

Bethel University

Spark

All Electronic Theses and Dissertations

2020

Teacher Evaluation: Methods Most Effective in Promoting Teacher Growth and Student Success

Jaclynn Bock
Bethel University

Follow this and additional works at: <https://spark.bethel.edu/etd>



Part of the [Education Commons](#)

Recommended Citation

Bock, J. (2020). *Teacher Evaluation: Methods Most Effective in Promoting Teacher Growth and Student Success* [Master's thesis, Bethel University]. Spark Repository. <https://spark.bethel.edu/etd/79>

This Master's thesis is brought to you for free and open access by Spark. It has been accepted for inclusion in All Electronic Theses and Dissertations by an authorized administrator of Spark.

TEACHER EVALUATION: METHODS MOST EFFECTIVE IN PROMOTING TEACHER GROWTH AND
STUDENT SUCCESS

A MASTER'S THESIS
SUBMITTED TO THE FACULTY
OF BETHEL UNIVERSITY

BY

JACLYNN A. BOCK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTERS OF ARTS

APRIL 2020

BETHEL UNIVERSITY

TEACHER EVALUATION: THE METHODS MOST EFFECTIVE IN PROMOTING TEACHER GROWTH
AND STUDENT SUCCESS

JACLYNN A. BOCK

APRIL 2020

ADVISOR'S NAME: Lisa Silmser, Ed. D

PROGRAM DIRECTOR'S NAME: Lisa Silmser, Ed. D

Abstract

This study investigates the effectiveness of teacher evaluation used across the United States and how evaluation tools can also improve the skills and development of educators. The literature review analyzes three forms of evaluation, including summative, formative, and emergent practices. Specific topics addressed in the research include the following: standards-based evaluation, Charlotte Danielson's *Framework for Teaching* (2011), Robert J. Marzano's *Marzano Teacher Evaluation Model* (2012), professional learning communities, collaboration among colleagues, walkthrough observations, and merit pay. The effective forms of evaluation identified by research include the summative and formative practices, with the possibility of all forms being effective when multiple forms are used simultaneously. However, the importance of timely and descriptive feedback is necessary to effectively communicate teacher effectiveness and opportunities for growth.

Table of Contents

Signature Page	2
Abstract	3
Table of Contents	4
Chapter I: Introduction	6
Context of Research and Application	6
Guiding Question	7
The Problem: Inconsistent Teacher Evaluation	7
Categories of Evaluations	8
Purpose of Teacher Evaluation	9
Rationale	10
Chapter II: Literature Review	12
Research Strategies	12
Summative Evaluation: Standards-Based Evaluation	12
Danielson's Evaluation	14
Danielson's Evaluation: Planning and Preparation	15
Danielson's Evaluation: Classroom Environment	15
Danielson's Evaluation: Instruction	17
Danielson's Evaluation: Professional Responsibilities	19
Danielson's Evaluation: Effectiveness of <i>Framework for Teaching</i>	20
Marzano's Evaluation	21
Formative Evaluation	24

	5
Professional Learning Communities and Collaboration	24
Walkthrough Observations	27
Emergent Evaluation	29
Merit Pay	29
Effectiveness of Summative, Formative, and Emergent Practices	34
Chapter III: Discussion and Conclusion	38
Summary of Literature	38
Professional Application	39
Limitations of the Research	41
Implications for Future Research	42
Conclusion	43
References	44

CHAPTER I: INTRODUCTION

Context of Research and Application

Every day teachers act as social workers, nurses, coaches, and leaders to their students. Thousands of decisions are made in the classroom daily, some resulting in great learning, while others ending in fires that need to be extinguished. Teaching can be challenging, but each day teachers inspire and lead upcoming generations. In order for students to become strong leaders for tomorrow's generation, teachers must become strong leaders today. Growth through evaluation and reflection are at the heart of creating strong teacher leaders in the classroom, and there are many ways in which teachers can be encouraged, challenged, and evaluated. This thesis is focused on evaluating and creating teacher evaluation methods to create leaders for the next generation.

Research has reflected that student learning is most impacted by teacher quality and efficacy (Berg, 2010; Stronge, Ward, & Grant, 2011; Weisburg, Sexton, Mulhern, Keeling, 2009). Each student needs teachers who have acquired sufficient knowledge and skills necessary to help each student learn and grow. Great teaching can also overcome even the most challenging home situations and outside factors that may hinder academic achievement. The question remains of how all students will be given teachers that will meet that need, seeing as many teachers enter the field without adequate experiences and "may have to learn while teaching" (Berg, 2010, p. 195). School evaluation coupled with strong feedback from school leaders provides teachers opportunities for teaching mastery.

As a fifth and sixth-grade teacher, I have experienced the impact of teaching evaluation; personally, some feedback very beneficial, while some unhelpful. To create a positive school

culture of growth, not only for students but also for teachers and staff, evaluation can be used as a tool to further the common mission of educating the whole child - mind, body, and spirit.

Guiding Question

A review of literature will seek to answer the question: What are the effective measures used in education to evaluate teacher performance? How do school communities improve the skills and growth of teachers, and can that be coupled with evaluation? The unique structure of teaching is the impact that school leaders – principals, assistant principals, teacher leaders and mentors – can have on teachers, who in-turn impact the students they interact with on a daily basis. The funnel flowing from school leader to teacher to students can result in greater student learning and performance if evaluation and reflection are done well.

Discussion about teacher evaluation will begin with the problem of teacher evaluation: historic inconsistencies in procedures, followed by the multiple categories of evaluation, standards-based evaluation, and additional evaluation assessments and tools that can be used in the classroom. Each of these pieces of teacher evaluation will provide a complete picture of effective teacher feedback and evaluation.

