to the next.  i-Ready creators advocated that vertical scale provided a

consistent metric for measuring and comparing student progress across grade levels

(Curriculum Associates, 2014, p. 6).

**District Teacher Evaluation Tool.**  In 2013 Missouri launched the Missouri

Educator Evaluation Model to meet with federal guidelines of NCLB and the then-current

theory of essential principles of effective evaluation (MODESE, 2013).  In compliance,

the researched school district began the planning stages of overhauling its teacher

evaluation program to align with Missouri's Educator Evaluation Model in 2013.  The

newly developed teacher evaluation was the instrument used to measure teacher

performance and was appropriate for the study's population and setting.  All teachers in

the selected school district participated in this teacher evaluation instrument, based on

pre-determined cycles.  Formal and/or informal teacher evaluations were conducted,

based on annual cycles over a three-year period (depending on teachers' years of

experience).  With new/non-tenured teachers receiving yearly comprehensive, summative

reviews (regardless of experience), teachers from years six through 14 receiving

comprehensive, summative evaluations every other year, and seasoned educators with

15+ years of teaching receiving comprehensive summative evaluations every three years.

Completion of a comprehensive summative evaluation took place during the teacher's

final year in the cycle.

The evaluation process included frequent observations (to provide formative

feedback to teachers for making instructional modifications and to monitor student

achievement data) during every year of the cycle.  For the observational component,

evaluators used a district-created rubric based on the Danielson's (2011) FFT model,

which described performance of each skill and practice at four levels:  Distinguished,

Proficient, Developing, and Emerging.

While teachers were accountable for each of the nine performance standards, they

were responsible for demonstrating growth in practice on just the two standards/

indicators specified in the PGP.  The professional growth was measured over the cycle

using a rubric, based on the rating scale, which included descriptors of Distinguished,

Proficient, Developing, and Emerging.  During the summative year of the cycle, a more

comprehensive evaluation of a teacher's performance was measured using a rubric, based

on the rating scale: No Concern, Possible Growth Opportunity, or Area of Concern.  The

researched district's Teacher Evaluation instrument detailed the performance

expectations for teachers for each standard and provided a general description of what a

rating entailed (Missouri School District, 2015).

The summative teacher evaluation instrument included the collection of data from four sources:  Areas of Concern (AOC) marked on the Summative Evaluation Feedback Form, the Final Indicator Rating (FIR) from the teacher's PGP form, the teacher Growth Factor Rating (GFA) from the teacher's PGP, and the teacher's Student Growth Measure (SGM) percentile determined from the assessment designated on the PGP to measure student growth and attainment levels (Missouri School District, 2015).   This process ensured input from multiple sources.

During the PGP process, each teacher set two personal growth goals.  These goals aligned on the district goal of differentiated instruction and one of his or her choosing on any of the nine standards.  At the building level, teacher teams set student achievement goals, that aligned with these personal growth goals.  These goals were based on student needs and projected growth potential assessed by looking at students' baseline achievement data.   Additionally, teachers collected artifacts to serve as a portfolio as part of the summative evaluation process.  The purpose of the portfolio was to collect pertinent data to demonstrate that each teacher was making contributions toward student growth and his or her own professional growth over the two-to-three-year cycle.

Finally, evaluators collected data via classroom observations and follow-up conferences with teachers.  A minimum of two-to-five opportunities for formal and/or informal feedback from evaluating administrators to teachers was required every year. The specific number was dependent upon the teacher's years of service (Missouri School District, 2015).  However, building principals were encouraged to frequent classrooms as much as possible, and more than the minimum number of feedback forms was encouraged.

The district's Human Resources officer provided the teacher evaluation rating data for the two years comprised in this study.  The district's Teacher Evaluation System had four basic components based upon targeted indicators in Teacher PGPs; multiple administrator observations employing a standards-based rubric for giving teacher feedback (Formative Feedback forms), teacher created portfolios of collected artifacts, and summative evaluation rubric and follow-up rating tool.

To accommodate the arduous task of reallocating all teachers to the new summative evaluation cycle, the district devised a phase-in schedule during the program's planning stages.  This schedule set the year of the first evaluation, based upon the number of years of teaching service within the district.  Teachers with one-to-five years of teaching experience in the district received a summative evaluation every year, while teachers with six-or-more years of teaching within the district received a summative evaluation every other year (Missouri School District, 2015).  While the district attempted to produce more robust observations and meaningful evaluations, limited resources increased the demand and workload placed upon the small number of administrators within the district, thus creating a particular challenge.

Both teachers and evaluators completed two-hour overview and training sessions during the back-to-school staff meetings.   Teacher representatives from the New Teacher Evaluation Committee, who volunteered to present the manual and new expectations for the then-new teacher evaluation process, facilitated these trainings.   The training emphasized the district's primary purpose of the Teacher Evaluation process to "promote growth in effective practice that ultimately increases student performance" (Teacher Evaluation System PowerPoint, 2014, page 3).  The training session consisted of a

District-created PowerPoint presentation explaining the process and communicating the

District's new process (Teacher Evaluation System PowerPoint, 2014), an opportunity to

ask clarifying questions which the trainer and building administrator fielded, and an

opportunity to review the new evaluation manual, forms, rubrics, and other related

documents.  Any questions the facilitators could not answer were directed to the

District's Human Resources Department.  Teachers then met with the building

administrator to set up individual PGPs.

*Professional growth plans*.  Each staff member was required to create a yearly

PGP, which addressed the goals and strategies for professional growth and learning

(Missouri School District, 2015).  Individual Teacher PGPs specified the specific

teaching standard, and sub-indicator, for which each teacher would be responsible for

showing progress on by the conclusion of the summative evaluation cycle.   Teachers and

building administrators worked in collaboration to determine the standard/indicator,

establish a baseline score for each, and to specify the strategies, action steps, and timeline

for completion.   The baseline score was determined by reviewing the Growth Guide

Level Descriptors, and determining the appropriate score on the range of proficiency

scale, 1 to 8   (Missouri School District, 2015).

In the 2014-2015 school year, the researched district launched a new teacher

evaluation system based upon the Missouri Model Evaluator (MODESE, 2013). The

purpose of this program was to evaluate the district's teaching staff's performance in-

and-out-of the classroom through multiple measures including: classroom observations, a

review of student growth measures on pre-determined assessments, and a review of work

products (artifacts) complied by the teacher.

***Classroom observations***. During the yearlong evaluation process, teachers were typically observed three-to-five times by an assigned school administrator (Missouri School District, 2015).  These observations could be of varying lengths of time.  Some observational drop ins were only five minutes long.  The majority of observations were unannounced.  However, teachers were encouraged to invite the administrator into his/her classroom to observe specific activities that demonstrated the teacher's implementation of one of his/her targeted standard on the PGP.  Additionally, principals could request/ schedule a formal or announced observation.

After each observation, administrators provided written feedback to the teacher using an online employee management system called Talent Ed.  Teachers and administrators were encouraged to meet frequently to discuss observations and the feedback provided by the administrator/evaluator. At the end of the evaluation school year, a final summative score was calculated by collaboration between the teacher and the administrator.  These final scores carried explicit consequences.  For new teachers (those within the induction stage of teaching in the district), a poor evaluation could result in a Professional Improvement Plan (PIP), which focused on intervention strategies for areas of concern (Missouri School District, 2015).  Successful evaluation could determine a teacher's eligibility for tenure status (or protection) within the district.  For already tenured teachers, poor evaluation scores could place the employee on a Professional Improvement Plan (PIP), or other employee assistance program, with a small risk for termination.  However, for tenured teachers receiving high evaluation scores, administrators could determine the employees' eligibility for professional advancement

Despite the training and detailed rubric provided to evaluators, the district teacher evaluation system did experience some leniency bias that was typical of other teacher evaluation programs.

**Sampling**

This current study employed a convenience sample, due to the small size.  The sample for this study included all third, fourth, and fifth grade students assessed by the MAP and i-Ready assessments in one medium-sized Missouri school district.  Since third grade was the first high stakes grade, there was no way to determine a growth score. Therefore, each teacher had a cluster of student scores in communication arts and mathematics, from which teacher effectiveness was calculated.  A mean score was calculated for each teacher using MAP test scores (in communication arts and mathematics) and i-Ready scores (in communication arts and mathematics).  The only foreseeable students with missing data would be in cases where attrition occurred.

An opportunity sample including all of the district's third, fourth, and fifth grade teachers of students assessed by the MAP and i-Ready assessments in the content areas of communication arts and mathematics were analyzed in this current study.  All of the teachers comprised in this study were also a part of the districts' then-new teacher evaluation process for the 2014-2015 and 2015-2016 school years.  The study utilized a sample size of 50 to 60 teachers, which was supported by the work of Fraenkel and Wallen (2006), based on the researcher's access.  The only foreseeable teachers with missing data would be in cases where teachers transferred to a different grade level or attrition occurred.

The qualitative portion of this study included teacher perception data. Participants in the surveys were general education elementary school teachers employed by the researched school district, with students assessed by both the MAP test and i-Ready test in both content areas of English language arts and mathematics.  The criteria for inclusion in the data included: participant was employed as a general education third, fourth, or fifth grade teacher during both years of the study (2014-2015 and 2015-2016); participants participated in the district's teacher evaluation system for both years of the study; and the participant taught the same grade level for both years comprised in this study.

The qualitative portion also included perception data of the evaluating administrators.  Participation in the administrator survey was voluntary.  Due to the small number of participants available (only nine total employed in the district), the data analysis performed was limited to descriptive statistics to explore emerging themes.

**Confidentiality**

For the secondary data portion of this study, the district's Director of Human Resources officer removed all personal identifiers from the student assessment data and teacher evaluation data before supplying the researcher with the data.  The researcher only knew the grade level of the students and the teachers when analyzing the data.  The researcher received a hard copy of this data, with personal identifiers removed for analysis.  The researcher entered the data into an Excel spreadsheet for analysis.  The information was saved on the researcher's personal computer.

For the primary data used in this study, the researcher contacted all of the district's general education third, fourth, and fifth grade classroom teachers of MAP-

tested content areas of communication arts and mathematics through district email to invite them to participate in the study, through the web-based survey manager program, Survey Monkey.  This invitation also informed participants of the confidentiality and gained informed consent (see Appendix D).  The web-based Survey Monkey program anonymously collected the data and delivered the results to the researcher for analysis.

Finally, no personal identifiers were be included in the research upon publication, nor would any information be attributable to any individual or the district be released.

**Limitations & Delimitations of the Study**

The objective of this study was to determine a relationship between student achievement growth (on the MAP and i-Ready assessments) and teacher performance ratings.  The focus of this study was on mathematics and English language arts (interchangeable with communication arts), since those were the content areas consistently assessed across the elementary grade levels.  This study compared MAP and i-Ready scores within a class based on archival data.  The focus of this study was teacher effectiveness, as research consistently found that the effectiveness of the teacher had the greatest impact on student achievement in these content areas.

The inherent nature of MAP data was a limitation of this study.  The MAP, like most assessments "often assume that [student achievement] scores are a direct and ambiguous measure of student achievement" (Koretz, 2000, p. 4).  Another complication to the use of the Missouri Assessment Program was, since 2008 the assessment program had undergone a period of transition that resulted in substantial changes to implementation.  In the 2014-2015 school year, Missouri schools administered the Smarter Balanced Interim Assessment, developed by the Smarter Balanced Assessment

Consortium, which aligned with the Common Core State Standards (MODESE, 2013).

However, after only one year, another revised assessment was implemented, again during

the 2014-2015 school year.  Not only did the 2014-2015 assessment test students over

new content standards in English language arts and Mathematics, the assessments were

also administered on computers through an online program for the first time.  The result

of this turbulence is the state of Missouri will have administered four different testing

systems in four years by 2017, when the new assessments were scheduled to be

implemented.  These changes also made it difficult for Missouri districts to compare their

growth over multiple years of data.  Instead, Missouri would only be able to compare its

performance from one year to the next.

  The lack of random student assignment was another limitation of the study.

Several researchers observed the non-random assignment of both teachers and students to

classrooms.  These experts propositioned that this fact posed a significant challenge for

teacher evaluation models (based on either statistical measures or purely observational

measures) to adequately separate a teacher's impact from all of the other factors

impacting a student's academic success (Darling-Hammond et al., 2013; Ladd, 2008).

According to Paufler and Amrein-Beardsley's 2013 study, these biases were

demonstrated most prevalently by the instability of value-added measures of teacher

effectiveness from year-to-year, and even from test-to-test (Paufler & Amrein-Beardsley,

2013).

**Assumptions of the study**

The first assumption of this study was that the distribution of students among the classrooms was relatively equal. (But they were NOT). The study also assumed that the MAP and i-Ready scores had a normal distribution and followed a standard bell curve.

**Conclusion**

This study included teachers employed within the researched district during the 2013-2014, 2014-2015 and 2015-2016 school years, evaluated using the new Teacher Evaluation System.  The study further focused primarily on those who were teaching the MAP assessed grades, third through fifth grades, during those years.  Additionally, the study focused on just the content areas of communication arts and mathematics.  For most other subjects, student achievement measures were not available.  The study combined the Teacher Evaluation System data (teacher follow-up ratings) and the teachers' student achievement data, both provided by the researched district, which allowed the researcher to match teachers to their students' achievement.

This chapter summarized the mixed methods research design used in this study to examine the relationship between teacher evaluation ratings on the researched districts' teacher evaluation tool and student achievement at the elementary level in a rural Missouri school district.  In addition, the chapter explained the population and data collection techniques used in the study.  The data utilized teacher evaluation ratings during the 2014-2015 and 2015-2016 school years, student MAP and i-Ready assessment data from 2013 through 2016, and teacher and administrator perception data from 2016. The chapter concluded with a synopsis of the instruments used in both the quantitative and qualitative components of the study.

The results of the data analysis is revealed in Chapter Four.   Conclusions,

implications for practice, and recommendations for future research are presented in

Chapter Five.

## Chapter Four: Results

Chapter Four presents the results of the study into two sub-sections: analysis of the quantitative data, and analysis of the qualitative data.   The qualitative analysis will enhance the interpretation of the quantitative results.  The qualitative analysis will include thematic categories developed throughout the data analysis process.

### Research Question and Null Hypotheses

This mixed-methods study was guided by the following question and null hypothesis:

**Q1:**  How do teachers and evaluating administrators perceive the district's teacher evaluation as having a potentially positive effect on improved teacher practice and professional growth?

**H10:**  Student achievement as measured on the MAP and i-Ready grade level assessments did not improve after the teacher evaluation system was implemented.

**H20:**  There is not a positive correlation between teacher performance ratings and student achievement on the MAP test.

### Results of Quantitative Data

**Null hypothesis one.**  Student achievement as measured on the MAP and i-Ready grade level assessments did not improve after the teacher evaluation system was implemented.  Null hypothesis one involved examination of student performance on the MAP English language arts and mathematics assessments from one year prior to implementation of the new Teacher Evaluation system (2012-2014 school year) with student performance on the MAP English language arts and mathematics assessments over the next two years after implementation of the new Teacher Evaluation system

(2014-2015 and 2015-2015 school years). Table 1 represents the data used to support the

outcome of hypothesis one.

Table 1

*MAP Data*

Mean of Students On or Above Grade Level

| | 2014 Results (x-bar) | 2015 Results (x-bar) | Change +/- Over First Year | 2016 Results | Change +/- Over Second Year |
|---|---|---|---|---|---|
| ELA | 53.23 (n=46) | 64.13 (n=46) | +10.90 | 69.18 (n=43) | +15.95 |
| Math | 57.71 (n=46) | 57.34 (n=46) | -0.37 | 65.95 (n=43) | +8.24 |

*Note*:  From MODESE, 2014, 2015, 2016

A two-sample *t*-test for difference in percentage was conducted comparing the

percentage of students scoring on grade level (combined total of students at the Proficient

and Advanced achievement levels) on the English language arts MAP assessment in the

year 2013-2014 (the year prior to implementation of the new Teacher Evaluation system)

and the percentage of students scoring on grade level (combined total of students at the

Proficient and Advanced achievement levels) on the English language arts MAP

assessment in the year 2014-2015 (the first year after the Teacher Evaluation system was

implemented).  The null hypothesis was rejected; the analysis revealed a significant

difference between the students' 2013-2014 English language arts MAP scores and the

students' 2014-2015 MAP scores, $t(44) =$  4.475, $p < .0001$ (*t*-critical $= 1.645$, $\alpha = .05$).

This suggested that student MAP achievement levels significantly improved in English

language arts after the first year of implementation of the then-new teacher evaluation

system.

A two-sample *t*-test for difference in percentage was conducted comparing the

percentage of students scoring on grade level (combined total of students at the Proficient

and Advanced achievement levels) on the English language arts MAP assessment in the

year 2014-2015 (the year after implementation of the new Teacher Evaluation system)

and the percentage of students scoring on grade level (combined total of students at the

Proficient and Advanced achievement levels) on the English language arts MAP

assessment in the year 2015-2016 (the second year after the Teacher Evaluation system

was implemented).  The null hypothesis was rejected; the analysis revealed a significant

difference between the 2014-2015 English language arts MAP scores and the students'

2015-2016 MAP scores, $t(42) = 2.342$, $p = 0.012$ ($t$-critical = 1.645, $\alpha = .05$).  This

suggested that student MAP achievement levels significantly improved again in English

language arts after the second year of implementation of the then-new teacher evaluation

system.

A two-sample $t$-test for difference in percentage was conducted comparing the

percentage of students scoring on grade level (combined total of students at the Proficient

and Advanced achievement levels) on the mathematics MAP assessment in the year

2013-2014 (the year prior to implementation of the new Teacher Evaluation system) and

the percentage of students scoring on grade level (combined total of students at the

Proficient and Advanced achievement levels) on the mathematics MAP assessment in the

year 2014-2015 (the first year after the Teacher Evaluation system was implemented).