The Problem: Inconsistent Teacher Evaluation

Teacher evaluation has been criticized for not adequately achieving the intended goals of evaluation. According to Marzano (2012), teacher evaluations fail for two reasons: they do not adequately measure teacher quality because the system does an inadequate job of discerning between effective and ineffective teachers, and evaluations have failed to promote growth in developing skilled educators. Toch (2008) criticizes evaluation systems for not providing teachers with feedback that fuels the development of highly proficient educators, and

instead focus on a checklist of “classroom conditions” (p. 32) that do not directly affect teacher instruction. Darling-Hammond (2013) contends that current evaluation systems lack consistent and clear standards, little focus on improving teacher practices, limited consideration of student outcomes and “cookie-cutter” (p. 5) procedures that do not consider teacher needs. In a study conducted by Kimball and Milanowski (2009), of the 23 school leaders included, results reflected that inconsistencies of evaluation practices, training, and oversight can affect principals’ judgment and therefore additional evaluation validity is required to promote consistent evaluation techniques when matching principal evaluation with compensation outcomes.

According to Darling-Hammond (2013), a new system must be developed to “ensure that teacher evaluation is connected to -- not isolated from -- preparation and induction, daily practices, and a productive instructional context (p. 7). A change in teacher evaluation is coming, but it must serve many purposes.

Categories of Evaluations

Evaluation can be categorized into three forms: summative, formative, and emergent (Boyland, Harvey, Quick, & Choi, 2014). Serving to meet accountability measures, summative evaluation includes scheduled and non-scheduled classroom observations, pre- and/or post-observation conferences, and rating scale and contribute to continued employment, tenure, and assignment. Principals act as the “judge” in evaluating teacher effectiveness and meeting district requirements. Formative evaluation hinges on collaboration between teachers and principals and focuses on improvement, goal setting, peer coaching, and principal walkthroughs. In this setting, principals act as a “coach” or “consultant” in supporting teacher

development. According to a study conducted by Ebmeier (2003) research reflects that when supervision of teachers is done well, teacher efficacy, job satisfaction, and commitment to practice improves, but primarily when feedback that is given is focused on growth goals.

Lastly, emergent evaluation is considered a high stakes form of evaluations and highlights value measures (Boyland et al., 2014). Including teacher evaluation surveys from students and/or parents, portfolio assessments, merit pay, and value-added measures that link student achievement to teacher efficacy. However, there are challenges associated with this practice, including inconsistent student scores year to year, statistical discrepancies of scores applied to the same teacher, and differences based on students assigned to teachers yearly. The pressure to meet student goals on standardized testing can also result in teachers focusing on teaching solely the material on the test without incorporating other necessary student skills (as cited in Boyland et al., 2014). Currently, Boyland et al. (2014) recommends using a combination of evaluations in addition to emergent evaluation to evaluate teacher effectiveness.

Purpose of Teacher Evaluation

Teacher evaluation serves multiple purposes, including the measurement of current teacher ability and areas of growth with a higher emphasis on teacher growth than measurement (Marzano, 2012). According to Marzano (2012), evaluation hinges on a system that meets three necessary characteristics: comprehensive and specific, includes a developmental scale and rewards academic growth. A comprehensive and specific evaluation system identifies the necessary skills students and teachers must attain in the classroom during a lesson. It extends beyond just providing clear and attainable objectives for students to attain,

but also incorporates the need for teachers to provide students the opportunity to learn using multiple methods, examples, and engaging materials.

A developmental continuum allows for teachers to guide teacher improvements. Using terms such as “not using, beginning, developing, applying, and innovating” (Marzano, 2012, p. 18) allow for a continuum for teachers to attain, instead of using a number system that does not describe the meaning of the stages. Using evaluation methods such as those listed above promote goals for higher levels of proficiency between teachers.

Teaching hinges on not remaining stagnant in where a teacher’s ability lies, but instead progressing forward and setting goals for achievement. A strong evaluation system acknowledges skills but also recognizes growth.

Rationale

Sankofa, originating in the Twi language of Ghana translates to “go back and get it,” referring to the ability to return to look back and learn from previous opportunities, experiences, and challenges. Teachers, students, and school leaders can take hold of this idea and learn from past opportunities. Teacher evaluation is rooted in the same principles, encouraging teaching professionals to look back at what they have accomplished and grow in their profession in order to affect the upcoming generations of learners.

In an attempt to encourage *sankofa* within the learning process on the part of both the teacher and the student, teacher evaluation must encourage the practice of self-reflection and improvement. An evaluation tool should include two pieces of information: the overall score of the teacher (effectiveness, ability, etc.) and areas in need of growth (Danielson, 2011; Sartain, Stoelinga, & Brown, 2011). According to Marzano (2012), the measurement of teacher ability is

an important aspect of evaluation, but a two-fold system that also allows for teacher growth is the most effective tool. Many tools can be used to evaluate teachers, but which provide the most effective form of feedback and which improve student learning? Teachers model *sankofa* to their students in their desire to improve their abilities, with the hope of not only being a role model for student behavior, but also helping students improve in their understanding and achievement.

CHAPTER II: Literature Review

Research Strategies

The literature for this thesis was located using a variety of search engines, including ERIC (EBSCOhost), Academic Search Premiere, EBSCO MegaFILE, Educator's Reference Complete, and Educational Journals. The research was narrowed to publications between the years 2000 and 2020. The parameters established for the of published research include peer-reviewed journals that contained information on the tools of teacher evaluation, the effectiveness of the multiple forms of evaluation tools, and professional development as it relates to teacher growth and development. The guiding questions of this thesis promoted the use of the keywords to locate appropriate research; the keywords include “teacher evaluation,” “summative evaluation,” “formative evaluation,” “emergent evaluation,” and “professional development.” The thesis is organized according to three forms of teacher evaluation: Summative Evaluation, Formative Evaluation, and Emergent Evaluation, with the conclusion synthesizing the use of these forms of evaluation collectively.