The null hypothesis was not rejected; there was not a significant difference between the

students' 2013-2014 mathematics MAP scores and the students' 2014-2015 MAP scores,

$t(44) = 0.142$, $p = 0.5563$ ($t$-critical = 1.645, $\alpha = .05$). This suggested there was not

enough evidence to conclude that student MAP achievement levels improved in

mathematics after the first year of implementation of the then-new teacher evaluation system.

A two-sample *t*-test for difference in percentage was conducted comparing the percentage of students scoring on grade level (combined total of students at the Proficient and Advanced achievement levels) on the mathematics MAP assessment in the year 2014-2015 (the year after implementation of the new Teacher Evaluation system) and the percentage of students scoring on grade level (combined total of students at the Proficient and Advanced achievement levels) on the mathematics MAP assessment in the year 2015-2016 (the second year after the new Teacher Evaluation system was implemented). The null hypothesis was rejected; the analysis revealed a significant difference between the students' 2014-2015 mathematics MAP scores and the students' 2015-2016 MAP scores, $t(42) = 3.578$, $p = 0.0004$ (*t*-critical $= 1.645$, $\alpha = .05$).  This suggested that student MAP achievement levels significantly improved in mathematics after the second year of implementation of the then-new teacher evaluation system.

Table 2 represents the data from null hypothesis one, which examined the correlation between student performance on the i-Ready reading and mathematics assessments from the first year of the new Teacher Evaluation system implementation (2014-2015 school year) and student performance on the i-Ready reading and mathematics assessments from the second year of the new Teacher Evaluation system's implementation (2015-2016 school year).

Table 2

*i-Ready Data*

| Mean of Students On or Above Grade Level | | | |
| --- | --- | --- | --- |
| | 2015 Results (x-bar) | 2016 Results (x-bar) | Change +/- After Second Year |
| ELA | 43.30 (n=46) | 44.41 (n=43) | +1.11 |
| Math | 58.5 (n=46) | 61.81 (n=43) | +3.31 |

*Note*: From i-Ready, 2015, 2016

A two-sample *t*-test for difference in percentage was conducted comparing the percentage of students scoring on grade level (combined total of students at the On and Above achievement levels) on the i-Ready reading assessment in the year 2014-2015 (the first year prior of the new Teacher Evaluation system's implementation) and the percentage of students scoring on grade level (combined total of students at the On and Above achievement levels) on the i-Ready reading assessment in the year 2015-2016 (the second year of the new Teacher Evaluation system's implementation).  The null hypothesis was not rejected; the analysis revealed no significant difference between the students' 2014-2015 i-Ready reading scores and the students' 2015-2016 i-Ready reading scores, $t(44) = 0.424$, $p = .3367$ (*t*-critical $= 1.645$, $\alpha = .05$).  This suggested there was insufficient data to conclude that student i-Ready achievement levels in reading improved over the first two years of the then-new teacher evaluation system's implementation.

A two-sample *t*-test for difference in percentage was conducted comparing the percentage of students scoring on grade level (combined total of students at the On and Above achievement levels) on the i-Ready mathematics assessment in the year 2014-2015 (the first year prior of the new Teacher Evaluation system's implementation) and

the percentage of students scoring on grade level (combined total of students at the On

and Above achievement levels) on the i-Ready mathematics assessment in the year 2015-

2016 (the second year of the new Teacher Evaluation system's implementation).  The

null hypothesis was not rejected; the analysis revealed no significant difference between

the students' 2014-2015 i-Ready mathematics scores and the students' 2015-2016 i-

Ready mathematics scores, $t(41) = 0.985$, $p = .1651$ ($t$-critical $= 1.645$, $\alpha = .05$).  This

suggested there was insufficient evidence to conclude that student i-Ready achievement

levels in Math improved over the first two years of the then-new teacher evaluation

system's implementation.

Null hypothesis one also involved examination of the implementation of the

researched district's new teacher evaluation system in the 2014-2015 and 2015-2016

school years and archival student assessment data MAP and i-Ready grade level as

generated by MODESE for the 2014-2015 and 2015-2016 school years.

Table 3 represents the data from hypothesis one which examined the correlation

between students' MAP English language arts performance for the 2014-2015 school

year and students' i-Ready reading performance for the 2014-2015 school year.

Table 3

*Results of Pearson Product Moment Correlation Coefficient for Student MAP
Performance and Student i-Ready Performance 2014-2015 School Year*

| | MAP ELA Performance | | i-Ready Reading Performance | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Results | *M* | *SD* | *M* | *SD* | *n* | *r* | *t* | *df* | *p* |
| | 64.13 | 11.35 | 43.30 | 12.85 | 46 | 0.428 | 3.105 | 44 | 0.0034 |

*Note.*  Statistical significance is noted at $p \le .05$.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean students' performance on the MAP English language arts assessment (*M* = 64.13, *SD* = 11.35) and the mean student performance on the i-Ready reading assessment (*M* = 43.30, *SD* = 12.85) for the 2014-2015 school year. The analysis revealed a significant, moderate, positive correlation between the MAP English language arts scores and the i-Ready reading scores for the 2014-2015 school year, *r*(44) = 0.428, *p* =0.0034 (*r*-critical = 0.288, *α* = .05). This suggested there was a relationship between these two variables.

Table 4 represents data from hypothesis one which examined the correlation between students' MAP mathematics performance for the 2014-2015 school year and students' i-Ready mathematics performance for the 2014-2015 school year.

Table 4

*Results of Pearson Product-Moment Correlation Coefficient for Student MAP Performance and Student i-Ready Performance 2014-2015 School Year*

| | MAP Math Performance | | i-Ready Math Performance | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Outcome | *M* | *SD* | *M* | *SD* | *n* | *r* | *t* | *df* | *p* |
| | 57.34 | 16.98 | 58.5 | 13.06 | 46 | 0.552 | 4.391 | 44 | 0.0001 |

*Note.*  Statistical significance is noted at *p* ≤ .05.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean students' performance on the MAP mathematics assessment (*M* = 57.34, *SD* = 16.98) and the mean student performance on the i-Ready mathematics assessment (*M* = 58.5, *SD* = 13.06) for the 2014-2015 school year.  The analysis revealed a significant, moderate,, positive correlation between the MAP mathematics scores and the i-Ready mathematics scores for the 2014-2015 school year, *r*(44) = 0.552, *p* =.0001 (*r*-critical = 0.288, *α* = .05). This suggested there was a relationship between these two variables.

Overall, the analysis revealed a significant relationship between MAP mathematics scores and i-Ready mathematics scores. This suggested increases in i-Ready mathematics scores were correlated with increases in MAP mathematics scores.

**Null hypothesis two.** There is not a positive correlation between teacher performance ratings and student achievement on the MAP test. Null hypothesis two involved examination of the correlation between the teachers' follow-up ratings on teacher evaluations generated by building administrators for the first two years of implementation and their students' MAP and i-Ready assessment data for the same two years' school years.

Table 5 displays the data examining the potential correlation between teacher's follow-up ratings for the 2014-2015 school year and student MAP English language arts assessment data for the 2014-2015 school year.

Table 5

*Results of Pearson Product-Moment Correlation Coefficient for Teacher Follow-Up Ratings and Student MAP ELA Performance 2014-2015 School Year*

| Outcome | Teacher Follow-Up Ratings | | MAP ELA Performance | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *n* | *r* | *t* | *df* | *p* |
| | 4.39 | 1.12 | 64.13 | 11.35 | 46 | -0.096 | -0.640 | 44 | 0.5257 |

*Note.* Statistical significance is noted at $p \le .05$.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean Teacher Follow-up Ratings on teacher evaluations (*M* = 4.39, *SD* = 1.12) for the 2014-2015 school year and the mean student performance on the MAP English language arts assessment (*M* = 64.13, *SD* = 11.35) for the 2014-2015 school year. The null hypothesis was not rejected; the analysis revealed no positive correlation between teacher Follow-Up Ratings and student MAP English language arts scores for the 2014-2015 school year,

$r(44) = -0.096$, $p = 0.5257$ ($r$-critical $= 0.288$, $\alpha = .05$).  This suggested there was not a

relationship between these two variables.

Table 6 displays the data examining the potential correlation between teacher's

follow-up ratings for the 2014-2015 school year and student i-Ready reading assessment

data for the 2014-2015 school year.

Table 6

*Results of Pearson Product-Moment Correlation Coefficient for Teacher Follow-Up*
*Ratings and Student i-Ready Reading Performance 2014-2015 School Year*

| Outcome | Teacher Follow-Up Ratings | | i-Ready Reading Performance | | | | | | |
|---------|------|------|------|------|----|--------|-------|----|--------|
|         | *M*  | *SD* | *M*  | *SD* | *n* | *r*   | *t*   | *df* | *p*  |
|         | 4.39 | 1.12 | 43.30 | 12.85 | 46 | -0.044 | 0.292 | 44 | 0.7715 |

*Note.*  Statistical significance is noted at $p \le .05$.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean

Teacher Follow-up Ratings on teacher evaluations ($M = 4.39$, $SD = 1.12$) for the 2014-

2015 school year and the mean student performance on the i-Ready reading assessment

($M = 43.30$, $SD = 12.85$) for the 2014-2015 school year.  The null hypothesis was not

rejected; t  The analysis revealed no positive correlation between teacher Follow-Up

Ratings and student i-Ready reading scores for the 2014-2015 school year, $r(44) = 0.044$,

$p = 0.7715$ ($r$-critical $= 0.288$, $\alpha = .05$).  This suggested there was not a relationship

between these two variables.

Table 7 displays the data examining the potential correlation between teacher's

follow-up ratings for the 2014-2015 school year and student MAP mathematics

assessment data for the 2014-2015 school year.

Table 7

*Results of Pearson Product-Moment Correlation Coefficient for Teacher Follow-Up Ratings and Student MAP mathematics Performance 2014-2015 School Year*

| Outcome | Teacher Follow-Up Ratings | | MAP Math Performance | | | | | | |
|---------|------|------|-------|-------|----|-------|-------|----|--------|
|         | *M*  | *SD* | *M*   | *SD*  | *n* | *r*  | *t*   | *df* | *p*  |
|         | 4.39 | 1.12 | 57.34 | 16.98 | 46 | 0.181 | 1.221 | 44 | 0.2287 |

*Note.* Statistical significance is noted at $p \leq .05$.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean Teacher Follow-up Ratings on teacher evaluations (*M* = 4.39, *SD* = 1.12) for the 2014-2015 school year and the mean student performance on the MAP mathematics assessment (*M* = 57.34, *SD* = 16.98) for the 2014-2015 school year. The null hypothesis was not rejected; the analysis also revealed no significant, positive correlation between teacher Follow-Up Ratings and student MAP mathematics scores for the 2014-2015 school year, *r*(44) = 0.181, *p* = 0.2287 (*r*-critical = 0.288, $\alpha$ = .05). This suggested there was not a relationship between these two variables.

Table 8 displays the data examining the correlation between teacher's follow-up ratings for the 2014-2015 school year and student i-Ready mathematics assessment data for the 2014-2015 school year.

Table 8

*Results of Pearson Product-Moment Correlation Coefficient for Teacher Follow-Up Ratings and Student i-Ready Mathematics Performance 2014-2015 School Year*

| Outcome | Teacher Follow-Up Ratings | | i-Ready Math Performance | | | | | | |
|---------|------|------|------|-------|----|-------|-------|----|--------|
|         | *M*  | *SD* | *M*  | *SD*  | *n* | *r*  | *t*   | *df* | *p*  |
|         | 4.39 | 1.12 | 58.5 | 13.06 | 46 | 0.225 | 1.532 | 44 | 0.1327 |

*Note.* Statistical significance is noted at $p \leq .05$.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean

Teacher Follow-up Ratings on teacher evaluations ($M = 4.39$, $SD = 1.12$) for the 2014-

2015 school year and the mean student performance on the i-Ready mathematics

assessment ($M = 58.5$, $SD = 13.06$) for the 2014-2015 school year. The null hypothesis

was not rejected; the analysis revealed no significant, positive correlation between

teacher Follow-Up Ratings and student i-Ready mathematics scores for the 2014-2015

school year, $r(44) = 0.225$, $p = 0.1327$ (*r*-critical $= 0.288$, $\alpha = .05$). This suggested there

was not a relationship between these two variables.

Overall, the data revealed no correlation between teacher follow-up ratings and

student achievement in mathematics after the first year of implementation of the new

teacher evaluation system.

Table 9 displays the data examining the correlation between teacher's follow-up

ratings for the 2015-2016 school year and student MAP English language arts assessment

data for the 2015-2016 school year.

Table 9

*Results of Pearson Product-Moment Correlation Coefficient for Teacher Follow-Up Ratings and Student MAP ELA Performance 2015-2016 School Year*

| Outcome | Teacher Follow-Up Ratings | | MAP ELA Performance | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *n* | *r* | *t* | *df* | *p* |
| | 5.17 | 1.03 | 69.18 | 13.57 | 46 | 0.192 | 1.237 | 44 | 0.2232 |

*Note.* Statistical significance is noted at $p \leq .05$.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean

Teacher Follow-up Ratings on teacher evaluations ($M = 5.17$, $SD = 1.03$) for the 2015-

2016 school year and the mean student performance on the MAP English language arts

assessment ($M = 69.18$, $SD = 13.57$) for the 2015-2016 school year. The null hypothesis

was not rejected; the analysis revealed no significant, positive correlation between

teacher Follow-Up Ratings and student MAP English language arts scores for the 2015-

2016 school year, $r(42) = 0.192$, $p = 0.2232$ ($r$-critical = 0.288, $\alpha = .05$).  This suggested

there was not a relationship between these two variables.

Table 10 displays the data examining the correlation between teacher's follow-up

ratings for the 2015-2016 school year and student i-Ready reading assessment data for

the 2015-2016 school year.

Table 10

*Results of Pearson Product-Moment Correlation Coefficient for Teacher Follow-Up*
*Ratings and Student i-Ready Reading Performance 2015-2016 School Year*

| Outcome | Teacher Follow-Up Ratings | | i-Ready Reading Performance | | | | | | |
|---------|------|------|------|------|----|------|------|------|--------|
|         | *M*  | *SD* | *M*  | *SD* | *n* | *r*  | *t*  | *df* | *p*    |
|         | 5.19 | 1.03 | 44.41 | 18.20 | 46 | 0.187 | 1.263 | 44 | 0.2134 |

*Note.*  Statistical significance is noted at $p \leq .05$.

A PPMCC (Pearson $r$) was calculated to assess the relationship between the mean

Teacher Follow-up Ratings on teacher evaluations ($M = 5.19$, $SD = 1.03$) for the 2015-

2016 school year and the mean student performance on the i-Ready reading assessment

($M = 44.41$, $SD = 18.20$) for the 2015-2016 school year.  The null hypothesis was not

rejected; the analysis revealed no significant, positive correlation between teacher

Follow-Up Ratings and student i-Ready reading scores for the 2015-2016 school year,

$r(42) = 0.187$, $p = 0.2134$ ($r$-critical = 0.288, $\alpha = .05$).  This suggested there was not a

relationship between these two variables.

Overall, the data revealed no correlation between teacher evaluation follow-up

ratings and student performance and reading assessments after the second year of

implementation of the new teacher evaluation system.  This suggested there was no

relationship between these two variables.

Table 11 displays the data examining the correlation between teacher's follow-up

ratings for the 2015-2016 school year and student MAP mathematics assessment data for

the 2015-2016 school year.

Table 11

*Results of Pearson Product-Moment Correlation Coefficient for Teacher Follow-Up*
*Ratings and Student MAP Mathematics Performance 2015-2016 School Year*

| | Teacher Follow-Up Ratings | | MAP Math Performance | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Outcome | *M* | *SD* | *M* | *SD* | *n* | *r* | *t* | *df* | *p* |
| | 5.16 | 1.06 | 65.95 | 13.94 | 46 | -0.118 | -0.761 | 44 | 0.4511 |

*Note.*  Statistical significance is noted at $p \leq .05$.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean

Teacher Follow-up Ratings on teacher evaluations ($M = 5.16$, $SD = 1.6$) for the 2015-

2016 school year and the mean student performance on the MAP mathematics assessment

($M = 65.95$, $SD = 13.94$) for the 2015-2016 school year.  The null hypothesis was not

rejected; the analysis revealed no positive correlation between teacher Follow-Up Ratings

and student MAP mathematics scores for the 2015-2016 school year, $r(41) = -0.118$, $p =$

0.451 (*r*-critical $= 0.288$, $\alpha = .05$).  This suggested there was insufficient evidence to

support a relationship between these two variables.

Table 12 displays the data examining the correlation between teacher's follow-up

ratings for the 2015-2016 school year and student i-Ready mathematics assessment data

for the 2015-2016 school year.

Table 12

*Results of Pearson Product-Moment Correlation Coefficient for Teacher Follow-Up Ratings and Student i-Ready Mathematics Performance 2015-2016 School Year*

| Outcome | Teacher Follow-Up Ratings | | i-Ready Math Performance | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | *M* | *SD* | *M* | *SD* | *n* | *r* | *t* | *df* | *p* |
| | 5.17 | 1.03 | 44.41 | 18.20 | 46 | 0.187 | 1.263 | 44 | 0.2134 |

*Note.* Statistical significance is noted at *p* ≤ .05.