Summative Evaluation: Standards-Based Evaluation

There are many forms of teacher evaluation used today, but at the center of current evaluation practices is the use of standards-based evaluations, a framework identifying key student learning outcome standards coupled with teaching expectations and guidelines that promote teacher learning and growth. Standards-based evaluations of practice generally include systematic observations that include feedback that indicates teacher effectiveness and professional standards (Darling-Hammond, 2013). There are many platforms for standards-based teacher evaluation used in schools today, including Danielson’s *Framework for Teaching*

(2011), Marzano Teacher Evaluation Model (2012), Race to the Top (2009), Kim Marshall's Rethinking Teacher Supervision and Evaluation (2013), and many other active models currently used in the United States. In each of these methods there is a desire to assist in helping teachers improve their instruction, but the methodology differs. Because of the need for more adequate teacher evaluations, as demonstrated by the inconsistencies and problems listed previously, a study of effectiveness is important. Danielson's (2011) and Marzano's (2017) evaluation structures have long been utilized in schools and their contributions and studies have been included in this literature review.

Danielson's (2011) methods of evaluation focus on teacher effectiveness. The four domains of evaluation include planning and preparation, classroom environment, instruction, and professional responsibilities. Marzano's (2017) form of evaluation is a five-point scale that is slightly different and focuses on teacher development through the following categories: classroom strategies and behaviors, planning and preparation, reflection on practices, and professionalism between coworkers.

Adapted by over 200 school districts worldwide (Roegman, Goodwin, Reed, & Scott-McLaughlin, 2007), Danielson's *Framework for Teaching* has been adopted by many school leaders and districts to support teachers in growth and development. Beneficial to both new teachers and veterans in the field, descriptive teacher evaluation can be especially helpful in guiding new teachers in the field. "The use of a framework such as Danielson's can support preservice teachers in identifying and developing effective teaching practices" (Roegman et al., p. 111, 2007). Providing details of good teaching, Danielson's framework for high-quality instruction extends beyond checklists of teaching responsibilities and recognizes other

responsibilities of teachers such as student needs, classroom cultures, and professional responsibilities.

Danielson's Evaluation

Charlotte Danielson – educator, author, speaker, and professional development trainer for Educational Testing Service (ETS), National Board of Professional Teaching Standards (NBPTS), and Association for Supervision and Curriculum Development (ASCD) – is a leader in the educational community. Her *Framework for Teaching* has been adapted to accommodate thousands of schools and has provided professional development resources to many (Kettler & Reddy, 2017).

Broken down into four comprehensive domains, Danielson highlights the important aspects of teaching: planning and preparation, classroom environment, instruction, and professional development. These four domains are broken down into 22 elements to be used as an evaluation tool for effective teaching. According to Danielson (2007), these domains are not specific steps teachers must take to become a “good teacher,” but are rather tools that can be used to focus on specific elements of effective teaching.

In each of the domains, teachers are evaluated on an incremental scale: unsatisfactory, basic, proficient, and distinguished. This performance level is then combined with two remaining parts of each scale, including the rationale with explanation and documentation of seen behavior. In 2011, Danielson revisited her *Framework for Teaching* while part of the Measures of Effective Teaching (MET) Project, as funded by the Bill and Melinda Gates Foundation, to substantiate her work through the research of over 20,000 classrooms using the *Framework for Teaching* model (Griffin, 2013). The results reflected no change in the domains

of her previous work, validating the domains Danielson created to evaluate teacher effectiveness.

Danielson's Evaluation: Planning and Preparation. The first domain of *The Framework for Teaching* is focused on teachers knowing what they are teaching and how to teach it well in a timely and appropriate manner. The teacher must not only know the content well but also the students as well in order to communicate the information being taught; student skill level, ability, cultural backgrounds, and interests are addressed in this area in communicating information (Danielson, 2007).

Another important component of this domain is goal setting resource knowledge and allocation. When necessary, students must be given additional tools to reach the intended lesson objective and supported and challenged appropriately. Assessment is necessary to evaluate if students were able to meet the previously identified objectives.

Danielson's Evaluation: Classroom Environment. The second domain of Charlotte Danielson's work is focused on the classroom environment and creating a culture of respect within the walls of the learning environment (Danielson, 2007). The first of five components within this domain is focused on respect between the students and the teacher. This sense of respect feeds into the second component of safety and security, which can be built and sustained with effective and consistent classroom practices. The fourth component of the classroom environment is appropriate student behavior management, including teacher response and redirection of student misbehavior. The fifth component of this domain is the organization of materials and classroom supplies in the physical classroom.

In an attempt to improve teacher growth and revamp teacher evaluation methods, Chicago Public Schools launched the Excellence in Teaching Pilot program in 2008 (Sartain, Stoelinga, & Brown, 2011). At the end of the 2007-2008 school year, Chicago Public Schools had identified 93 percent of teachers as effective or superior, while only 44 percent of schools were meeting state standards in Illinois testing. In an effort to change these statistics, the Excellence in Teaching Pilot program was launched and utilized Danielson's *Framework for Teaching*, providing teachers with feedback on strengths and weaknesses while driving instructional improvement forward in an attempt to differentiate between teachers and improve teacher and student growth. For two years, elementary schools in the Chicago Public Schools district were evaluated and data was collected using statistical models. In the study, 44 elementary schools participated during the first year (2008-2009), and during the second year participation was increased to 101 elementary schools. The program included training and resources for principals and teachers on observational methods and requirements, principal observations of teacher practices using the Danielson's framework that occurred twice during the school year, and pre- and post- conferences to provide teachers feedback on lessons. Domains two and three were specifically studied as a form of validity and reliability in this study because of how they relate directly to the focus on Excellence in Teaching Pilot program.

Because Danielson's *Framework for Teaching* was a new form of evaluating teachers, the Chicago Public Schools wanted to encourage principals and vice principals in their abilities to effectively implement this form of evaluation. There were challenges associated with implementing the new program, including weak instructional coaching skills and lack of acceptance of the new program from principals and/or school leaders. As a way to improve the

principal or school leader's ability to evaluate teachers consistently and according to the model, the Chicago Public Schools using the Excellence in Teaching Pilot framework required that two raters observed the same lesson simultaneously. One of the raters was a highly trained observer who was able to give feedback to the principal or occasionally vice-principal on how to rate a teacher's performance accurately. Using this as a form of validity of school leader evaluation practices, Sartain et al. (2011) stated that, "Principals and observers gave similar proportions of Unsatisfactory and Basic ratings in most of the components, though principals were more likely to call practice Distinguished" (p. 14). Principals and school leaders were able to grow in their ability to evaluate teachers well and provide effective feedback in the process using this model established by the Chicago Public Schools.

After data was collected during the two years of the pilot program, the validity of the scores of teachers were compared to that of student achievement as shown on the Illinois state testing. According to Sartain et al. (2011), there is a strong correlation between teacher scores on Danielson's *Framework for Teaching* and value-added measures in the subject areas of mathematics and reading. Also, data showed that in classrooms of highly effective teachers as demonstrated by scores "Distinguished" (4) on the Danielson's *Framework for Teaching*, student growth was the greatest; while in classrooms of lower effective teachers ("Unsatisfactory" with a score of 1), student growth was the smallest.

Danielson's Evaluation: Instruction. Instruction is a key element to teaching and should be delivered in multiple ways, including interpersonal, kinesthetic, visual, and more, because students learn in a variety of ways (Gardner, 1999). The third domain of Danielson's *Framework for Teaching* is broken down into five components. The first aspect of instruction begins with

clear communication of content, both oral and written. Through effective communication, high-level questions and discussions should emerge, as they are important components of student understanding. With adequate questions and class discussions, student engagement and participation in content should be created. Lesson structure, material, and student grouping is part of this aspect of instruction. When students are engaged and producing materials through discussion, activities, or other resources, feedback must also be given and is a key element of this domain. Lastly, flexibility and adaptability is an important aspect of teaching; at any given time, not all students will be ready to learn the same content in the same way.

In a study of Cincinnati's Public Schools during the academic years of 2003-2004 through 2008-2009, Kane, Taylor, Tyle, and Wooten (2011), focused on determining the correlation between teacher effectiveness in Domains Two and Three, "Creating an Environment for Learning" and "Teaching for Learning" (which are the only two domains centered on classroom practices) and student achievement levels. Of the 572 teachers sampled, the results reflected an existing correlation between student achievement and teacher effectiveness in classroom practices. Looking specifically at grade levels three through eight, students were evaluated on the Ohio state exam in the subjects of reading and mathematics. The students' test scores were then compared to teacher evaluations and overall proficiency levels as defined by the *Framework for Teaching*.

The final results of Kane et al. (2011) reflected that teacher proficiency is incredibly important in students' overall test scores. According to the Cincinnati Teacher Evaluation System (TES) and Kane et al. (2011), a student who began at the 50th percentile in reading will improve approximately three percentile points working with a teacher in the top-quartile of

proficiency as measured by the TES using Danielson's *Framework for Teaching*; these same students would improve approximately two percentile points in mathematics. Ultimately, "The nature of the relationship between practices and achievement, as estimated here, supports teacher evaluation and development systems that make use of multiple measures" (Kane et al., 2011, p. 611). The comparison between teachers' evaluations and student achievement scores can be used together to determine professional development strategies moving forward, remedial strategies to support developing teachers, and professional connections for teachers in the same school setting. In summation, as defined by Danielson's *Framework for Teaching*, teachers' classroom practices do predict student achievement growth.

Danielson's Evaluation: Professional Responsibilities. The last domain of Danielson's Framework requires teachers to reflect on current and past professional practices and evaluate the effectiveness of different learning experiences for students (Danielson, 2007). Containing six components, the first element of this domain is self reflection, followed by maintaining accurate records of student learning and classroom practices. By completing the second element of this domain, teachers must also communicate the information they have gathered regarding student understanding to families. By creating relationships with students and parents, teachers are contributing to the overall school environment, which is the next element of this domain. The fifth element of this domain is professional development through classes, continuing education, and additional learning opportunities and resources. Lastly, teachers must demonstrate professionalism, serving students well and advocating for them when necessary and seeking opportunities to support them adequately.

Danielson's Evaluation: Effectiveness of the *Framework for Teaching*. According to Kettler and Reddy (2017), Danielson's *Framework for Teaching* has become one of the most widely accepted and used systems for observational evaluations of teachers. In a study of 12 high-poverty charter schools consisting of 156 teachers and 34 trained administrators using the *Framework for Teaching* (Danielson, 2013), the effectiveness of Danielson was evaluated (Kettler & Reddy, 2017). Focused on studying the reliability and validity of Danielson's four domains, Kettler and Reddy use a composite scoring approach coupled with standardized testing of student scores rather than the commonly discussed traditional domain methodology.

The teachers involved in this study were evaluated three times a year by school leaders who had been involved in a six-day observer training that focused on four key elements: background information on *Framework for Teaching*, how to use the framework effectively, critical observer skills and responsibilities, and the necessary skills to train upcoming observers. An additional seven and a half hour online training was also administered to ensure observer efficiencies.