A PPMCC (Pearson *r*) was calculated to assess the relationship between the mean Teacher Follow-up Ratings on teacher evaluations (*M* = 5.17, *SD* = 1.03) for the 2015-2016 school year and the mean student performance on the i-Ready mathematics assessment (*M* = 44.41, *SD* = 18.20) for the 2015-2016 school year.  The null hypothesis was not rejected; the analysis revealed no significant, positive correlation between teacher Follow-Up Ratings and student i-Ready mathematics scores for the 2015-2016 school year, *r*(44) = 0.187, *p* = 0.2134 (*r*-critical = 0.288, *α* = .05).   This suggested there was no relationship between the two variables.

Overall, the data revealed there was no correlation between teacher follow-up ratings and student mathematics performance after the second year of implementation of the new teacher evaluation system.

**Question one.**  Teachers and evaluating administrators perceived that the District's teacher evaluation system does not have a positive effect on improving teacher practice and professional growth. Null hypothesis three involved analysis of teachers' and administrators' perceptions of the effect of the new teacher evaluation system on improvement in teacher practice, through a five point Likert-scale survey.  The survey yielded informative quantitative data from both teachers and administrators.

**Teacher Survey data**.  The researcher used inferential data from teachers to determine what the teachers' perceptions were to the new evaluation system.  Questions on the survey were analyzed to see if the scores were significantly higher than neutral (3).

Table 13 displays the data examining the perception of teachers on the then-new teacher evaluation system, as measure through the five-point Likert survey.

Table 13

*Teacher Survey Data Analysis*

| Survey Question | Mean | Median | Mode | Range |
|---|---|---|---|---|
| 1.  The current teacher evaluation process is a fair and objective measure of my teaching ability and performance. | 2.96 | 3 | 3 | 4 |
| 2.  The current teacher evaluation framework allows my administrator to assess a more accurate picture of my teaching ability than the previous teacher evaluation model used in the district. | 2.92 | 3 | 2, 4 | 4 |
| 3.  I have received more useful feedback from my evaluator under the current teacher evaluation process than I received under the previous teacher evaluation process. | 2.8 | 3 | 4 | 4 |
| 4.  The current teacher evaluation process requires me to focus more on strategies to achieve specific development and student achievement goals than the previous evaluation process. | 3.44 | 4 | 4 | 4 |
| 5.  The current teacher evaluation process is implemented across the district with consistency. | 2.64 | 3 | 4 | 3 |
| 6.  Overall, I believe the current teacher evaluation process is more effective for evaluating teachers and determining effective teaching than the previous process. | 2.84 | 3 | 3 | 4 |

**Continued**

Table 13 continued.

| | | | | |
|---|---|---|---|---|
| Overall | 2.92 | | | |
| Survey Question | Mean | Median | Mode | Range |
| 8. Focusing on specific goals/indicators within the current teacher evaluation framework has led to changes in my practice that have improved my planning, preparation, and instruction. | 3.6 | 4 | 4 | 4 |
| 9. The feedback from my principal during formal and/or informal observations was helpful and constructive, and the feedback resulted in changes or improvement in my instructional practice. | 3.44 | 4 | 4 | 4 |
| 10. The current teacher evaluation process has had a positive impact on my students' learning and academic achievement. | 3.04 | 3 | 3 | 4 |
| 11. As a result of the student growth measure component of the current teacher evaluation process, I use more student assessment data to guide my instruction. | 3.32 | 4 | 4 | 4 |
| 12. As a result of the Professional Growth Plan (PGP) component of the current teacher evaluation process, I reflect more on my impact on student learning, growth, and achievement. | 3.2 | 4 | 4 | 4 |
| Overall | 3.32 | | | |

*Note*: Questions had 27 respondents, Likert Scale: 1—Strongly Disagree, 2—Disagree, 3—Neither Disagree nor Agree, 4—Agree, 5—Strongly Agree

The mean for the multi-item construct for questions 1 through 6, which dealt with comparing the then-current teacher evaluation process with the previous evaluation process used in the district was 2.92 on the Likert scale, indicating that teachers generally had a negative perception of the then-new evaluation process. As seen on Table 13, with a mean of 3.44, the item receiving the highest overall ratings in this category was item 4, which was related to the then-current teacher evaluation system's increased focus on

specific strategies to target specific student needs.  Item 1, concerning the then-current

teacher evaluation system's ability to provide a fair and objective measure of teachers'

ability and performance, received the second highest mean of 2.96.  Item 5, which

pertained to the consistency of implementation of the then-current teacher evaluation

system's implementation across the district, received the lowest mean of 2.64.  The Likert

average scale that appeared most often in these questions was 4.

The mean for the multi-item construct for questions 8 through 12, which dealt

with specific components of the then-current teacher evaluation process (e.g. PGP goals,

administrator feedback, impact on student and teacher growth), was 3.32 on the Likert

scale, indicating that teachers generally had a positive perception of the individual

components comprised in the then-current teacher evaluation process.  There were 27

teachers that completed this portion of the survey.  As can be seen in Table 13, the item

that received the highest overall ratings in this category was item 8, with a mean of 3.6,

concerning the then-current teacher evaluations' focus on specific teaching goals that

resulted in improved planning, preparation and instruction.  The item receiving the

second highest mean of 3.44 was item 9, which was related to evaluator feedback, which

resulted in teachers' reflecting more on improving instructional practices.  The item that

yielded the lowest mean of 3.04 was item 10, concerning the then-current evaluation

system resulting in positive impact on student learning and achievement.

**Administrator survey data.**  Table 14 displays the data examining the perception

of administrators on the then-new teacher evaluation system.  The sample of

administrator responses was too small for analyses; however, the descriptive statistics are

reported in the following. Table 14.

Table 14

*Administrator Survey Data Analysis*

| Survey Question | Mean | Median | Mode | Range |
|---|---|---|---|---|
| 1.  The current teacher evaluation process is a fair and objective measure of my teaching ability and performance. | 4.25 | 4 | 4 | 1 |
| 2.  The current teacher evaluation framework allows my administrator to assess a more accurate picture of my teaching ability than the previous teacher evaluation model used in the district. | 4.25 | 4 | 4 | 1 |
| 3.  I have received more useful feedback from my evaluator under the current teacher evaluation process than I received under the previous teacher evaluation process. | 4.5 | 4.5 | 4, 5 | 1 |
| 4.  The current teacher evaluation process requires me to focus more on strategies to achieve specific development and student achievement goals than the previous evaluation process. | 4.25 | 4 | 4 | 1 |
| 5.  The current teacher evaluation process is implemented across the district with consistency. | 4.25 | 4.5 | 5 | 2 |
| 6.  Overall, I believe the current teacher evaluation process is more effective for evaluating teachers and determining effective teaching than the previous process. | 4.25 | 4 | 4 | 1 |
| Overall | 4.29 | | | |
| 8.  Focusing on specific goals/indicators within the current teacher evaluation framework has led to changes in my practice that have improved my planning, preparation, and instruction. | 4.75 | 5 | 5 | 1 |

Table 14 continued.

| Survey Question | Mean | Median | Mode | Range |
|---|---|---|---|---|
| 9.  The feedback from my principal during formal and/or informal observations was helpful and constructive, and the feedback resulted in changes or improvement in my instructional practice. | 4.25 | 4 | 4 | 1 |
| 10.  The current teacher evaluation process has had a positive impact on my students' learning and academic achievement. | 4 | 4 | 4 | 0 |
| 11.  As a result of the student growth measure component of the current teacher evaluation process, I use more student assessment data to guide my instruction. | 4 | 4 | 4 | 2 |
| 12.  As a result of the Professional Growth Plan (PGP) component of the current teacher evaluation process, I reflect more on my impact on student learning, growth, and achievement. | 4.5 | 4.5 | 4, 5 | 1 |
| Overall | 4.3 | | | |

*Note*:  Questions had 4 respondents, Likert Scale: 1—Strongly Disagree, 2—Disagree, 3—Neither Disagree nor Agree, 4—Agree, 5—Strongly Agree

The mean for the multi-item construct for questions 1 through 6, which dealt with comparing the then-current teacher evaluation process with the previous evaluation process used in the district, was 4.29 on the Likert scale, indicating that administrators generally had a positive perception of the then-new evaluation process over the previous evaluation system used in the district.  Four administrators participated in this portion of the survey.  As seen on Table 14, with a mean of 4.5, the item receiving the highest overall ratings in this category was item 3, concerning the then-current teacher evaluation system process resulting in administrators providing more useful feedback to teachers than under the previous teacher evaluation system.  All other items in this category

received a mean score of 4.25, with item 5, concerning implementation consistency of the then-current teacher evaluation process across the district, yielding the greatest range within participant responses. The Likert average scale that appeared most often in these questions was 4.

The mean for the multi-item construct for questions 8 through 12, which dealt with specific components of the then-current teacher evaluation process (e.g. PGP goals, administrator feedback, impact on student and teacher growth), was 4.3 on the Likert scale, indicating that administrators also had a positive perception of the individual components comprised in the then-current teacher evaluation process. Four administrators completed this portion of the survey. As can be seen in Table 14 the item that received the highest overall ratings in this category was item 8, with a mean of 4.75, concerning the then-current teacher evaluations' focus on specific teaching goals and belief that this resulted in improved planning, preparation and instruction. The item receiving the second highest mean of 4.5 was item 12, related to the PGP component of the then-current teacher evaluation system resulting in teachers to reflect on and improve his/her instructional practices. The Likert average scale that appeared most often in these questions was 4.

Question one. How do teachers and evaluating administrators perceive the district's teacher evaluation as having a potentially positive effect on improved teacher practice and professional growth? Question one involved comparing Teacher perceptions with Administrator perceptions as measured by the survey instruments. Table 15 represents the data comparing teacher's perceptions on each survey question compared with the administrator's perceptions on the same questions.

Table 15

*Results of Teacher Survey and Administrator Survey*

| Outcome | Teacher Perception Survey Results (n = 25) | | | | Administrator Perception Survey Results (n=4) | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Min | Max | Mode | *x-bar* | Min | Max | Mode | *x-bar* |
| Question 1 | 1 | 5 | 3 | 2.96 | 4 | 5 | 4 | 4.25 |
| Question 2 | 1 | 5 | 2,4 | 2.92 | 4 | 5 | 4 | 4.50 |
| Question 3 | 1 | 5 | 4 | 2.80 | 4 | 5 | 4,5 | 4.50 |
| Question 4 | 1 | 5 | 4 | 3.44 | 4 | 5 | 4 | 4.25 |
| Question 5 | 1 | 4 | 3 | 2.64 | 3 | 5 | 5 | 4.25 |
| Question 6 | 1 | 5 | 3 | 2.84 | 4 | 5 | 4 | 4.25 |
| Questions 1-6 Overall | | | | 2.933 | | | | 4.333 |
| Question 8 | 1 | 5 | 4 | 3.60 | 4 | 5 | 5 | 4.75 |
| Question 9 | 1 | 5 | 4 | 3.44 | 4 | 5 | 4 | 4.25 |
| Question 10 | 1 | 5 | 3 | 3.04 | 4 | 4 | 4 | 4.00 |
| Question 11 | 1 | 5 | 4 | 3.32 | 4 | 5 | 4 | 4.00 |
| Question 12 | 1 | 5 | 4 | 3.20 | 4 | 5 | 4,5 | 4.50 |
| Questions 8-12 Overall | | | | 3.32 | | | | 4.30 |

*Note.* Min = Minimum; Max = Maximum; and *M* = Mean

As Table 15 illustrates, a significant difference between the teachers' perception and the administrators' perception of the then-new teacher evaluation system (as it pertains to the implementation process of the evaluation system), which were analyzed through survey questions 1 through 6.  Table 15 also indicates that administrators perceive the effectiveness (as it pertains to student and teacher outcomes) of the then-new teacher evaluation system much higher than the teachers, which were analyzed through survey questions 8 through 12.

**Qualitative Data**

The research methodology included a qualitative component to provide a better understanding of the issues and to add insight to the research questions investigated from both the administrator and teacher point of view.   The following research question guided the qualitative portion of this study:  What are the teachers' and administrators' perceptions of the researched district's new teacher evaluation process (based on the MO Teacher Evaluation Model) as a method for improving professional practice (as measured by the Teacher Evaluation Rating tool) and influencing student achievement (as measured by the MAP and i-Ready assessments)?

The qualitative portion of this study involved administering a customized survey (Appendix E; Appendix F) to 45 teachers and nine administrators in six elementary schools at the end of the 2015-2016 school year; the same teachers and administrators were involved in the quantitative portion of this study.  The survey was comprised of 11 items requesting participants to rate their level of agreement to statements related to their experiences with, and perceived effectiveness of, the new teacher evaluation system on a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

Implementation of the survey served three goals: (1) to identify how teachers and their evaluating administrators in the researched school district viewed the districts' teacher evaluation tool and process compared to the previous teacher evaluation model; (2) to gain insight into teachers and their evaluating administrators in the researched school districts' perception of the new teacher evaluation tool as an accurate method for evaluating a teacher's quality and effectiveness; and (3) to assess the perception of the researched district's teachers and evaluators perception of the effectiveness of the new

teacher evaluation tool  for improved educator practice and increased student learning

and achievement.

**Teacher survey data.**  In addition to the Likert scale survey items, the survey

also included three optional, open-ended questions included for the purpose of collecting

qualitative data to further explore the research question.  Questions 7, 13, and 14 were the

open-ended survey questions, which allowed participants to share their thoughts,

concerns, and suggestions regarding the implementation of the researched districts'

teacher evaluation model.  The survey responses were coded by searching through the

data to identify commonly cited responses and to represent the findings supported by the

evidence and substantiated with quotes from the survey participants.

Survey questions #1 through 6 were focused on teachers and administrators

comparing the previous evaluation system and the then-current system (e.g. the

objectivity, effectiveness, and accuracy of the evaluation tool for achieving its desired

results) on a Likert scale. The quantitative data were discussed in the Quantitative data

section of this chapter.  Question 7 allowed participants to provide comments to further

explain their personal experiences or perceptions of these aspects of the district's then-

current teacher evaluation process.

Survey questions #8 through 12 were focused on determining teachers' and

administrators' perceptions of the specific components of the then-current teacher

evaluation process and the impact of these on improved teacher instructional practice and

increased student achievement.  Some of these components included focusing on specific

goals/indicators, which resulted in changes to teachers' instructional practices and

resulted in a positive impact on students' learning and academic achievement.  Question

13 allowed participants to provide comments to further explain their personal experiences

or perceptions of these aspects of the district's then-current teacher evaluation process.

From a qualitative perspective, the survey yielded useful information. Some of the

survey participants offered comments.  Three major themes emerged from the open-

ended survey responses:  (1) there was apprehension regarding the implementation

process across buildings (or administrators) in the district, (2) there were perceived

concerns regarding specific components of the then-current teacher evaluation model,

and (3) there were perceived advantages and disadvantages to the district's new teacher

evaluation model.  Each of these three themes contained subthemes that supported them.

On occasion, the researcher was able to draw connections from the quantitative data to

the qualitative data.  Table 8 displays the data related to themes and sub-categories

identified from the participants' responses.

Table 16.

*Themes*

| Theme Title | Categories |
| --- | --- |
| 1.  Apprehension Regarding the Implementation Process | Subjectivity; Administrator Bias; Inconsistencies across Administrators/District |
| 2.  Concerns Regarding Specific Components of the Evaluation Process | Lack of Clarity on Process; Assessment Data Used; Other Factors to Consider |
| 3.  Perceived Advantages and Disadvantages to the Evaluation Process | Focus on Specific Strategies; Increased Teacher Involvement; Increased Stress on Teachers; Increased Teacher Workload; Over-Emphasis on Assessment |

**Apprehension regarding the implementation process.**  The most prevalent

responses in the open-ended comments regarded the consistency of implementation of the

new teacher evaluation model across the district.  Particularly, concerns addressed

included: the subjectivity of the evaluation tool for assessing teacher effectiveness,

evaluator bias interfering with the validity of the evaluation process, and an inconsistency

of expectations, procedures, and implementation of the process across the buildings and

administrators in the district.

   *Subjectivity of the evaluation process*.  One area of concern expressed by the

study participants was concern that aspects of the evaluation process were too subjective.

One teacher specifically cited student growth goals and BAS assessment data as too

subjective to be considered in the teacher evaluation process.

   *Evaluator bias*.  Another area of concern identified by teacher participants was a

perceived bias of evaluating administrators that interfered with the validity of the

evaluation process.  One teacher explained, 'In my experience, evaluations are very

inconsistent and based more on principal-teacher friendships and personal relationships

rather than on effective teaching.'

   *Inconsistency of implementation*.  Another area of concern presented by the

survey participants was a lack of clarity and inconsistency of the implementation of the

teacher evaluation process.  Comments regarding clarity of the process and procedures

included: 'I don't think the language of the desired standards is easy to understand,' and

'If I had known that my growth goals were going to be averaged I would have picked

something different or rated myself lower at the beginning of the year.'  Other survey

participants felt the process lacked consistency across the buildings in the district,

particularly; with the way the building administrators interpreted the process.  One

respondent explained, 'I have heard from other teachers that each administrator in the

district interprets the evaluation tool differently.'  Another participant proposed, 'Each principal is different and wants different things for their building to focus on.'  Teachers also responded that they could not attest to how the evaluation process was implemented across the district.

*Concerns regarding specific components.*  Additionally, teachers proposed concerns regarding some of the specific components of the district's then-current teacher evaluation process.  Particularly, concerns proposed included: a perceived over-emphasis on student assessment data in teacher evaluations, dissatisfaction regarding the timeline and types of student assessments used, inconsistent interpretation of student assessment data used, and the need to consider other factors in the teacher evaluation process.