School leaders part of a mid- Atlantic state observed lessons of the instructors and provided feedback in each domain of the Framework. Composite scores for all domains were then calculated through averages and compared to the MAP assessment testing (NWEA, 2011) results of the students within that grade level and classroom. MAP assessment testing is a test that indicates student growth over the course of one year and can be used to measure growth over the span of many years for students.

In summation, the results of this study indicate that on average the teachers observed in the 12 schools observed met proficiency level on the *Framework for Teaching* (rating of three

on the scale) and taught students who were near the national average score for mathematics and reading achievement. By averaging teacher ratings on the *Framework for Teaching*, “The composite scores were found internally consistent and more stable than traditional scoring” (Kettler & Reddy, 2017, p. 77). The composite scores also were predictive of student growth in mathematics and reading achievement.

When the scores are averaged within each domain, teachers can seek ways to improve that specific aspect of teaching. For example, if teachers do not meet proficiency or distinguished levels within a specific domain, such as classroom environment there are resources or professional learning opportunities available to that teacher to seek out to develop his/her ability in creating a nurturing learning environment; a development plan or lesson can be created between the teacher and school leader. “Collectively, these reliability estimates support the use of the composite scores as a justifiable, and perhaps preferable, way to use the FFT” (Kettler & Reddy, 2017, p. 77). One aspect Kettler and Reddy (2017) highlighted was that *Framework for Teaching* should be used in combination with other resources, such as testing or other observable practices to evaluate teacher effectiveness.

Marzano’s Evaluation

Marzano is an educational researcher who has created tools to evaluate and develop teacher competencies. With professional development resources available and training for educational leaders, Marzano provides education on how to use his model of evaluation known as the Marzano Teacher Evaluation Model (MTEM). This form of evaluation includes a five-point scale and encourages areas of growth for teachers in 41 elements. According to Marzano (2012), the most effective forms of evaluation have three components. First of all, an effective

evaluation model is comprehensive and specific. Classroom strategies and behaviors elicited by the teacher should provide students with opportunities for growth and if not, the model should provide evidence and reasoning for lack of performance. The second aspect of an effective evaluation tool includes a progressive rubric or scale system that measures development including “Not Using” (0), “Beginning” (1), “Developing” (2), “Applying” (3) and “Innovating” (4). The last characteristic of a strong evaluation model recognizes and rewards teacher growth; encouraging teachers to create goals and pursue resources to develop skills. At the end of the year, an evaluation tool should describe two goals: overall score and growth areas. According to Marzano (2012), the measurement of teacher ability is an important aspect of evaluation, but a two-fold system that also allows for teacher growth is the most effective tool. “These distinctions are crucial to the effective design and implementation of current and future teacher evaluation systems” (Marzano, 2012, par. 30).

In a study conducted by The Marzano Research Laboratory (2011), a sample size of 61 schools and between 19 and 54 teachers in Oklahoma were used to measure the effectiveness of the Marzano Teacher Evaluation Model (MTEM). In this study, teacher observation scores and student value-added measures in math and reading were compared. Twelve of the schools were high schools (5 were non-improvement status, while 7 were improvement status), 28 middle schools (12 non-improvement status, while 16 were improvement status), and 21 elementary schools (11 non-improvement status and 10 improvement status). Non-improvement status or improvement status is defined by meeting nine essential elements, including the following elements: (1) Instruction, (2) Curriculum, (3) Classroom Evaluation, (4) School Culture, (5) Professional Growth and Development, (6) Student/Community Supports,

(7) Organizational Structure/Resources, (8) Leadership, (9) Effective Planning. According to these results deemed by the state of Oklahoma, the 61 identified schools were studied in relation to teacher efficacy and student growth.

To begin this study, principals and/or vice principals at each of the 61 schools were interviewed during this process to evaluate observation techniques being used at each school. On-site classroom observations were also conducted at ten of the 61 schools used in this study to evaluate school leader's mode of evaluation.

Data was collected to indicate teachers' proficiency in all 41 elements of the Marzano Teacher Evaluation Model (MTEM). In math, all 41 of 41 correlations were positive while 39 of the 41 correlations were positive in English. This data of the Oklahoma school systems can be used to defend that instructional strategies matter. In summary, three conclusions can be made from this study. According to The Marzano Research Laboratory (2011), the "more strategies that were typically used in a school, the better the achievement in mathematics and reading" (p. 29). Both non-improvement and improvement schools alike need to focus on instructional strategies to improve student understanding. Secondly, a school-wide focus improves teaching ability and student growth. Schools in this study that were on an improvement plan and were able to successfully move out of the plan had a school-wide focus and goals that teachers and school leaders implemented. PLCs improved teacher growth and school development. Lastly, engaged teaching and vocabulary development improve student success of state-mandated testing. Based on Marzano Research Laboratory visits to schools on Improvement status, teachers that engaged students in learning produced higher test results.

According to Basileo and Toth (2019), Marzano Teacher Evaluation Model (MTEM) is widely used across the state of Florida but has received little recognition because of the possibility of the model's inefficiency to predict teacher effectiveness. With lack of evidence for the effectiveness of MTEM, Basileo and Toth (2019) collected data up to 13,316 teachers in the state of Florida for three consecutive school years beginning in the 2012-2013 school year.

During the years of evaluation, teachers were evaluated using the MTEM (obtained through the Florida Department of Education (FLDOE)) and teachers' scores were compared to value-added measures of students' scores in English and Language Arts (ELA) and math. Student growth was measured on how well they did on Florida state testing compared with prior testing results and grade-level predictions.

Ultimately, the results of the comparisons between the value-added measures and the MTEM scores of teacher in the state of Florida were positive and showed correlation between the two sets of data, with a higher correlation in student growth in math than ELA. Basileo and Toth's (2019) data reflected "small, positive, and statistically significant correlation coefficients between the average teacher observation score and value-added measures of teacher effectiveness" (p. 9). Data reflected that MTEM scores were the largest predictor of value-added measures, including student and teacher growth, the observation system, and school characteristics. Basileo and Toth (2019) warned that additional research is required to assess whether observation scores forecast the specific teacher value-added measures of school poverty rates, teacher characteristics, and observational system differences in schools.

Formative Evaluation

Professional Learning Communities and Collaboration

Often principals have relied on teacher evaluation to be indicators of teacher abilities, but according to DuFour and Mattos (2013), evaluation only provides a glimpse into a teacher's ability. Instead, teachers can grow most when they can participate in Professional Learning Communities (PLCs). "The most powerful strategy for improving both teaching and learning, however, is not by micromanaging instruction but by creating the collaborative culture and collective responsibility of a professional learning community (PLC)" (DuFour & Mattos, 2013). This allows for collaboration between colleagues and also principals to engage in meaningful conversations and relationships with teachers. According to Green (2017), one of the most important relationships in creating a healthy community is the collaboration that happens between school leaders.

Teamwork and collaboration in a school setting fosters community between teachers and school leaders, which in turn produces more effective teachers (Leana & Pil, 2009; Schiff, Herzog, Farley-Ripple, & Iannuccilli, 2015). "The most effective teachers do not work in isolation" (Schiff, et al., 2015, p. 1). Collaboration and trust among teachers and coworkers (social capital) have a great influence on student achievement, possibly as much as teachers' classroom competency and experience (human capital) (Leana & Pil, 2009; Bryk & Schneider, 2002). In a study of 1,013 teachers in 202 schools in the northeastern United States, Leana and Pil (2009) describe a model to evaluate teachers' social capital and human capital. In March 2004, surveys were distributed to the teachers by school leaders (principals, department leaders, etc.) who had been trained in distributing surveys and collecting data from departments of people. Interviews of teachers and school administrators were also conducted by the survey committee to analyze the teachers' effectiveness in the classroom. Information

collected included questions and information regarding the teacher's level of education, current position and time spent in the described position.

Interviews and surveys indicated that teachers who remained in the same position for extended amounts of time had a greater impact on student ability given adequate resources and training opportunities for development, specifically in the content area of math. Leana and Pil (2009) addressed the importance of collaboration between coworkers, indicating that it is a gauge of effective teaching.

In the survey distributed to teachers, questions of social interactions between coworkers were addressed. Teachers were asked about their horizontal ties at the school, including questions regarding the number of connections they had within the grade level or department, how frequently they collaborated among their specific grade level or team, and the closeness between the teams. Using a scale of varying degrees (1 reflecting no closeness; 5 meaning very close), teachers were asked to evaluate their interactions with coworkers as well as with principals and school leaders (vertical ties).

Ultimately, results reflected positive correlations between teacher collaboration and student achievement. "At the teacher level, formal education, experience, and ability, as well as both horizontal and vertical tie strength, are positively and significantly correlated with student performance" (Leana & Pil, 2009, p. 1111). According to Leana and Phil (2009), teacher collaboration, especially in the content area of mathematics, is an important aspect of teacher efficacy and can be used as a form of evaluation when relationships have been established and trust, both in terms of horizontal and vertical ties, has been established. Student growth and ability is promoted and enhanced.

Walkthrough Observations

Walkthrough observations are an informal form of evaluation where school leaders, such as principals, instructional coaches, or department leaders are encouraged to enter a teacher's classroom for a short amount of time, approximately five to 20 minutes, and provide feedback following the observation to provide the teacher with information regarding the lesson and potential professional development opportunities for growth (Garza, Ovando, & O'Doherty, 2016). Instead of prescheduled classroom observations with pre- and post-conference meetings, walkthroughs can be an effective and authentic observation tool of a teacher's practices (Garza et al., 2016; Zepeda, 2005). However, in order for these practices to be effective, feedback must be provided in a timely manner and achievable for teachers; "Data suggested that building transparent interactions and professional relationships to create a culture of trust is necessary to improve the process of walkthrough observations as a collaborative endeavor" (Garza et al., 2016, p. 10). Teamwork and edification of the whole community is an important aspect of teacher evaluation.

There is currently limited information on to the effects of walkthrough observation as an evaluation tool for educators (Bushman, 2006). As principal, Buschman (2006) forewent structured observation tools such as the *Framework for Teaching* and conducted walkthroughs of the teachers in his high school; however, instead of completing these observations independently, he was accompanied by other teachers as a way for teachers to observe the practices of other teachers. Following these observations of math and science classrooms (approximately 8-10 observations with one accompanying teacher), Buschman and math and science teams discussed the instructional strategies of the team within specific department

teams and in professional development settings as well. At the end of the 2003-2004 school year, teachers were collaborating more frequently and cross-curriculum lessons and connections were established.

Walkthrough observations can vary in length, some lasting as long as 20 minutes while others may only take three minutes, but at the center of these observations is the intent to gather intel on teacher effectiveness and write an action plan for continued improvement (Garza et al., 2016). In a study conducted by Garza et al. (2016), aspiring teacher leaders share their perceptions of walkthrough observations and its potential to be an adequate form of teacher evaluation and promote student growth.

Garza et al. (2016) included current participants or recent graduates from an educational leadership program for prospective school principals, 59 individuals were invited to participate, while 22 completed the survey. The questionnaire included questions that related to the purpose, advantages, and challenges of conducting and using walkthrough observations as evaluation tools.