*Over-emphasis on student assessment data.*  Teachers expressed concern that the new evaluation tool placed too much weight on student assessment data for determining teacher effectiveness.  One teacher respondent stated, 'I feel the new evaluation tool focuses too much on student assessment scores (in many cases i-Ready).'  Another participant responded, 'I feel that using one test score to evaluate a teacher's whole year is not necessarily accurate.'  In like manner, a teacher indicated, 'I feel that 1/3 of the emphasis on our summative evaluations should not be placed on this specific [i-Ready] assessment.'  Another wrote, 'I just do not agree with how much of an impact i-Ready has on your evaluation.'  Even the use of student assessment data in general was questioned.  'I strongly feel, even having strong student scores, that we cannot judge and grade a teacher on student achievement,' explained one teacher.

*Inappropriate timeline and use of student assessment data*.  Teacher participant respondents revealed some trepidation that the mid-year data point is used to report

whole year student growth, which many teachers argued is misrepresentation of actual student growth over the year, because an end of year assessment is also given, but not consistently used due to a misalignment between end of year assessments and when summative evaluations are due.  For example, one teacher stated, 'I am concerned to hear from others throughout the district that teachers are being judged by middle of the year data. This process should be judged on what students accomplish in an entire year.'   A similar criticism was 'Evaluating me on where my students are at in Feb./Mar. makes me feel like April and May don't count and that my students should be at their end of the year goal by them.  Base my evaluation on where they were in Dec. on i-Ready and BAS not compared to my end of the year numbers, which can't be configured because students are still learning.'

   ***Concerns over assessments***.  Another alarm revealed from teacher participant statements related to the particular assessments that are utilized in the teacher evaluation system to determine teacher effectiveness and ratings.  One teacher expounded, 'I was scored a 1 in the area of student growth because only 66% of my students made the required growth on i-Ready.  I have many other pieces of evidence that can prove that they have made growth.  Some of the students that didn't make the required assessments were already performing ABOVE grade level.'

   ***Inconsistent interpretation of student data***.  Teacher participants disclosed concern over the consistency in which Student Growth Goals, and measurement thereof, were determined and implemented across the grades and buildings in the district.  One contributor explained their understanding of the process as follows 'The [Student Growth] goals were to be written showing student growth as the evaluation tool of the

teacher.'  Teachers identified this as a problem when they felt the expectations were not

consistent across the district.  For example, one participant stated, 'Factors used to

evaluate are often subjective.  Our building goals are based on 100% of students

achieving the goal, other buildings' percents vary considerably. Some buildings have

unmeasurable (sic) completely subjective goals with lower percentiles, while others have

100% expectations for student achievement.'  Another stated, 'Our evaluation is not the

same, in that the qualifiers for student achievement is not 100% across the district, which

does impact our evaluation.'  Another further explained a possible discrepancy, 'By

allowing teachers to set up their own goals, there is no consistency in which they are

evaluated.'  An additional teacher responded with dissatisfaction in the following way

'The expectation that all students will reach 100% of their goal is unrealistic.  I have

students that have not made a year's growth since Kindergarten and yet when they are in

third grade they are expected to make a year's growth or more.'

     ***Other factors that should be considered***.  Teachers felt that the teacher evaluation

process should take into consideration additional factors.  One responded, 'We have

numerous measures [of student growth] including BAS, common assessments, and daily

work/observation.  Why can't all of these be a part of it [teacher evaluation]?  We need to

look at the whole picture.'   Another participant mentioned the possibility for dips in

results after implementation of new teaching strategies.  They explained, 'Some new

practices take longer to show results than others and this should be reflected in the

process.'

     Another complication highlighted by a survey participant was that teachers might

not always be evaluated by the same administrator from year to year, which might

introduce a compounding factor in the teacher evaluation process.  One participant explained, 'Having two principals during this time has been hugely different, with one showing distinct favorites and riding some teachers while ignoring other teacher behaviors.'  In a similar fashion another teacher expressed frustration, 'It is regretful that a teacher could get exemplar reviews and at the discretion of another principal they would be reviewed poorly.'

Additionally, concerns were revealed pertaining to assessment data on students with various learning difficulties.  They explained, 'For many low students, ELL students, IEP students, or poor test takers, this assessment [i-Ready] did not reflect their true ability and according to the [Student Growth] goals, showed the teachers as [demonstrating] poor growth in reaching their goal.  I believe other factors should be taken into affect (sic).'  Another teacher responded in a similar vein, 'Until student histories, DFS situations, IQ's IEP's, behaviors, home issues, gifted, number of students and other uncontrolled variables are taken into consideration, teachers should be evaluated with a tool similar to the previous tool.'  Still another teacher acknowledged, 'There are so many factors that go into how students test.'

**Perceived advantages and disadvantages of the new process.**  Teacher respondents also revealed what they perceived as advantages and advantages to the new teacher evaluation process.  Specifically, advantages offered surrounded the PGP component of the teacher evaluation process.  The disadvantages cited included: added stress and workload on teachers, potential for 'gaming the system' by teachers, potential failure of system to accurately identify teacher performance, lack of adequate training on the system, and limitations on teacher performance categories.

***Focus on specific strategies***.  Teachers felt positively about the PGP requirement for teachers to focus on specific strategies to improve his/her instructional practice related to improving student learning and achievement in demonstrated areas of need.  Teachers also commented that they appreciated the teacher voice in establishing what strategies to focus on in their PGP, and teacher participation allowed when setting their PGP growth goals, and establishing their benchmark and follow-up ratings.  It is also important to note that teacher participation in establishing goals and ratings, teacher contributions to the goal-setting and rating process were also described as a con by teacher participants.

***Added stress and teacher workload***.  Negative perceptions were revealed about increased stress and workload on teachers due to the new teacher evaluation system and processes.  One participant explained, 'I feel that the burden of 'proving' we are effective teachers has been added to our workload.  I don't feel that this makes me a better teacher either.' Another explained this in a slightly different manner, 'I disagree with the way teachers are told to rate themselves at the beginning of the school year.  We are experienced teachers and we rate ourselves low only so we can show growth.'  Another described the system as 'Unrealistic and stressful for the teacher.'  Finally, one teacher stated, 'I worry more about my job security.  I feel pressured to ensure my students perform well on i-Ready.'

***Potential for 'gaming the system****.'*  Teachers also responded negatively to what they considered as a possibility that teachers could undermine the process.  One participant explained, 'Some [PGP] goals can be poorly written and can set up a good teacher for failure.  Also a weaker teacher could set up an easy to attain goal and make themselves look like a proficient teacher.'  Another example of this potential was the

response, 'If I had known that my growth points were going to be averaged I would have picked something different or rated myself lower at the beginning of the year.  This system does not promote success.'

*Inaccurately distinguishing teacher performance*.  Teachers identified an undesirable possibility that the teacher evaluation process could inaccurately or misidentify teacher performance.  One participant stated, 'It [evaluations] is very subjective and not consistent from school to school.  This puts teachers at a disadvantage if they aren't on a friendly basis with their principal.'

In addition, teachers proposed the possibility of strong teachers being penalized or rated poorly due to the inclusion teacher growth percentile on PGP goals as a component of the new evaluation process.  One explained, 'I also believe that more experienced teachers show 'less growth' or 'no growth' because many are already proficient in certain areas.'  Another teacher wrote, 'My own growth rated me at a 2 because I went up an average of 2 points in each area, one of which I wouldn't have chosen, but it was a building goal.'

*Inadequate training and support for the new process*.  Some responses indicated a lack of clarity on processes and protocols involved with the teacher evaluation system.  One teacher stated, 'I feel there could have been a little more clarity on the assessment data that should have been utilized.'  One participant explained confusion with the language of expectations related to specific standards.  'I think the desired standards should explicitly identify what the desired outcome or look-for is, such as 'anchor charts or visual displays clearly show student thinking or learning' or 'teacher provides rigorous tasks that challenge students'.'  Another wrote, 'I am not sure how the evaluation process

is implemented across the district.'  Another concern revealed was relating to inadequate

training or support for the newly implemented teacher evaluation process.  'I have heard

from other teachers that each administrator in the district interprets the evaluation tool

differently and has a different level of interaction with each teacher in setting goals.'

*Limitations of teacher performance categories*.  Another concern represented by

a teacher participant related to teacher performance level descriptors for teachers on the

summative evaluation document.  This participant wrote, 'Another concern that I have is

that the only way to be 'distinguished' is for one to present your knowledge to others.  I

don't believe presenting information to others makes me a better teacher.'

**Administrator survey data.**  While the administrator survey included the same

optional open-ended questions as the teacher survey, none of the four administrator

respondents elected to respond; and therefore, yielded no usable data to report.

**Summary**

In Chapter Four, the results from this mixed-methods study were presented.  The

quantitative data and descriptive statistics were reported in tables, and the qualitative data

were reported in thematic categories developed throughout the data analysis process.

Chapter Five includes a summary of the findings, conclusions, recommendations for

further research, and implications for practice.

## Chapter Five: Discussion and Reflection

The focus of this mix methods study was to determine if a relationship existed between archival student MAP and i-Ready performance data when compared to the researched districts' evaluation data of classroom teachers, generated by a newly-implemented evaluation tool and process. .

### Research Question

The research question that guided the work of this dissertation was:

What are the teachers' and administrators' perceptions of the researched district's new teacher evaluation process (based on the MO Teacher Evaluation Model) as a method for improving professional practice (as measured by the Teacher Evaluation Rating tool) and influencing student achievement (as measured by the MAP and i-Ready assessments)?

This question addressed whether teachers and administrators buy-in and perceived value/effectiveness of the evaluation process and tools on their practice/ professional growth and on student outcomes.

### Research Question and Hypotheses

This mixed methods study was guided by the following question and hypotheses:

$Q_1$: Teachers and evaluating administrators perceive the district's teacher evaluation having a positive effect on improved teacher practice and professional growth.

$H_1$: Student achievement as measured on the MAP and i-Ready assessments will improve after implementation of the new teacher evaluation system.

$H_2$: There is a positive correlation between teacher performance ratings and student achievement on the MAP and i-Ready assessments.

This mixed-methods research study was undertaken to provide scientifically valid and reliable data to assist school district administration in assessing the effectiveness of the researched district's then-new teacher evaluation system.  This study provided an initial dataset and baseline point for evaluation of the local districts' teachers over time as to the quantifiable determination of classroom effectiveness compared to changes in annual standardized test scores by classroom sets of students.

**Findings**

The analysis concluded the i-Ready and MAP assessments as moderately correlated, with reading (English language arts) showing a stronger correlation than mathematics.  This finding suggests that the students' performance on one could be predictive of performance on the other assessment.

The analysis of student MAP data revealed a statistically significant increase in students' achievement scores each year after implementation of the new teacher evaluation system in English language arts (reading).  However, in mathematics, the MAP data revealed no significant difference between students' achievement after the first year of the new teacher evaluations system's implementation; but revealed a significant increase between students' mathematics achievement scores after the second year of implementation.  In contrast, the analysis of i-Ready assessment data revealed no statistically significant difference between students' i-Ready reading or mathematics achievement scores over the first two years of the then-current teacher evaluation system. Therefore, in response to hypothesis one, the researcher found that overall student performance increased over the first two years of the new teacher evaluation system's implementation and rejected null hypothesis one.

The analysis of data seeking a relationship between teacher evaluations and students' assessment data revealed a lack of significance and non-linear fit between the Teacher Follow-Up Ratings and student performance on English language arts (reading) and mathematics assessments for both years comprised in the study.  Therefore, the researcher also rejected null hypothesis two.

The analysis of the teacher and administrator perception data revealed diverging views on how the new teacher evaluation system was implemented across the district, the new teacher evaluation system's ability to improve student and teaching performance, and the new teacher evaluation system's ability to accurately determine a teacher's effectiveness.  Generally, teachers viewed the system with more skepticism than the administrators.  Therefore, in response to question one, the data, while useful, remained inconclusive.

**Conclusions**

When interpreting data for hypothesis two, the data did not illustrate a correlation or predictive relationship between teacher performance and student achievement through the variables examined in this study.  A modified research design, which used a different measure for rating teacher performance, may have been better suited for establishing a link.  However, at the onset of this study, an alternative measure of teacher effectiveness was not an available component of the researched district's teacher evaluation system, and therefore, would have dramatically altered the timeline of the study.

When interpreting qualitative data for question one, the data suggested teachers perceived the then-new teacher evaluation system as having an influence on use of instructional strategies and improvement in professional practice.  While there was no

quantitative evidence to support a direct impact, it appeared that the majority of the

teacher participants indicated that the system had an impact on planning their instruction.

Improvements in student assessment data could be a result of the changes to the district's

teacher evaluation process.  However, this change in student achievement could also be

related to other factors not examined in this study.  One could argue that improvement of

instructional practices was a result of increased collaborative discussions through

Professional Learning Committees, which were also established in the district during this

time period.  Others may suggest that the improvement was a nod toward improved

feedback to teachers from evaluating administrators.  In any case, with the call for

increased accountability on educators, districts should support teachers in developing

innovative methods and strategies for advancing student learning and academic

achievement.

**Implications for Practice**

The correlational data generated by this mixed-methods study comparing student

performance on standardized assessments with teacher evaluation ratings, was based on

evaluators utilizing one district's newly-adopted teacher evaluation system.

At a micro level, the results of this study may be used as a baseline for further

investigation and modifications of the researched district's teacher evaluation system.  A

direct application of this research may be to guide professional development plans to

improve teacher performance and affect student achievement.  Employing this research

study as a framework, the researched district could continue examination of quantifiable,

successive years of performance data, which may allow administrators to identify

established patterns to be utilized for making teaching assignments.

At the macro-level, the findings from this research may add to the growing body of knowledge and research on the development of effective teacher evaluation systems for promoting improved educator practice, which could then affect student achievement growth. A main benefit of the study may be to substantiate the complexity of tailoring a teacher evaluation tool with sensitive enough measures to accurately assign a discrete number to a teacher's effectiveness or impact on student learning.  The absence of research available on how to isolate a teacher's direct impact on student learning (for the purpose of measuring individual teacher quality) within the context of a specific teacher evaluation system is particularly problematic.

Other points to consider for the improvement of teacher performance and students' assessment achievement would be a trajectory data analysis over multiple years while considering specific teachers, grade-level groupings, and content areas.  It is worth noting that the teacher evaluation survey in this study was not utilized by a representative sample that could scientifically support the notion that all teachers in the researched district had the same perspectives of the teacher evaluation system.

**Recommendations for the Program**

In short, the qualitative feedback for the researched district's then-new teacher evaluation model revealed conflicting perceptions of the new evaluation.  The results indicated teachers had a negative opinion of the system's implementation and introduction.  However, teacher participants also indicated a generally positive view of the individual components of the then-current teacher evaluation in regards to its impact on both professional improvement and students' academic growth.  In contrast, the qualitative feedback from administrators indicated a highly positive perception of the

district's new teacher evaluation system for both the implementation and its impact on student and instructional improvement.

It is important to note that only 27 out of over 40 teachers, about two thirds of the total district's staff population, opted not to participate in this study.  Perhaps, one could conclude the staff may have felt uncomfortable with expressing feelings regarding the district without repercussion with regard to their expressed views.

The number of teachers evaluated at this level identified a concern regarding partial understanding of the newly-adopted teacher evaluation system.  This may be an indication that building administrators who conducted the teacher evaluations may be at capacity, or suggest a lack of training with the new system.  Additional faculty may need to be assigned to this task.  One possibility could be a building liaison, or staff member dedicated to fielding questions and/or monitoring the implementation of the program across the district, which would reduce the pressure on building principals and maintain the integrity of the program.  One prominent study by Baker et al. (2010) implied a teacher's lack of understanding of a system, or failure to have a shared belief in the system, may result in the improper use of the tool in a manner that facilitates professional growth or fosters student improvement.

The program should continue to offer the teacher evaluation model, but additional options should be investigated, such as alternative methods for determining a teacher's overall rating and a more objective rubric for determining education benchmark and follow-up ratings on the PGP.  The district should also offer support meetings or workshops for new teachers on the teacher evaluation process, and for any teachers that request further assistance or training. Communication with all staff and consistency of the

evaluation programs implementation across administrators and buildings across the district is key.  Finally, the district should explore other assessment options for measuring student growth, especially as it pertains to the measurement of teacher effectiveness and impact on students.

**Recommendations for Future Research**

Additional recommendations for future research components may include expanding examination of teacher follow-up ratings to a different, more sensitive metric for evaluating teacher effectiveness.  Analysis of the teacher evaluation scores could also be expanded to the other middle school and secondary levels, which may provide more statistically relevant results.

A single district study is not sufficient for truly evaluating a new evaluation format, especially one designed to facilitate evaluation of teacher performance.  Follow-up studies could investigate teachers' improvement after their summative evaluation year and perhaps find reasons why teachers do not make a district-desired impact upon their students' growth.  Additional investigations on how the district's teacher evaluation system is monitored, supported and systematically evaluated is also a potential avenue for future research.  In addition, having adequate time to provide frequent observations and confer with teachers to provide actionable feedback was identified as a challenge for administrators, upon implementing the revised teacher evaluation processes.  Having adequate time is crucial for administrators to observe, evaluate, provide feedback, and have meaningful discussion with teachers.  Teachers, as well, need adequate time to reflect on practice and feedback, and strategize for improvements in their performance.  Therefore, as previously mentioned, future research should investigate alternative models

and roles within school administration to provide more focus on instructional leadership

by accommodating more frequent observations and meaningful feedback sessions.

Finally, for a better measure of attaining stakeholder feedback it is  recommended

that the district administer a validated survey instrument at the beginning of the school

year and again at the end — in a pre/post model — and use this feedback to make needed

adjustments to the evaluation system.