Results of the study reflected two types of walkthrough observations, which Garza et al. (2016) coined the Bureaucratic Approach and the Collaborative Approach. The Bureaucratic approach is a principal-dominated observational tool that affirms that the principal is the only one who has the expertise to evaluate teacher effectiveness and offer recommendations for teacher improvement. Participants of Garza et al. (2016) also emphasized the importance of observations being unscheduled and unannounced as they provide authentic teaching opportunities. The principals or school leaders providing feedback for walkthrough observations, therefore, need training and resources to provide teachers with adequate

feedback regarding teaching instruction and professional responsibilities. With the multitude of responsibilities principalship demands, observations with adequate feedback is frequently being neglected, resulting in teacher evaluation that is not promoting the growth of teacher abilities or student progress (Garza et al., 2016; Granada & Vriesenga, 2008).

The Collaborative Approach allows for team members and peers to observe the teaching techniques of coworkers and provide feedback. This shared-responsibility approach to teacher evaluation allows teachers to receive feedback in a collaborative manner; “Rather than passively receiving feedback, teachers were described as active members who generated, analyzed, reflected, and acted on walkthrough observation data” (Garza et al., 2016, p. 9). As a team, data from walkthrough observations are analyzed and discussed and a professional development plan is shared between parties. Findings of Garza et al. (2016) suggested that classroom walkthroughs conducted in a collaborative manner can promote teacher and student growth. “Data suggested that building transparent interactions and professional relationships to create a culture of trust is necessary to improve the process of walkthrough observations as a collaborative endeavor” (Garza et al., 2016, p. 10). Walkthrough observations can be an important indicator of teacher effectiveness in the classroom and can be used as a tool for evaluation when completed frequently and with feedback included.

Emergent Evaluation

Merit Pay

Positive classroom rewards influence the behavior of some students, but not all. Charts, stickers, and rewards for doing the right thing are nice, but at the heart of doing the right thing should be the interior desires of the individual, not the reward itself. Positive classroom

rewards and merit pay are similar, rewarding teachers for great test scores instead of focusing on the teaching behind the test scores. The additional pay for a job well done is appreciated and sometimes reflects the hard work of a teacher, but is creating an environment where teachers feel safe, respected, and cared for more important than an additional stipend in their paycheck? Many teachers do not receive enough for what they do in the classroom and how they care for students, but many teachers are working the required amount of hours of their position and do not deserve the same additional pay as others.

Research shows that teachers have a direct effect on the achievement of their students (Berg, 2010; Stronge, et al., 2011; Weisburg, Sexton, Mulhern, & Keeling, 2009). Qualified and educated teachers work hard for their students, fight for those they work with, and find success as their students do. According to DuFour and Mattos (2013) merit pay does not “improve student outcomes or change teacher behavior in a positive way,” rather it may unintentionally lower teacher standards (par. 8). When added pressure of pay is on the line according to student achievement on tests, some teachers will alter their teaching teach material that

Another important aspect of merit pay to consider is based on Maslow’s hierarchy of needs. On this pyramid, the top identifies the importance of self-actuation, the ability of one to self start and become the best he or she can be. Teachers crave this for their students and should desire it for themselves as well. Students achieve and struggle for a variety of reasons, especially if their basic psychological or safety needs are not being met, and to say that the teacher is the sole individual responsible for their achievement is painting a large picture using a small paintbrush. Many factors play into student success and one test cannot adequately define a student’s understanding or a teacher’s ability.

Ultimately, when making the decision of using merit pay, school leaders must evaluate the school environment, the motives of the teachers in the building, and the implications of utilizing merit pay. As identified by Green (2017), there are many different forms of decision making and each school and learning community is vastly different. There are many factors to consider when implementing merit pay structures.

In 2006, the Teacher Incentive Fund (TIF) was created by Congress to compensate teachers and principals who met certain standards in improving student growth and achievement in high-need schools (Chiang et al., 2018). As a form of incentive to improve teacher performance and therefore improve student achievement, this grant was awarded in 2010 to over 130 districts with high needs school in the country. Of those 130 districts, the study conducted by Chiang et al. (2018) evaluated ten of those districts that administered this program between the course of four years (between the school years of 2011-2012 through 2014-2015). Most of these districts were in southern parts of the country and urban in demographics.

The results of this study are threefold and reflect student improvement as a result of the implementation of this program. Of the ten sampled districts, results indicated a two percent increase in student achievement in reading during the second year of implementation of the program compared to schools that did not implement merit pay. This trend continued into the last two years of the study. In mathematics, the results were significant only in year one and indicated only slight increases in student achievement in the last years of the study. However, in both content areas the differences “were equivalent to about three to four weeks of learning” (Chiang et al., p. 7, 2018).

To encourage teacher improvement by providing extra bonuses for teachers who met specific goals, the goals were required to be challenging in achievement, differentiated according to grade level and/or subject, and substantial in size. However, approximately 70 percent of teachers and principals received their bonus, data inconsistent with the desire that the goal is challenging to obtain. Bonuses were differentiated between teachers, with teachers of high performance and test results earning more than the average. Unfortunately, many of the teachers, approximately 50-60 percent, were unaware of their opportunity to earn additional pay for performance, which did not improve for the remainder of the TIF program.

Secondly, there were four required components of the grant that districts were required to implement. The first of the four mandated that teachers are observed and student growth testing be used to evaluate teacher and school effectiveness. The second required that teachers receive bonuses based on performance. The next requirement was that teachers be offered additional professional responsibilities and receive compensation for their contributions. Lastly, teachers must be offered professional development opportunities to improve their performances on areas principals or school leaders have identified as areas for improvement. Beginning the first year and continuing for all following years, 90 percent of districts offered pay-for-performance opportunities and 85 percent reported having offered opportunities for additional pay. However, a smaller percent of districts (80 percent for teachers and 60 percent for principals) implemented the required training and measures for teacher and principal effectiveness. Fewer still, 59 to 74 percent of the districts offered teachers opportunities for professional development in their field.