**Conclusion**

The importance of the teacher' preparation and effectiveness for student

achievement has been researched in many formats and by numerous researchers over the

decade previous to this writing.  The link between the effectiveness of the teacher and

student learning is undeniable (Bracey, 2004; Goodwin, 2011; Hattie, 2009; Marzano,

2007; Rivkin et al., 2005; Rockoff et al., 2008; Tucker & Stronge, 2005; Wright et al.,

1997).  Research confirmed the lasting impact of a good teacher beyond academics and

for several consecutive years (Chetty et al., 2014; Hanushek & Rivkin, 2012; Schmoker,

2011).   In light of these understandings, experts reached a consensus that the purpose of

teacher evaluations is to improve teaching and learning for all students.  These findings

served as the impetus for federal and state government mandates for schools to improve

their teacher evaluation systems, using multiple measures to evaluate teachers' impact on

students, including student assessment data.  For many districts, including the one

comprised in this research study, this mandate required a complete overhaul of the

teacher evaluation system.  In accordance with the then-current best practices for the

evaluation of teachers, the researched school district established the goal for the revised

teacher evaluation system to serve as a tool for improving classroom instruction and student learning.

The purpose of this study was to examine the relationship between a specific teacher evaluation rating and student achievement.   To accomplish this, student assessment data was collected from the researched district's six general elementary classrooms, and teacher evaluation rating data was collected for the teachers of these same classrooms.  This study also examined teacher and administrator perceptions of the evaluation tool's ability to meet its intended goals of improving both classroom instruction and student learning.

A *t*-test for difference in dependent means was conducted to compare the mean percentage of students performing Proficient (On-grade Level) and Advanced (Above grade level) on the MAP assessment over the two years of the study.   A PPMCC was calculated comparing the overall mean percentage scores in the two achievement categories (in the content areas of communication arts and mathematics) to detect whether a relationship existed between teacher's Follow-Up evaluation ratings and student achievement levels on the MAP and i-Ready assessments.  The results of the study demonstrated that, while there was an overall increase in student achievement after the first two years of implementation, there was not a statistically significant relationship between a teacher's follow-up rating and the assessment scores of his or her students in English language arts and mathematics on either the MAP or i-Ready assessments.   The results of the survey demonstrated a disconnect between how teachers and administrators perceived the ability of the revised system to meet its intended goals, with administrators feeling much more positively than teachers that it was a valid measurement tool.

The conclusions of this study suggest three implications for practice.  First, the district should continue to evaluate, monitor, and make modifications to the teacher evaluation system, based on issues that arose with its initial implementation and concerns addressed throughout its first two years of application.  Next, due to the complexity of isolating an individual teacher's direct impact on his or her students' learning, additional research should be conducted seeking a valid and reliable evaluation tool.   Finally, identifying the effectiveness of teachers should include student assessment data in addition to other measures of a teacher's instructional quality.  More research is needed to determine how to incorporate these multiple measures into a final rating.

## References

Aldeman, C., & Chuong, C. (2014). *Teacher evaluations in an era of rapid change. From "unsatisfactory" to "needs improvement."* Bellwether Education Partners. Retrieved from http://bellwethereducation.org/sites/default/files/Bellwether_ TeacherEval_Final_Web.pdf

American Educational Research Association. (2015). AERA statement on use of value added models (VAM) for the evaluation of educatiors and educator preparation programs. *Educational Researcher, 44*(8), 448-452. Retrieved from http://edr. sagepub.com/content/early/2015/11/10/0013189X15618385.full.pdf+html

American Federation of Teachers. (2012). *It's elemental: A quick guide to implementing evaluation and development systems.* [Issue brief]. Washington, DC: Author. Retrieved from the American Federation of Teachers website: http://www.aft.org/ sites/default/files/tde_ie_itselemental_2012.pdf

American Federation of Teachers. (2015). *Moving beyond compliance: Lessons learned from teacher development and evaluation.* [Issue brief]. Washington, DC: Author. Retrieved from the American Federation of Teachers website:

American Statistical Association. (2014, April). ASA statement on using value-added models for educational assessment. Retrieved from http://www.aft.org/sites/ default/files/i3_movingbeyondtcompliance2015.pdfhttp://www.amstat.org/policy/ pdfs/asa_vam_statement.pdf

Amrein-Beardsley, A., Collins, C., Polasky, S. A., & Sloat, E. F. (2013). Value-added model (VAM) research for educational policy: Framing the issue. *Educational*

*Policy Analysis Archives, 21*(4). Retrieved from http://files.eric.ed.gov/fulltext/ EJ1015315.pdf

The Aspen Institute. (2015). *Roles and responsibilities of the State Education Agency: Discussion guide for state leaders.* Washington, DC: Author. Retrieved from http://www.aspendrl.org/portal/browse/DocumentDetail?documentId=2846&download

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D. . . . Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers* (Briefing Paper #278). Washington, DC: Economic Policy Institute. Retrieved from http://www.epi.org/publication/bp278/

Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher, 44*(2), 77-86. Retrieved from http://journals. sagepub.com/doi/pdf/10.3102/0013189X15574904

Banks, T. (2015, January, 22). Improving teacher evaluation in 2015 [Web log post]. Retrieved from http://tntp.org/blog/post/improving-teacher-evaluation-in-2015

Barton, P. E., & Coley, R. J. (2009). Parsing the achievement gap II. Policy Information Report. *Educational Testing Service.* Retrieved from http://files.eric.ed.gov/full text/ED505163.pdf

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

Berk, R. A., (2005).  Survey of 12 strategies to measure teaching effectiveness.

    *International Journal of Teaching and Learning in Higher Education. 17*(1), 48-

    62.  Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.

    454.3400&rep=rep1&type=pdf

Berliner, D. (2009).  Poverty and potential:  Out-of-school factors and school success.

    East Lansing, MI:  The Great Lakes Center for Education Research and Practice.

    Retrieved from http://educationnext.org/in-schools-teacher-quality-matters-most-

    coleman/

Berliner, D. (2013). Effects of inequality and poverty vs. teachers and schooling on

    America's youth.  *Teachers College Record, 115*(12), 1-26.

Betebenner, D. (2009).  Norm-and criterion-referenced student growth. *Educational

    Measurement: Issues and Practice, 28*(4), 42-51.

Betebenner, D. W. (2011). A technical overview of the student growth percentile

    methodology:  Student growth percentiles and percentile growth projections/

    trajectories.  The National Center of Educational Assessment.  Retrieved from

    http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf

Bill & Melinda Gates Foundation.  (2013).  *Ensuring fair and reliable measures of

    effective teaching:  Culminating findings from the MET Project's three-year

    study.*  Seattle: WA: Author.  Retrieved from http://metproject.org/downloads/

    MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Black, P., & Wiliam, D., (2004).  Working inside the black box:  Assessment for learning

    in the classroom.  *Phi Delta Kappan, 86*(1), 9-21.  Retrieved from http://datause.

    cse.ucla.edu/DOCS/pb_wor_2004.pdf

Bracey, G. (2004, December). Value-added assessment findings: Poor kids get poor

teachers. *Phi Delta Kappan , 86*(4), 331-333.

Braun, H., Chudowsky, N., & Koenig, J. (Eds.). (2010). Getting value out of value added:

Report of a workshop. Committee on Value-Added Methodology for Instructional

Improvement, Program Evaluation, and Accountability. Washington, D.C.: The

National Academies Press.  Retrieved from http://rlhoover.people.ysu.edu/Class

Connections/core-readings/Getting%20Value%20Out%20of%20Value-

Added.pdf

Burnett, A., Cushing, E., & Bivona, L. (2012).  Uses of multiple measures for

performance-based comprehension.  *Center for Educator Compensation Reform.*

Retrieved from http://files.eric.ed.gov/fulltext/ED533704.pdf

Campbell, J., Kyriakides, L., Muijs, D., & Robinson, W. (2012). *Assessing teacher

effectiveness: Different Models*.  New York, NY: Routledge.

Center for American Progress, & Council of Chief State School Officers. (2014). Next-

*generation accountability systems:  An overview of current state policies and

practices.*  Washington, DC:  Author.  Retrieved from https://cdn.american

progress.org/wp-content/uploads/2014/10/Accountability-report.pdf

Center for Public Education. (2005). *Teacher quality and student achievement: Research

review.* Washington, DC:  Author.  Retrieved from http://www.centerforpublic

education.org/Main-Menu/Staffingstudents/Teacher-quality-and-student-

achievement-At-a-glance/Teacher-quality-and-student-achievement-Research-

review.html

Center for Public Education. (2012). *Teaching the teachers:  Effective professional*

    *development in an era of high stakes accountability.*  Washington, DC:  Author.

    Retrieved from http://www.centerforpubliceducation.org/Main-

    Menu/Staffingstudents/Teaching-the-Teachers-Effective-Professional-

    Development-in-an-Era-of-High-Stakes-Accountability/Teaching-the-Teachers-

    Full-Report.pdf

Center for Public Education.  (2013). Trends *in teacher evaluation:  How states are*

    *measuring performance.*  Washington, D.C.: Author.  Retrieved from

    http://www.centerforpubliceducation.org/Main-Menu/Evaluating-

    performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-

    Evaluation-Full-Report-PDF.pdf

Center for Research on Learning and Teaching.  (2016). Guidelines for evaluating

    teaching.  Retrieved from http://www.crlt.umich.edu/ tstrategies/guidelines

Chetty, R., Friedman, J., & Rockoff, J. (2011). *The long-term impacts of teachers:*

    *Teacher value-added and student outcomes in adulthood.* [NBER working paper

    No. 17699].  Retrieved from the National Bureau of Economic Research website:

    http://www.nber.org/papers/w17699.pdf

Chetty, R., Friedman, J. N., & Rockoff, J. (2014). Measuring the impacts of teachers:

    Ealuating bias in teacher value-added estimates. *American Economic Review,*

    *104*(9), 2593-2632.  Doi: 10.1257/aer.104.9.2593

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014).  What makes great teaching?

    Review of the underpinning research.  *Sutton Trust.*  Retrieved from http://www.

suttontrust.com/wp-content/uploads/2014/10/What-Makes-Great-Teaching-
REPORT.pdf

Cogan, M. L. (1973).  Clinical supervision.  Boston, MA:  Houghton Mifflin Co.

Cohen, J., & Goldhaber, D. (2016). Observations on evaluating teacher performance:
Assessing the strengths and weaknesses of classroom observations and value-
added measures.  In J.  A. Grissom & P. Youngs (Eds.), *Improving teacher
evaluation systems* (pp. 8-21).  New York, NY: Teachers College, Columbia
University.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. . . . York,
R. (1966).  *Equality of Educational Opportunity:  The Coleman Report.*
Washington, DC: D.C.L Government Printing Office.  Retrieved from http://files.
eric.ed.gov/fulltext/ED012275.pdf

Council of Chief State School Officers. (2008). *Performance expectations and indicators
for education leaders:  An ISLLC-Based guide to implementing leader standards
as a companion guide to educational leadership policy standards: ISLLC 2008*.
Washington, D.C.: Author.  Retrieved from http://www.ccsso.org/Documents/
2008/Peformance_Indicators_2008.pdf

Council of Chief State School Officers.  (2013, April).  Interstate teacher assessment and
support consortium.  In*TASC model core teaching standards and learning
progressions for teachers 1.0: A resource for ongoing teacher development*.
Washington, DC:  Author.  Retrieved from http://www.ccsso.org/documents/
2013/2013_intasc_learning_progressions_for_teachers.pdf

Council of Chief State School Officers.  (2016). *Principles for teacher support and*

      *evaluation systems*.  Washington, DC: Author.  Retrieved from http://www.ccsso.

      org/Documents/2016/Principles%20for%20Teacher%20Support%20and%20Eval

      uation%20Systems.pdf

Corcoran, S. P., (2010).  *Can teachers be evaluatied by their student's test scores?*

      *Should they be? The use of value-added measures of teacher effectiveness in*

      *policy and practice.*  New York, NY:  Annenberg Institute for School Reform.

      Retrieved from http://files.eric.ed.gov/fulltext/ED522163.pdf

Corcoran, S. P., (2016). Improving teacher evaluation systems: Making the most of

      multiple measures.  In J. A. Grissom & P. Youngs (Eds.), *Potential pitfals in the*

      *use of teacher value-added data.*  (pp. 51-62).  New York, NY: Teachers College

      Press.

Curriculum Associates, (2014).  *The science behind i-Ready's adaptive diagnostic*

      [SETDA Position Paper].  Retrieved from the State Educational Technology

      Directors Association (SETDA) website: http://www.setda.org/ls2013/wp-

      content/uploads/sites/8/2014/12/Adaptive-Diagnostic-Science-SETDA.pdf

Daley, G., & Kim, L. (2010). *A teacher evaluation system that works.* [NIET working

      paper].  Retrieved from National Institute for Excellence in Teaching website:

      http://www.niet.org/assets/Publications/working-paper-teacher-evaluation.pdf?

      processed=1

Danielson, C. (1996).  *Enhancing professional practice: A framework for teaching.*

      Alexandria, VA:  Association for Supervision and Curriculum Development.

Danielson, C., (2001). New trends in teacher evaluation. *Educational Leadership. 58*(5),

      p. 12-15.  Retrieved from http://eds.a.ebscohost.com.gatekeeper2.lindenwood.

      edu/ehost/pdfviewer/pdfviewer?vid=6&sid=722c1e99-5517-4133-8eaa-

      c17226a625db%40sessionmgr4009&hid=4213

Danielson, C. (2011).  *Enhancing professional practice:  A framework for teaching,* 2nd

      ed..  Alexandria, VA:  Association for Supervision and Curriculum Development.

Danielson, C. (2012).  Observing classroom practice. *Educational Leadership, 70*(3), 32-

      37.  Retrieved from http://www.ode.state.or.us/wma/teachlearn/commoncore/

      observing-classroom-practice---edleadership-article.pdf

Danielson, C. & McGreal, T. L. (2000).  *Teacher evaluation:  To enhance professional

      practice.*  Alexandria, VA:  Association for Supervision and Curriculum

      Development.

Danielson Group. (2013). *The framework.* Retrieved from the Danielson Group website:

      http://danielsongroup.org/framework/

Darling-Hammond, L. (1990). Teacher evaluation in transition: Emerging roles and

      evolving methods. In J. Millman, & L. Darling-Hammond (Eds.) *The new teacher

      handbook of teacher evaluation: Assessing elementary and secondary school

      teachers* (pp. 17-32). Newbury Park, CA: Sage.

Darling-Hammond, L. (1997). Doing what matters most:  Investing in quality teaching.

      New York, NWL, National Commission on Teaching and America's Future.

      Retrieved from http://nctaf.org/wp-content/uploads/DoingWhatMattersMost.pdf

Darling-Hammond, L. (2000).  Teacher quality and student achievement:  A review of

state policy evidence.  *Education Policy Analysis Archives, 8*(1).  Retrieved from

http://epaa.asu.edu/ojs/article/view/392/515

Darling-Hammond, L. (2010).  *Evaluating teacher effectiveness:  How teacher

performance assessments can measure and improve teaching.*  Washington, DC:

Center for American Progress.  Retrieved from http://www.highered.nysed.gov/

TELDH.pdf

Darling-Hammond, L. (2012).  *Creating a comprehensive system for evaluating and

supporting effective teaching.*  Stanford, CA: Stanford Center for Opportunity

Policy in Education.  Retrieved from https://edpolicy.stanford.edu/sites/default/

files/publications/creating-comprehensive-system-evaluating-and-supporting-

effective-teaching.pdf

Darling-Hammond, L. (2013).  *Getting teacher evaluation right:  What really matters for

effectiveness and improvement.*  New Your, NY:  Teachers College Press.

Darling-Hammond, L. (2015).  Can value added add value to teacher evaluations?

*Educational Researcher*, *24*(2), 132-137. doi:10.3102/0013189X15575346

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J.  (2013).

Evaluating teacher evaluation.  *Phi Delta Kappan, 93*(6), 8-15.  Retrieved from

http://schottfoundation.org/sites/default/files/resources/evaluating-teacher-

evaluation-ldh.pdf

Darling-Hammond, L., Cook, C., Jaquith, A., & Hamilton, M. (2012).  *Creating a

comprehensive system for evaluating and supporting effective teaching.*  Stanford,

CA:  Stanford Center for Opportunity Policy in Education.  Retrieved from

https://edpolicy.stanford.edu/sites/default/files/publications/creating-

comprehensive-system-evaluating-and-supporting-effective-teaching.pdf

Darling-Hammond, L., & Snyder, J. (2015).  Accountability for resources and outcomes:

An introduction.  *Education Policy Analysis Archives, 23*(20), 1-5.  Retrieved

from http://www.educationjustice.org/newsletters/ej_newsblast_150617_

Hammond-Snyder.pdf

Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014).  Accountability for college

and career readiness:  Developing a new paradigm.  *Education Policy Archives,

22*(86), 1.  Retrieved from file:///Users/maggiemathus/Downloads/1724-6068-2-

PB.pdf

Davies, B. (2008).  Accountable teacher evaluation.  *International Journal of

Educational Management, 22*(1), 107-107.

Diaz-Bilello, E. & Briggs, E. (2014).  *Using student growth percentiles for educator

evaluations at the teacher level:  Key issues and technical considerations for

school districts in Colorado.* Retrieved fromhttp://www.nciea.org/ publication

_PDFs/GrowthPercentileReport%20EDB073114.pdf

Doherty, K. M., & Jacobs, S. (2013).  State of the states 2013 connect the dots:  Using

evaluations of teacher effectiveness to inform policy and practice.  *National

Center on Teacher Quality.*  Retrieved from http://www.nctq.org/dmsView/State_

of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

Donaldson, M. L., (2012).  *Teachers' perspectives on evaluation reform*.  Center for

American Progress. Retrieved from http://files.eric.ed.gov/fulltext/ED539750.pdf

Donaldson, M. L., & Papay, J. P. (2012).  *Reforming teacher evaluation:  One district's story*.  Center for American Progress.  Retrieved from http://files.eric.ed.gov/full text/ED539748.pdf

Donaldson, M. L., & Peske, H. G. (2010).  Supporting effective teaching through teacher evaluation:  A study of teacher evaluation in five charter schools.  *Center for American Progress.*  Retrieved from https://www.americanprogress.org/issues/ education/reports/2010/03/10/7381/supporting-effective-teaching-through- teacher-evaluation/

Eaton, J. S. (2010).  Acredidation and the federal future of higher education.  *Academe*, *96*(5), 21.  Retrieved from the American Association of University Professors [aaup] site:  http://www.aaup.org/article/accreditation-and-federal-future-higher- education#.VyAIEqMrJsM

Education Trust.  (2012, January).  *Report: Teaching counts:  Recommendations for reforming California's teacher evalaution system.*  Oakland, CA: Author. Retrieved from http://west.edtrust.org/wp-content/uploads/sites/3/2015/01/ Teaching-Counts.pdf

Education Trust. (2016). The Every Student Succeeds Act:  What's in it?  What does it mean for equity? Overview.  Washington, DC:  Author.  Retrieved from https:// edtrust.org/resource/the-every-student-succeeds-act-whats-in-it-what-does-it- mean-for-equity/

Eisenbach, B. B., (2014). Words that encourage.  *Educational Leadership*, *71*(5), 70-72. Retrieved from http://eds.a.ebscohost.com.gatekeeper2.lindenwood.edu/ehost/pdf

viewer/pdfviewer?sid=205cbc39-4533-49ac-aed7-0f1b42454b93%40sessionmgr

4004&vid=4&hid=4210

Fetters, J. (2013).  *High fidelity:  Investing in evaluation training.*  Retrieved from

Center on Great Teachers and Leaders website:  http://www.gtlcenter.org/sites/

default/files/docs/GTL_AskTeam_HighFidelity.pdf

Fraenkel, J. R, & Wallen, N. E. (2006).  *How to design and evaluate research in*

*education.*  Boston, MA:  McGraw-Hill.

Galey, S. (2015), Education politics and policy: Emerging institutions, interests and

ideas.  *Policy Studies Journal*, 43(1), S12 - S39. doi: 1111/psj.12100.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W. . . . Zhu, P. (2008).

*The impact of two professional development interventions on early reading*

*instruction and achievement.  NCEE 2008-4030.*  National Center for Education

Evaluation and Regional Assistance.  Retrieved from http://files.eric.

ed.gov/fulltext/ED502700.pdf

Glazerman, S., Loeb, S., Goldhaber, D. D., Raudenbush, S., & Whitehurst, G. J. (2010).

*Evaluating teachers:  The important role of value-added* (Vol. 201. No. 0).

Washington, DC:  Brown Center on Education Policy at Brookings.  Retrieved

from http://www.leg.state.vt.us/WorkGroups/EdOp/Brookings%20Value%20

ADDED1117_evaluating_teachers.pdf

Goe, L. (2007).  *The link between teacher quality and student outcomes:  A research*

*synthesis*.  National Comprehensive Center for Teacher Quality.  Retrieved from

http://files.eric.ed.gov/fulltext/ED521219.pdf

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality.  Retrieved from http://files.eric.ed.gov/fulltext/ED521228.pdf

Goe, L., & Croft, A. (2009).  *Methods of evaluating teacher effectiveness.  Research-to-Practice Brief.*  National Comfrehensive Center for Teacher Quality.  Retrieved from  http://www.gtlcenter.org/sites/default/files/docs/RestoPractice_Evaluating TeacherEffectiveness.pdf

Goe, L., & Holdheide, L. (2011).  *Measuring teachers' contributions to student learning growth for nontested grades and subjects* [Research & Policy Brief]. Washington, DC: National Comprehensive Center for Teacher Quality.  Retrieved from http:// www.gtlcenter.org/sites/default/files/docs/MeasuringTeachersContributions.pdf

Goe, L., Holdheide, L., & Miller, T. (2014).  *Practical guide to designing comprehensice teacher evaluation systems.*  [Revised edition].  Washington, DC: Center on Great Teachers & Leaders.  Retrieved from http://files.eric.ed.gov/fulltext/ED 555655.pdf

Goodwin, B. (2011). *Simply better: Doing what matters most to change the odds for student success.* Alexandria, VA: Association for Supervision and Curriculum Development.

Goldhaber, D. (2016).  In schools, teacher quality matters most.  *Education Next, 16*(2). Retrieved from http://educationnext.org/in-schools-teacher-quality-matters-most-coleman/

Goldhammer, R., Anderson, R. H. & Krajewski, R. J.  (1980). *Clinical supervision:*

    *Special methods for the supervision of teachers,* 2nd ed.  New York, NY: Holt,

    Rinehart, and Winston, Inc.

Grant, L., Hindman, J., & Stronge, J. (2013).  *Planning, instruction, and assessment:*

    *Effective teaching practices.*  Larchmont, NY: Routledge.

Grissom, J. A., Loeb, S., & Doss, C. (2016).  The multiple dimensions of teacher quality:

    Does value-added capture teachers' non-achievement contributions to their

    schools?  In J.A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation*

    *systems* (pp. 37-50).  New York: NY: Teachers College Press.

Grissom, J. A., & Youngs, P. (2016). Making the most of multiple measures.  In J. A.

    Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems* (pp. 1-7).

    New York, NY: Teachers College Press.

Guggenheim, D, Kimball, B., Chilcott, L, Strickland, B., Canada, G., Rhee, M. . . .

    Paramount Home Entertainment (Film).  (2011). *Waiting for "Superman".*

    Hollywood, CA: Paramount Home Entertainment.

Haertel, E. H. (2013).  *Reliability and validity of inferences about teachers based on*

    *student test scores*.  Educational Testing Service.  Retrieved from

    https://www.ets.org/Media/Research/pdf/PICANG14.pdf

Hall, E., Diaz-Bilello, E., & Marion, S. (2015). *Considerations for establishing*

    *performance standards for educator evaluation systems*. Retrieved from http://

    www.nciea.org/publication_PDFs/Establishing%20Performance%20Standards%2

    0for%20EES_EH2015.pdf

Hanushek, E. A. (2014). Boosting teacher effectiveness. In C. E. Finn, Jr. & R. Sousa
(Eds.), *What lies ahead for America's children and their schools* (pp. 23-35).
Stanford, CA: Hoover Institution Press.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for
teacher quality.* (NBER working paper No. 11154).  Retrieved from the  National
Bureau of Economic Research website: http://www.nber.org/papers/w11154

Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and
implications for policy. *Annual Review of Economics, 4*(1), 131-157.
doi:10.1146/ annurev-economics-080511-111001.

Hansen, M., Lemke, M., & Sorensen, N. (2013).  *Combining multiple performance
measures:  Do common approaches undermine districts' personnel evaluation
systems?* Retrieved from American Institutes for Research website: http://www.
air.org/sites/default/files/downloads/report/Combining_Multiple_Performance_M
easures_0.pdf

Harris, H. N., & Herrington, C. D. (Eds). (2015). Value added meets the schools:  The
effects of using test-based teacher evaluation on the work of teachers and leaders
[Special issue].  *Educational Researcher, 44*(2), 71-76.

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014).  How teacher evaluation methods
matter for accountability:  A comparative analysis of teacher effectiveness ratings
by principals and teacher value-added measures.  *American Educational Research
Journal, 51*(1), 73-112.  Doi:10.3102/0002831213517130

Hattie, J. (2003, October). *Teachers make a difference: What is the research evidence?*
Paper presented at the Austrailian Council for Educational Research Annual

Conference on Building Teacher Quality, Melbourne, Australia.  Abstract retrieved from http://www.educationalleaders.govt.nz/Pedagogy-and-assessment/Building-effective-learning-environments/Teachers-Make-a-Difference-What-is-the-Research-Evidence

Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Hattie, J. A. (2012).  *Visible learning for teachers:  Maximizing impact on learning.*  New York, NY: Routledge.

Hattie, J. (2015).  *What doesn't work in education:  The politics of distraction.*  London, England:  Pearson.  Retrieved from https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/hattie/150602_DistractionWEB_V2.pdf

Hattie, J., & Anderman, E. M. (Eds). (2013).  *International guide to student achievement.*  New York, NY: Routledge.

Henry, G. T., & Guthrie, J. E. (2016).  Using multiple measures for development teacher evaluation.  In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems* (pp. 143-155).  New York, NY: Teachers College, Columbia University.

Hess, F. M., & Fullerton, J. (2010).  The numbers we need:  How the right metrics could improve K-12 education. *Education Outlook,* (2), 1-8.  Retrieved from https://www.aei.org/publication/the-numbers-we-need-how-the-right-metrics-could-improve-k-12-education/

Hinchey, P. H. (2010).  *Getting teacher assessment right:  What policymakers can learn from research.*  Boulder, CO:  National Education Policy Center.  Retrieved from http://files.eric.ed.gov/fulltext/ED513908.pdf

Hull, J. (2013a).  Trends in teacher evaluation:  At a glance [Report Summary].

> Retrieved from Center for Public Education website: http://www.centerforpublic

> education.org/teacherevalreview

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

> *performance.*  [Research Report].  Retrieved from Center for Public Education

> website:  http://www.centerforpubliceducation.org/Main-Menu/Evaluating-

> performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-

> Evaluation-Full-Report-PDF.pdf

Hunter, M. (1982).  *Mastery teaching.*  Thousand Oaks, CA: Corwin Press.

Jacob, B. A., & Lefgren, L. (2008).  Can principals identify effective teachers?  Evidence

> on subjective performance evaluation in education.  *Journal of Labor Economics,*

> *26*(1), 101-136.

Jerald, C. D. (2012). *Movin' it and improvin' it!  Using both education strategies to*

> *increase teaching effectiveness.*  Washington, DC:  Center for American Progress.

> Retrieved from http://files.eric.ed.gov/fulltext/ED535645.pdf

Jordan, H. R., Mendro, R. L., & Weerasinghe, D.  (1997).  *Teacher effects on*

> *longitudinal student achievement:  A report on research in progress.* Paper

> presented at the National Evaluation Institue (July, 1997), Indianapolis, IN.

> Retrieved from http://www.dallasisd.org/cms/lib/TX01001475/Centricity/Shared/

> evalacct/research/articles/Jordan-Teacher-Effects-on-Longitudinal-Student-

> Achievement-1997.pdf

Jorgenson, O., (2012).  What we lose in winning the test score race:  Value-added

assessment models could be a win-win solution.  *Principal, 9*(5), 12-15.

Retrieved from http://www.naesp.org/sites/default/files/MJ12%20Jorgenson.pdf

Kane, T. J.,  & Staiger, D. O. (2008).  *Estimating teacher impacts on student

achievement:  An experimental evaluation.*  Cambridge, MA: National Bureau of

Economic Resarch,.  Retrieved from http://www.nber.org/papers/ w14607.pdf

Kane, T. J., & Staiger, D. O. (2012, January).  *Gathering feedback for teaching:

Combining high-quality observations with student surveys and achievement gains*

[Research paper].  Seattle, WA:  Bill and Melinda Gates Foundation.  Retrieved

from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research

_Paper.pdf

Kane, T. J., Taylor, E. S., Tyler, J. H, & Wooten, A. L. (2011a, Summer).  Evaluating

teacher effectiveness:  Can classroom observations identify practices that raise

achievement? *Education Next, 11*(3), 54-60.  Retrieved from http://education

next.org/files/ednext_20113_research_kane.pdf

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011b). Identifying effective

classroom practices using student achievement data. *Journal of human

Resources*, *46*(3), 587-613.

Klein, J. (2011, June). The failure of American schools.  *The Atlantic*.  Retrieved from

http://www.theatlantic.com/magazine/archive/2011/06/the-failure-of-american-

schools/308497/

Koretz, D. (2000). *Limitations in the use of achievement tests as measures of educators'*

    *productivity*. RAND Corporation. Retrieved from http://www.irp. wisc.edu/news

    events/other/symposia/koretz.pdf

Krasnoff, B. (2014).  *What the research says about class size, professional development,*

    *and recruitment, induction, and retention of highly qualified teachers:  A*

    *compendium of the evidence on Title II, part A, program—funded strategies*.

    Retrieved from Education Northwest website: http://www.k12.wa.us/titleiia/

    program/pubdocs/CompendiumofT2AStrategies.pdf

Labaree, D. F. (2008).  *An uneasy relationship: the history of teach eduation in the*

    *university*.  Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=

    10.1.1.462.6342&rep=rep1&type=pdf

Ladd, H. F. (2008).  Teacher effects:  What do we know?  In G. Duncan & J. Spillane

    (Eds.), *Teacher quality:  Broadening and deepening the debate* (pp. 3-12).

    Evanston, IL:  Northwestern University.

Ladd, H. F. (2012).  Education and poverty:  Confronting the evidence.  *Journal of Policy*

    *Analysis and Management, 31*(2), 203-227.

Levin, C. (Producer/Director).  (2001).  *Only a teacher* [video recording]. Princeton, NJ:

    Films for the Humanities and Sciences.  Retrieved from PBS.org website:

    http://www.pbs.org/onlyateacher/index.html

Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher*

    *effectiveness*.  *National Comprehensive Center for Teacher Quality*.  Retrieved

    from http://files.eric.ed.gov/fulltext/ED543776.pdf

Linn, R. (2011).  Test-based accountability.  *The Gordon Commission on the future of*

   *assessment in education.*  Retrieved from http://www.gordoncommission.net/

   rsc/pdfs/linn_testbased_accountability.pdf

Looney, J. W. (2011). Integrating formative and summative assessment: progress toward

   a seamless system. (OECD Education Working Papers, No. 58), Retrieved from

   OECD Publishing: http://dx.doi.org/10.1787/5kghx3kbl734-en

Marshall, K. (2012). Teacher evaluation: What's fair? What's effective? *Educational*

   *Leadership, 70*(3), p. 50-53.  Retrieved from http://www.marshallmemo.com/

   articles/Ed-Leadership-Nov-2012.pdf

Marzano, R. J. (2003).  *What works in schools:  Translating research into action.*

   Alexandria, VA:  Association for Supervision and Curriculum Development.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for*

   *effective instruction.* Alexandria, VA: Association for Supervision and

   Curriculum Development.  Retrieved from http://www.ascd.org/

   publications/books/110019/chapters/A-Brief-History-of-Supervision-and-

   Evaluation.aspx

Marzano, R. (2012, November). The two purposes of teacher evaluation. *Educational*

   *Leadership , 70* (3), 14-19.  Retrieved from http://www.ascd.org/publications/

   educational-leadership/nov12/vol70/num03/The-Two-Purposes-of-Teacher-

   Evaluation.aspx

Marzano, R. J., Frontier, T., & Livingstone, D. (2011).  *Effective supervision:*

   *Supporting the art and science of teaching.*  Alexandria, VA:  Association for

   Supervision and Curriculum Development.

Marzano, R. J., & Toth, M. D. (2013). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Toth, M., & Schooling, P. (2012). *Examining the role of teacher evaluation in student achievement: Contemporary research base for the Marzano Causal Teacher Evaluation Model*. Retrieved from the Marzano Center website: http://www.marzanocenter.com/files/MC_White_Paper_20120424.pdf/

Maslow, V. J., & Kelley, C. J. (2012).  Does evaluation advance teaching practice?  The effects of performance on teaching quality and system change in large diverse high schools.  *Journal of School Leadership, 22*(3), 600-632.

Mathers, C., Olivia, M., & Laine, S. W. M. (2008). *Improving instruction through effective teacher evaluation:  Options for states and districts* [TQ Research & Policy Brief].  Washington, DC: National Comprehensive Center for Teacher Quality.  Retrieved from http://files.eric.ed.gov/fulltext/ED520778.pdf

McCaffrey, D. F., Koretz, D, Lockwood, J. R., & Hamilton, L. S. (2004).  The promise and peril of using value-added modeling to measure teacher effectiveness.  Santa Monica, CA:  RAND Corporation.  Retrieved from http://www.rand.org/content/dam/rand/pubs/research_briefs/2005/RAND_RB9050.pdf

McCaffrey, D. F., Lockwood, J. R., Koretz, D. & Hamilton, L. S. (2003).  *Evaluating value-added models for teacher accountability.* Santa Monica, CA:  RAND Corporation. Retrieved from http://www.rand.org/pubs/monographs/MG158.html

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal

variability of teacher effect estimates. Education and Finance Policy, *4*(4), 572-

606.  Retrieved from http://utla.net/system/files/mccaffrey_study.pdf

McGuinn, P. (2012). *The state of teacher evaluation reform:  State education agency*

*capacity and the implementation of new teacher-evaluation systems.*  Center for

American Progress.  Retrieved from http://files.eric.ed.gov/fulltext/ED539744.pdf

McGuinn, P. (2015). *Evaluating progress: State education agencies and the*

*implementation of new teacher evaluation systems*. [White Paper] (#WP2015-09).

Philadelphia, PA: Consortium for Policy Research in Education, University of

Pennsylvania.  Retrieved from http://www.cpre.org/sites/default/files/wp_

mcguinn_2015.pdf

McMillan, J. H., Mohn, R. S., & Hammack, M. V. (2013).In L. H. Meyer (Ed.), Oxford

bibliographies in education. New York, NY: Oxford University Press.

McRel Industries. (2009). *What matters most.*  Aurora, CO: Author.  Retrieved from

http://www.mcrel.org/about-us/~/media/72F07B06A80E4FBF90E512A6DC

15CE28.ashx

Measures of Effective Teaching. (2013). *Ensuring fair and reliable measures of effective*

*teaching: Cumulating findings from the MET Project's three-year study*. Bill and

Melinda Gates Foundation.  Retrieved from http://www.metproject.org/down

loads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Melvin, L. (2011). How to keep good teachers and principals:  Practical solutions to

today's classroom problems.  Lanham, MD: Rowan & Littlefield Publishers, Inc.

Retrieved from http://www.slideshare.net/alfeliz/lonnie-melvin-howtokeepgoodteachersandprinbookfiorg

Milanowski, A. T., Kimball, S. M., & White, B. (2004).  *The relationship between standards-based teacher evaluation scores and student achievement: Replication and extensions at three sites*.  CPRE-UW Working Paper Series TC-04-01, Madison, WI: University of Wisconsin-Madison, Wisconsin.  Centr for Education Research, Consortium for Policy Research in Education.  Retrieved from http://www.cpre-wisconsin.org/papers/3site_long_TE_SA_AERA04TE.pdf

Milner, H. R. (2010). *Start where you are but don't stay there:  Understanding diversity, opportunity gaps, and teaching in today's classrooms.*  Cambridge, MA:  Harvard Education Press.

Missouri Department of Elementary and Secondary Education. (2011).  *Missouri model teacher and leader standards:  A resource for state dialogue*.  Retrieved from Missouri Department of Elementary and Secondary Education website: https://dese.mo.gov/sites/default/files/StandardsInformationDocument.pdf

Missouri Department of Elementary and Secondary Education. (2013). *Missouri's educator evaluation system.* Jefferson City, MO: Author.  Retrieved from Missouri Department of Elementary and Secondary Education website http://dese.mo.gov/educator-quality/educator-effectiveness/educator-evaluation-system

Missouri Department of Elementary and Secondary Education. (2014).  *Missouri Assessment Program grade-level assessments: Technical Report 2014 Final.* Jefferson City, MO: Author.  Retrieved from Missouri Department of Elementary

and Secondary Education website https://dese.mo.gov/sites/default/files/asmt-gl-2014-tech-report.pdf

Missouri Department of Elementary and Secondary Education.  (2015). Grade level. Retrieved from https://dese.mo.gov/college-career-readiness/assessment/grade-level

Missouri Department of Elementary and Secondary Education.  (2016).  *LEA guide to the Missouri Assessment Program 2016-2017 v1.*  Retrieved from https://dese.mo. gov/sites/default/files/asmt-guide-to-missouri-assessment-program_1617.pdf

Missouri School District.  (2014). *Teacher evaluation system powerpoint*. High Ridge, MO:  author.

Missouri School District. (2015). *Teacher evaluation manual.*  High Ridge, MO:  author.

Missouri Senate. (2009). Missouri Senate Bill 291, HSC #2 §§ SB 291. Retrieved from http://www.senate. mo,gov/09info/BTS_Web/Bill.aspx?Session Type=R&BillID =683252

Murphy, J., Hallinger, P., & Heck, R. H. (2013).  Leading via teacher evaluation:  The case of the missing clothers?  *Edicational Researcher, 42*(6), 349-354.  Retrieved from file:///Users/maggiemathus/Desktop/EDUCATIONAL%20RESEARCHER-2013-Murphy-349-54.pdf

National Academy of Engineering, Steering Committee for Evaluating Instructional Scholarship in Engineering. (2009).  *Developing metrics for assessing engineering instruction:  What gets measured gets improved.*  Retrieved from The National Academies Press website: http://www.nap.edu/read/12636/chapter/1#iv

National Commission on Teaching and America's Future. (1996).  *What matters most: Teaching for America's future.*  New York, NY:  Author.  Retrieved from http:// nctaf.org/wp-content/uploads/WhatMattersMost.pdf

National Conference of State Legislators. (2015).  *Summary of the Every Student Succeeds Act, legislation reauthorizing the Elementary and Secondary Education Act.*  Retrieved from the National Conference of State Legislators website: http://www.ncsl.org/documents/capitolforum/2015/onlineresources/summary_12_ 10.pdf

National Council on Teacher Quality. (2011). *State of the states: Trends and early lessons on teacher evaluation and effective policies.*  National Center on Teaching Quality website. Retrieved from http://www.nctq.org/dmsView/State_of_the_ States_Teacher_Evaluation_and_Effectiveness_Policies_NCTQ_Report

National Council on Teacher Quality. (2013). *State of the states 2013 connect the dots: Using evaluations of teacher effectiveness to inform policy and practice.*  National Center on Teaching Quality website. Retrieved from http://www.nctq.org/dms View/ State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

National Council on Teacher Quality. (2015). *State of the Stats 2015:  Evaluating teaching, leading, and learning.*  National Center on Teaching Quality website. Retrieved from http://www.nctq.org/dmsView/StateofStates2015

National Education Association. (1995). Summative evaluation and formative feedback. Washington, DC: Author.  Retrieved from http://www.nea.org/assets/docs/HE/ upv1no5.pdf

National Education Association.  (2010). Teacher assessment and evaluation:  The

    National Education Association's framework for transforming education systems

    to support effective teaching and improve student learning [White paper].

    Washington, DC: Author.  Retrieved from http://www.nea.org/assets/docs/HE/

    TeachrAssmntWhtPaperTransform10_2.pdf

National Governors' Association. (2015). Governors seeking to ensure students' long-

    term success. Retrieved from https://www.nga.org/cms/home/news-room/news-

    releases/page_2005/col2-content/main-content-list/governors-seeking-to-ensure-

    stud.html

Nelson, J. A. (2012). *Effects of teacher evaluations on teacher effectiveness and student

    achievement.*  (Doctoral dissertation).  Retrieved from https://www.nmu.edu/

    education/sites/DrupalEducation/files/UserFiles/Nelson_Julie_MP.pdf

The New Teacher Project. (2010, May 25).  *Teacher Evaluation 2.0.*  Boulder, CO:

    National Education Policy Center.  Retrieved from http://tntp.org/assets/

    documents/Teacher-Evaluation-Oct10F.pdf

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110 § 115, Stat. 1425

    (2002).  Retrieved from https://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Odden, A. (2004).  Lessons learned about standards-baed teacher evaluation systems.

    *Peabody Journal of Education, 79*(4), 126-137.  Retrieved from https://www.

    researchgate.net/publication/248942763_Lessons_Learned_About_Standards-

    Based_Teacher_Evaluation_Systems

Okafor, P. (2012). *Leadersip in instructional supervision.*  Retrieved from http://patrick

    okafor.com/files/ClinicalSupervision.pdf

Organization for Economic Cooperation and Development. (2009). *Creating effective teaching and learning environments:  First results from TALIS- Executive Summary.*   Paris, France: OECD.  Retrieved from https://www.oecd. org/edu/school/43044074.pdf

Paufler, N. A., & Amrein-Beardsley, A. (2013). The random assignment of students into elementary classrooms:  Implications for Value-Added analyses and interpretations.  *American Educational Research Journal, 51*(2), 328-362.

Pellegrino, J. W. (2004).  *The evolution of educational assessment:  Considering the past and imagining the furture.*  Retrieved from Educational Testing Service, Policy Evaluation and Reseach center website:  https://www.ets.org/Media/Research/ pdf/PICANG6.pdf

Pennington, K. (2014).  *ESEA waivers and teacher-evaluation plans:  State oversight of district-designed teacher-evalaution systems*.  Center for American Progress. Retrieved from http://files.eric.ed.gov/fulltext/ED561085.pdf

Pianta, R. C., & Hamre, B. K., (2016). Implementing rigorous observations of teachers: Synchronizing theory with systems for implementation and support. In J. A. Grissom & P. Youngs (Eds), *Improving teacher evaluation systems* (pp. 22-36). New York, NY:  Teachers College, Columbia University

Piro, J., & Mullen, L. (2013).  Outputs as educator effectiveness in the United States: Shifting towards political accountability.  *International Journal of Educational Leadership Preparation, 9*(1), 59-77.  Retrieved from https://www.researchgate. net/profile/Piro_Jody/publication/267034789_Outputs_as_educator_effectiveness

_in_the_United_States_Shifting_towards_political_accountability/links/54427ac5

0cf2a76a3ccb005d.pdf

Popham, W. J. (2013).  *Evaluating Americ's teacher:  Mission possible?*  Thousand

Oaks, CA: Corwin Press

RAND Education. (2012a).  *Multiple choices:  Options for measuring teaching*

*effectiveness.*  RAND Corporation (Report No. PC-693/2 (09/12)).  Retrieved

from http://www.rand.org/education/projects/measuring-teacher-effectiveness/

multiple-choices.html

RAND Education. (2012b).  *Student growth percentiles 101:  Using relative ranks in*

*student test scores.*  RAND Corporation (Report No. CP-693/1 (09/12)).

Retrieved from http://www.rand.org/pubs/corporate_pubs/CP693z5-2012-09.html

Raudenbush, S. W. (2015).  Value added:  A case study in the mismatch between

education research and policy.  *Educational Researcher, 44*(2), 138-141.

Ravitch, D. (2002). Testing and accountability, historically considered.  In W. M. Evers

& H. J. Walberg (Eds.), *School accountability:  An assessment by the Koretz task*

*force on K-12 education* (pp. 9-22).  Palo Alto, CA: Hoover Institution, Stanford

University.

Ravitch, D. (2010).  *The death and life of the great American school system: How testing*

*and choice are undermining education*.  New York, NY:  Basic Books.

Reardon, S. F. (2011).  *The widening academic achievement gap between the rich and the*

*poor:  New evidence and possible explanations*. Retrieved from http://citeseerx.

ist.psu.edu/viewdoc/download?doi=10.1.1.359.7217&rep=rep1&type=pdf

Reform Support Network. (2013, July).  *Measuring teaching matters:  What different ways of looking at student results tell us about teacher effectiveness.*  [Issue Brief].  Retrieved from https://www2.ed.gov/about/inits/ed/implementation-support-unit/tech-assist/measuring-teaching-matters.pdf

Riordan, J., Lacireno-Paquet, N., Shakman, K., Bocala, C., & Chang, Q. (2015). *Redesigning teacher evaluation: Lessons from a pilot implementation* (REL 2015–030). Washington,DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

Ritter, G. W., & Shuls, J. V. (2012, November).  If a tree falls in a forest, but no one hears . . .  *Phi Delta Kappan, 94*(3), 34-38.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica , 73*(2),  417-458.  Retrieved from http://citeseerx.ist. psu.edu/viewdoc/download?doi=10.1.1.322.4872&rep=rep1&type=pdf

Robelen, E. (2013). Classroom portfolios used as alternative teacher-evaluation measure. *Education Week.* Retrieved from http://www.edweek.org/ew/articles/2013/09/ 18/04 arts_ep.h33.html

Rockoff, J. E. (2004).  The impact of individual teachers on student achievement: Evidence from panel data.  *The American Economic Review 94*(4), 247-252.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2008).  *Can you recognize an effective teacher when you recruit one?*  (NBER Working Paper No. 14485).

Cambridge, MA;  National Bureau o Economic Research.  Retrieved from

http://www.nber.org/papers/w14485.pdf

Rothman, R. (2009).  Behind the classroom door:  A rare glimpse indicates the extent—

and persistence—of variation in teacher practice.  *Harvard Education Letter*

*25*(6).  Cambridge, MA:  Harvard Education Publishing Group.  Retrieved from

http://hepg.org/hel-home/issues/25_6/helarticle/behind-the-classroom-door_427

Rothstein, J. (2008).  *Teacher quality in educational production: Tracking, decay, and*

*student achievement.* (NBER Working paper No. 1442).  Retrieved from National

Bureau of Economic Research website  http://www.nber.org/papers/w14442.pdf

Rothstein, J. (2011).  Review of "Learning about teaching:  Initial findings from the

Measures of Effective Teaching Project."  Boulder, CO: National Education

Policy Center.  Retrieved from http://nepc.colorado.edu/thinktank/review-

learning-about-teaching

Santiago, P., & Benavides, F. (2009). *Teacher evaluation: A conceptual framework and*

*examples of country practices*. Towards a Teacher Evaluation Framework in

Mexico: International Practices, Criteria and Mechanisms, Mexico.  Retrieved

from the Organization for Economic Co-operation and Development [OECD]

website: http://www.oecd.org/edu/school/44568106.pdf

Schmoker, M. (2011).  *Focus: Evaluating the essentials to radically improve student*

*learning.* Alexandria, VA: Association for Supervision and Curriculum

Development.

Schagen, I., & Hodgen, E. (2009).  *How much difference does it make? Notes on*

*understanding, using, and calculating effect sizes for schools.*  Retrieved from the

New Zealand Center for Educational Research (NZCER) website:  http://www. educationcounts.govt.nz/publications/schooling/36097/36098

Schooling, P., Toth, M., & Marzano, R. (2010).  *Creating an aligned system.*  Retrieved from http://www.iobservation.com/files/whitepapers/Marzano-Race-to-the-Top-White-Paper.pdf

Schul, J. E. (2011). Unintended consequences:  Fundamental flaws that plague the No Child Left Behind act.  *JEP: EJournal of Education Policy.*  Retrieved from http://nau.edu/uploadedFiles/Academic/COE/About/Projects/Unintended%20Con sequences.pdf

Shakman, K., Breslow, N., Kochanek, J., Riordan, J., & Haferd, T. (2012). *Changing cultures and building capacity: An exploration of district strategies for implementation of teacher evaluation systems*. Waltham, MA: Educational Development Center.  Retrieved from http://ltd.edc.org/sites/ltd.edc.org/files/ District%20Strategies%20for%20Implementation%20of%20Teacher%20Evaluati on%20Systems.pdf

Shakman, K., Riordan, J., Sanchez, M. T., Cook, K. D., Fournier, R., & Brett, J. (2012). An examination of performance-based teacher evaluation systems in five states. Issues & Answers. REL 2012-No. 129.  *Regional Educational Laboratory Northeast & Islands.*  Retrieved from http://files.eric.ed.gov/fulltext/ED529800. pdf

Schochet, P. Z. & Chiang, H. S. (2010).  *Error rates in measuring teacher and school performance based on student test score gains*.  NCEE 2010-4004.  National

Center for Education Evaluation and Regional Assistance.  Retrieved from

http://files.eric.ed.gov/fulltext/ED511026.pdf

Spring, J. (2001). *The American school 1642-2000 (5th ed.)*.  New York, NY: McGraw-

Hill Publishing Companies.

Stanford Center for Opportunity Policy in Education.  (2013).  *Excellent teachers for*

*each and every child:  A guide for state policy.*   Stanford, CA: Author.  Retrieved

from https://edpolicy.stanford.edu/sites/default/files/ publications/excellent-

teachers-each-and-every-child-guide-state-policy.pdf

Steinberg, M. & Donaldson, M. (2014).  *The new educational accountability:*

*Understanding the landscape of teacher evaluation in the post-NCLB era*.  [Policy

Brief].  Retrieved from http://cepa.uconn.edu/wp-content/uploads/sites/

399/2014/02/The-New-Educational-Accountability_policy-brief_8-19-14.pdf

Steinberg, M., & Garrett, R. (2016).  Classroom composition and measured teacher

performance:  What do teacher observation scores really measure? *Educational*

*Evaluation and Policy Analysis,* Doi: 10.3102/0162373715616249

Stronge, J. H. (1993). Evaluating teachers and support personnel. In B. S. Billingsley

(Ed.),  Program leadership for serving students with disabilities (pp. 445-464),

Richmond, VA: Virginial Department of Education.

Stronge, J. H. (2006).  Teacher evaluation and school improvement:  Improving the

educational landscape.  In J. H. Stronge (Ed.)  *Evaluating teaching:  A guide to*

*current thinking and best practice.* (pp. 1-23, 2nd ed).  Thousand Oaks, CA:

Corwin.

Stronge, J. (2013). *Evaluating what good teachers do: Eight research-based standards for assessing teacher excellence*. New York, NY: Routledge.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education , 62*(4), 339-355.  Retrieved from https://www.strongeandassociates.com/files/articles/Stronge%20et%20al%20201 1_What%20makes%20good%20teachers%20good.pdf

Taylor, E. S., & Tyler, J. H. (2011).  *The effect of evaluation on performance:  Evidence from longitudinal student achievement data of mid-career teachers*.  (NBER Working Paper No. 16877).  National Bureau of Economic Research.  Retrieved from http://www.nber.org/papers/w16877.pdf

Taylor, E. S. & Tyler, J. H. (2012).  The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628-51.  Doi: 10.1257/aer.102.7.3628

Toch, T., & Rothman, R. (2008).  *Rush to judgement: Teacher evaluation in public education.* Washington, DC: Education Sector.

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

U.S. Department of Education.  (2009). Race to the top program: Preamble and major changes. Washington, DC: Author.  Retrieved from http://www2.ed.gov/ programs/racetothetop/major-changes.pdf

U.S. National Commission on Excellence in Education.  (1983). A *nation at risk:  The imperative for educational reform: A report to the Nation and the Secretary of*

*Education, United States Department of Education.* Washington, D.C.: The

Commission.  Retrieved from http://www2.ed.gov/pubs/NatAtRisk/ index.html

Weber, K. (Ed.). (2010).  *Waiting for superman:  How we can save America's failing*

*public schools.* New York, NY: Public Affairs, 2010.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our

national failure to acknowledge and act on differences in teacher effectiveness.

Brooklyn, NY: New Teacher Project.  Retrieved from http://tntp.org/publications/

view/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness

Whitaker, T. (2013).  Help teachers be their best:  Sow the seeds for duplicating teacher

excellence.  *Principal.*  Retrieved from https://www.naesp.org/sites/default/files/

Whitaker_JF13.pdf

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context

effects on student achievement: Implications for teacher evaluation. *Journal of*

*Personal Evaluation in Education, 11*, 57-67.  Retrieved from http://bulldogcia.

com/Documents/Articles/Wright_Horn_Sanders_1997.pdf

Yoon, K. S., Duncan, T., Lee, S. W-Y., Scarloss, B., & Shapley, K. (2007).  Reviewing

the evidence on how teacher professional development affects student

achievement.  (Issues & Answers Report, REL 2007—No. 033).  Retrieved from

http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel_2007033_sum.pdf

Youngs, P. (2013). *Using teacher evaluation reform and professional development to*

*support Common Core assessments*.  Washington, D.C. Center for American

Progress.  Retrieved from http://files.eric.ed.gov/fulltext/ED539747.pdf

Young, S., Range, B. G., Hvidston, D. L., & Mette, I. M.  (2015). Teacher evaluation

    reform:  Principal's beliefs of newly adopted teacher evaluation systems.

    *Planning and Changing, 46*(1/2), 158-174.

## Appendix A

### Lindenwood University ● St. Charles, Missouri

# Educational Leadership - Prospectus

**Date** ___December 4, 2015_____

**Chair** __Dr. Terry Stewart_____          **Student** __Maggie Mathus_____

Maggie Mathus,

Your prospectus has been approved. Please proceed with the next step, which is to prepare and send your IRB DRAFT to swisdom@lindenwood.edu for review.

Please, work with your chair to transfer information to your draft of the IRB protocol application. Include consent forms, permission notes, survey / interview / focus group questions, if appropriate, and NIH certificate.

When ready, your chair will send the IRB DRAFT (all items pasted into one Word document) for review of items related to ethics and bias to (swisdom@lindenwood.edu).

Thank you,

Sherrie Wisdom, EdD

Associate Professor - Education Leadership
Supervisor of Graduate Research

**Appendix B**

# Northwest R-I
## SCHOOL DISTRICT

**INFORMED CONSENT: PERMISSION TO USE
PREMISES, NAME, AND/OR SUBJECTS**

August 3, 2015

To Lindenwood University,

As a representative of the Northwest R-1 School District, I confirm that the school district grants permission for Maggie Mathus, a Lindenwood Doctoral Student, to conduct the proposed research titled, *A Mixed Method Study of the Relationship Between Teacher Evaluation Ratings and Student Achievement as Measured by MAP, iReady, and Teacher and Principal Perceptions.* I authorize Maggie Mathus to use the premises, name, and/or subjects requested to conduct the study.

| | |
|---|---|
| *Mary Thomasson* | *NWR-1 Chief Academic Officer* |
| Printed Name of School District Official | Title of School District Official |
| | |
| Signature of School District Official | August 3, 2015    Date |

2843 Community Lane ● High Ridge, MO  63049  |  Phone (636) 677-3473  |  Fax (636) 677-5480  |  www.nwr1.k12.mo.us

**Appendix C**

# LINDENWOOD

LINDENWOOD UNIVERSITY ST.CHARLES, MISSOURI

| | |
|---|---|
| DATE: | February 19, 2016 |
| TO: | Maggie Mathus |
| FROM: | Lindenwood University Institutional Review Board |
| STUDY TITLE: | [868192-1] A mixed methods study of the relationship between Teacher Evaluation Ratings and Student Achievement |
| IRB REFERENCE #: | |
| SUBMISSION TYPE: | New Project |
| ACTION: | APPROVED |
| APPROVAL DATE: | February 19, 2016 |
| EXPIRATION DATE: | February 19, 2017 |
| REVIEW TYPE: | Expedited Review |

Thank you for your submission of New Project materials for this research project. Lindenwood University Institutional Review Board has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to the IRB.

This project has been determined to be a Minimal Risk project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the completion/amendment form for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of February 19, 2017.

Please note that all research records must be retained for a minimum of three years.

Generated on IRBNet

**Appendix D**

# LINDENWOOD

## ADULT - INFORMED CONSENT FOR PARTICIPATION IN RESEARCH ACTIVITIES

"A mixed methods study of the relationship between
Teacher Evaluation Ratings and Student Achievement"

Principal Investigator:  Margaret Mathus
       Telephone:  314-607-7875   E-mail: mam905@lindenwood.edu

---

1.  You are invited to participate in a research study conducted by Margaret Mathus under the guidance of Dr. Terry Stewart.  The purpose of this research is to investigate a possible relationship between teacher's profession growth, measured by the district's teacher evaluation system (Professional Growth Plans) and student academic performance at the elementary level.  This study will analyze teacher and administrator perceptions of the district's teacher evaluation system.  The study will also analyze teacher evaluation growth ratings and their students' academic progress as measured by MAP and i-Ready assessments.

2.  a) Your participation will involve
    - ➢ Completion of a 14-question (Likert Scale) survey through Survey Monkey.
    - ➢ Completion of the survey will indicate that you have: read and understand the information provided in this Informed Consent Letter, that you willingly agree to participate, and that you are aware of your right to withdraw your consent and discontinue participation at any time.
    - ➢ The data will be collected, managed, and aggregated through the online data management program, Survey Monkey. The researcher will receive data without personal identifiers to preserve the anonymity of the participants.
    - ➢ Participants may access the survey through the link provided in the invitational email.  Directions for this one-time survey will precede the survey questions in the on-line survey system (Survey Monkey).  The survey may be completed at the participant's convenience any time before the survey deadline, which will be communicated at the time of the invitation to participate (May 1, 2016).  Participant data will be stratified by the following categories: administrator or teacher.  The data will also consider the following demographics of participants: the number of years of teaching experience, and number of years of teaching experience in this school district.

b) The amount of time involved in your participation will be approximately 10 minutes.  Participants will receive no compensation, but a $5.00 donation will be made to the Northwest Lions Educational Alliance For Funding (NWLEAFF) organization for every submitted from a staff member (with a maximum of $200 total donation).

Approximately 70 participants will be involved in this research.

3.  There are no anticipated risks associated with this research but there may be certain discomforts associated with this research. They include reluctance to answer survey questions honestly or provide free-form feedback in the comments section, for concern that doing so may make the respondent personally identifiable to the researcher and that the feedback may be communicated to district or building supervisors. To mitigate this risk it will be important to communicate to research respondents that the survey responses are completely anonymous, and that the feedback is being collected by a third party tool to which the district does not have access. Additionally, respondents should be informed that any comments entered into the survey will be summarized by the researcher, rather than having the participants' verbatim responses appear in the final study. Unless the respondent self-identifies through responses to the open-ended questions, neither the researcher nor district staff will have any way to discern the identity of the respondents.

4.  There are no direct benefits for you participating in this study. However, your participation will contribute to the knowledge about how teacher evaluations can support educator growth and may help society.

5.  Your participation is voluntary and you may choose not to participate in this research study or to withdraw your consent at any time. You may choose not to answer any questions that you do not want to answer. You will NOT be penalized in any way should you choose not to participate or to withdraw.

6.  We will do everything we can to protect your privacy. As part of this effort, your identity will not be revealed in any publication or presentation that may result from this study and the information collected will remain in the possession of the investigator in a safe location. In some studies using small sample sizes, there may be risk of identification.

7.  If you have any questions or concerns regarding this study, or if any problems arise, you may call the Investigator, Margaret Mathus (314-607-7875) or the Supervising Faculty, Dr. Terry Stewart (636-949-4656).  You may also ask questions of or state concerns regarding your participation to the Lindenwood Institutional Review Board (IRB) through contacting Dr. Marilyn Abbott, Interim Provost at mabbott@lindenwood.edu or 636-949-4912.

**I have read this consent form and have been given the opportunity to ask questions.  I will also be given a copy of this consent form for my records.  I consent to my participation in the research described above.**

_____          _____

Participant's Signature                Date          Participant's Printed Name


_____          _____

Signature of Principal Investigator   Date          Investigator Printed Name

**Appendix E**

**SURVEY QUESTIONS FOR TEACHERS**

*I plan to set up this survey on Survey Monkey.  This document lists the questions and answer options I will provide, but the actual formatting may differ according to the capabilities of Survey Monkey's tool.*

Thank you for participating in my research study.  Data collected from this survey will remain anonymous and be used solely for my dissertation project.

**Overall Evaluation Program Rating**

*Please reflect on the evaluation process in your school for this current school year. Consider the entire evaluation process including goal setting (PGP), self-assessment, meetings with evaluator, formal and informal observations, and other procedures or feedback to rate the **overall quality of the NWR-1 teacher evaluation process.***

Please indicate the degree to which you agree or disagree with the following statements. (1 = strongly disagree,  5 = strongly agree)

A. The new NWR-1 teacher evaluation framework has provided more useful feedback from evaluators than the previous model of teacher evaluation used by the district.

   | 1 | 2 | 3 | 4 | 5 |

   Strongly                                              Strongly
   Disagree                                              Agree

B. The new NWR-1 teacher evaluation framework allows the evaluator to assess a more accurate picture of my teaching ability than the previous teacher evaluation model used in the district.

   | 1 | 2 | 3 | 4 | 5 |

   Strongly                                              Strongly
   Disagree                                              Agree

C. The implementation of the new NWR-1 evaluation framework has provided greater opportunity for professional growth for me than the previous evaluation model used in my district.

   | 1 | 2 | 3 | 4 | 5 |

   Strongly                                              Strongly
   Disagree                                              Agree

D. Overall, I feel the new NWR-1 evaluation framework is a more effective evaluation model than the evaluation model previously used in the district in determining quality teaching.

   | 1 | 2 | 3 | 4 | 5 |

Strongly                                                 Strongly
Disagree                                                 Agree

E.   The new NWR-1 evaluation framework requires me to focus more on
     strategies related to specific goals than the previous evaluation model used in
     my district.

1               2               3               4               5

Strongly                                                 Strongly
Disagree                                                 Agree

If you disagreed or strongly disagreed with any of these statements, please explain
why.

[insert text box for open-ended comments]

**Specific Attributes of Evaluation Rating**
*Please reflect on your Professional Growth Plan for the current year.  Consider the
entire PGP process including your district goal/indicator and your personal goal/
indicator, your benchmark and follow-up ratings, the specific strategies (artifacts) you
used to work towards those goals/indicators, and student assessment data used to rate the
impact of the evaluation process on a **your professional growth**.*

1.  Please indicate the degree to which you agree or disagree with the following
    statements.  (1 = strongly disagree,  5 = strongly disagree)

    A.   Focusing on specific goals/indicators within the NWR-1 Evaluation
         Framework led to changes that have improved my teaching.

         1               2               3               4               5

         Strongly                                                 Strongly
         Disagree                                                 Agree

    B.   I have made changes in the way I plan lessons as a result of the district goal
         of differentiated instruction.

         1               2               3               4               5

         Strongly                                                 Strongly
         Disagree                                                 Agree

    C.  I have made changes in my teaching practice as a result of my personal goal.

         1               2               3               4               5

Strongly                                                                Strongly
Disagree                                                                Agree

D. The feedback from my principal during formal or informal observation conferences was helpful and constructive.

| 1 | 2 | 3 | 4 | 5 |

Strongly                                                                Strongly
Disagree                                                                Agree

E. The feedback from my principal during formal or informal observation conferences led me to make improvements in my teaching.

| 1 | 2 | 3 | 4 | 5 |

Strongly                                                                Strongly
Disagree                                                                Agree

F. The evaluation system has a positive impact on my students' learning.

| 1 | 2 | 3 | 4 | 5 |

Strongly                                                                Strongly
Disagree                                                                Agree

G. The evaluation system has a positive impact on my teaching quality.

| 1 | 2 | 3 | 4 | 5 |

Strongly                                                                Strongly
Disagree                                                                Agree

H. I use more student assessment data to guide my planning of lessons than I did prior to the implementation of the new NWR-1 evaluation framework.

| 1 | 2 | 3 | 4 | 5 |

Strongly                                                                Strongly
Disagree                                                                Agree

I. The new NWR-1 teacher evaluation framework requires me to focus more on strategies related to my goals (PGP) than the previous evaluation program used in my district.

| 1 | 2 | 3 | 4 | 5 |

Strongly                                                                Strongly
Disagree                                                                Agree

        If you disagreed or strongly disagreed with any of these statements, please explain why.

        [insert text box for open-ended comments]

**Additional Comments**

Please provide comments on any aspects of the NWR-1 Evaluation Framework you

believe were not addressed in the survey questions.

_____

_____

**Appendix F**

**SURVEY QUESTIONS FOR ADMINISTRATORS**
*I plan to set up this survey on Survey Monkey. This document lists the questions and answer options I will provide, but the actual formatting may differ according to the capabilities of Survey Monkey's tool.*

Thank you for participating in my research study. Data collected from this survey will remain anonymous and be used solely for my dissertation project.

**Overall Evaluation Program Rating – District Level**
*Please reflect on the evaluation process in your school for this current school year. Consider the entire evaluation process including goal setting (PGP), self-assessment, meetings with evaluator, formal and informal observations, and other procedures or feedback to rate the **overall quality of the NWR-1 teacher evaluation process.***

2. Please indicate the degree to which you agree or disagree with the following statements. (1 = strongly disagree, 5 = strongly agree)

   F. The new NWR-1 teacher evaluation framework provides more useful feedback from evaluators than the previous model of teacher evaluation used by the district.

   | 1 | 2 | 3 | 4 | 5 |
   |---|---|---|---|---|
   | Strongly Disagree | | | | Strongly Agree |

   G. The new NWR-1 teacher evaluation framework allows me, as an evaluator, to assess a more accurate picture of my staff's teaching ability than the previous teacher evaluation model used in the district.

   | 1 | 2 | 3 | 4 | 5 |
   |---|---|---|---|---|
   | Strongly Disagree | | | | Strongly Agree |

   H. The implementation of the new NWR-1 evaluation framework has provided greater opportunity for professional growth for my staff than the previous evaluation model used in my district.

   | 1 | 2 | 3 | 4 | 5 |
   |---|---|---|---|---|
   | Strongly Disagree | | | | Strongly Agree |

   I. Overall, I believe the new NWR-1 evaluation framework is a more effective evaluation model than the evaluation model previously used in the district in determining teacher quality.

|   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|

Strongly                                         Strongly
Disagree                                         Agree

J.   The new NWR-1 evaluation framework requires teachers to focus more on
strategies related to specific goals than the previous evaluation model used in
the district.

|   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|

Strongly                                         Strongly
Disagree                                         Agree

If you disagreed or strongly disagreed with any of these statements, please explain
why.

[insert text box for open-ended comments]

**Specific Attributes of Evaluation Rating – Building Level**
*Please reflect on teacher's Professional Growth Plans for the current year.  Consider the
entire PGP process including the district goal/indicator and personal goal/ indicator,
benchmark and follow-up ratings, the specific strategies (artifacts) used, and student
assessment data used to rate the impact of the evaluation process on **teacher's
professional growth**.*

3.   Please indicate the degree to which you agree or disagree with the following
statements.  (1 = strongly disagree,  5 = strongly disagree)

J.   Focusing on specific goals/indicators within the NWR-1 Evaluation
Framework led to changes that have improved my staff's teaching.

|   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|

Strongly                                         Strongly
Disagree                                         Agree

K.  My staff has made changes in the way they plan lessons as a result of the
district goal of differentiated instruction.

|   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|

Strongly                                         Strongly
Disagree                                         Agree

L.  My staff has made changes in their teaching practices as a result of their
personal goals.

|   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|

Strongly                                                    Strongly
Disagree                                                    Agree

M. I am able to provide helpful and constructive feedback to my staff during formal or informal observation conferences.

    1              2              3              4              5

Strongly                                                    Strongly
Disagree                                                    Agree

N. My staff has made improvements in their teaching practices as a direct result of the feedback I provided through formal or informal observation conferences.

    1              2              3              4              5

Strongly                                                    Strongly
Disagree                                                    Agree

O. The evaluation system has a positive impact on student learning in my building.

    1              2              3              4              5

Strongly                                                    Strongly
Disagree                                                    Agree

P. The evaluation system has a positive impact on the overall teaching quality in my building.

    1              2              3              4              5

Strongly                                                    Strongly
Disagree                                                    Agree

Q. I use more student assessment data in my evaluations of teachers I did prior to the implementation of the new NWR-1 evaluation framework.

    1              2              3              4              5

Strongly                                                    Strongly
Disagree                                                    Agree

R. The new NWR-1 teacher evaluation framework holds teachers more accountable to their goals in their individualized Professional Growth Plan (PGP) than the previous evaluation program used in the district.

    1              2              3              4              5

Strongly                                                    Strongly
Disagree                                                    Agree

If you disagreed or strongly disagreed with any of these statements, please explain why.

[insert text box for open-ended comments]

**Additional Comments**

Please provide comments on any aspects of the NWR-1 Evaluation Framework you

believe were not addressed in the survey questions.

_____

_____

**Vitae**

Maggie Mathus has been an educator for 16 years.  She earned a Bachelor of Science in Elementary Education from Southeast Missouri State University in 2000 and a Master of Arts in Elementary Education from Lindenwood University in 2010.  Maggie is currently an elementary teacher at Cedar Springs Elementary school in House Springs, Missouri. She has served as a teacher leader and mentor, and facilitated professional development within her school district.  Her experience includes collaborating with Missouri's Department of Elementary and Secondary Education on State Assessments. During the 2014-2015 school year, Maggie was recognized as a Peabody Energy Leader in Education Award recipient.