At the end of the study, teachers were asked about their perception of the TIF program and their participation satisfaction. In each year of the program, approximately two-thirds of teachers reported being satisfied with their participation in the program and desire to continue with the requirements of the grant.

Lastly, the implementation of the grant is not sustainable for many of the districts currently participating in the program because of varying reasons. In response to these challenges, only 47 percent of the schools in the ten districts studied report continuing to offer pay-for-performance measures once the grant ends during the 2015-2016 school year. However, the other requirements of the TIF program will continue to be utilized by the districts, including measuring teacher effectiveness with feedback and rubrics (at least 80 percent), offering additional pay for extra responsibilities (74 percent), and providing professional development for teachers based on performance ratings (90 percent).

In regards to teacher effectiveness, Chiang et al. (2018) reported three main impacts. First of all, pay-for-performance teachers had slightly higher observational scores by the third year of implementation. The teachers who participated in this program improved their overall scores from 0.05 to 0.09 points higher on an observational score of 1 through 4 compared to the control group who did not receive a performance bonus. Secondly, the impacts of teacher improvement on observational scores did not seem to explain the effects on student achievement. Schools that had great numbers of pay-for-performance success did not have a strong positive correlation of student achievement scores. "Therefore, although classroom observation ratings detected some improvements in practices due to pay-for-performance,

they did not identify the improvements that were actually associated with student achievement” (Chiang et al., p. 15, 2018).

Effectiveness of Summative, Formative, and Emergent Practices

In a study of 477 Indiana principals regarding their practices in teacher evaluations and the effectiveness of formative, summative and emergent evaluation methods, Boyland, Harvey, Quick, and Choi (2014) determined that some forms of these are perceived as effective while others are not as effective. Elementary and secondary principals reported high fidelity with many summative methods and several formative methods, but emergent methods should be used with caution.

The form that principals utilized is known as the Indiana RISE rubric with two main components: student learning and professional methods. There are four domains with 23 teacher performance criteria, similar to Danielson. The domains include planning, instruction, leadership, and professional practices.

Overall, participants are using the summative practices nearly 80 percent of the time; these practices include pre and/or post conferences, prescheduled classroom observations, unannounced classroom observations, and rating scales with descriptions; however, rating scales without narratives are used a lesser amount.

Of the six formative evaluations practices, two of them are being used frequently, including non-scheduled classroom walkthroughs and goal setting by the teacher. Goal setting that is tied to school-wide goals is also being used but to a lesser degree (64 percent in elementary and 66 percent in secondary).

In regard to emergent practices, these methods are overall seen as ineffective and low fidelity. The only emergent practice that principals rated positively was the use of teacher effectiveness rubrics, such as that produced by Danielson in the *Framework for Teaching*. Though, in some cases, informal student and/or parent survey input have been seen as effective.

In summary, teacher evaluation is important and improves teacher effectiveness (Boyland et al., 2014). “Elementary and secondary participants strongly agree that instruction is a critical component in assessing teacher effectiveness, followed by professionalism, planning, and teacher leadership” (Boyland et al., par. 44, 2014). Teacher evaluation is an important aspect of education, but without clear expectations, feedback, and collaboration between staff members, teacher evaluation and growth does not happen as easily or frequently.

According to Chaplin, Gill, Miller and Regional Educational Laboratory Mid-Atlantic (ED) (2014), in a study of Pittsburg Public Schools during the 2011-2012 school year, school leaders evaluated the effectiveness of three teacher evaluation practices. Because data was tracked for multiple years, the number of teachers studied by Chaplin et al. varies, but approximately 8,332 teachers were studied in evaluating Pittsburg Public Schools’ evaluation methods. The first of which was the Research-based Inclusion System of Evaluation (RISE), based on Danielson’s *Framework for Teaching*. The second form of evaluation was student surveys referred to as 7Cs, founded on Ronald Ferguson’s research presented on the Tripod Project. Lastly, the third evaluation form was a value-added measure that used changes in student test scores to determine teacher effectiveness in student learning up to three years of teaching. These results

were used independently to measure teacher effectiveness, but also in unison to evaluate teacher effectiveness.

The first evaluation form known as RISE is composed of 24 components among four domains of Danielson's *Framework for Teaching* (which were previously listed in this review), but of the 24 domains listed, the Pittsburg schools identified 12 of these domains as "power components" (p. 1, 2014) that were thought to have the greatest impact on student learning. Each teacher received feedback through a rating scale of "Distinguished" (4), "Proficient" (3), "Basic" (2) or "Unsatisfactory" (1), as well as principal feedback given during conference time between the principal and teacher. The second evaluation known as the 7Cs and based on Ferguson's research involved students rating teachers according to a seven section survey. The measurable areas included: care for students, classroom management, clear delivery, challenge of learning, engagement of students, student dialogue and confer with students, and time management. Lastly, the third evaluation referred to as the value-added measure (VAM) was developed in association with Mathematical Policy Research and was calculated by grade level (including grades 4-12) and subject (science, math, reading/English, social studies).

This study of Pittsburg Public Schools had two main focuses; the first of which, was to differentiate low-performing teachers as well as high-performing teachers. The second was to identify the correlation between professional practice (RISE), student surveys (7Cs), and value-added measure of student scores. The premise of all three of these evaluation methods are focused on different criteria of an effective teacher, and therefore, when combined should not only determine the effectiveness of a teacher, but also show a positive correlation between all three measures.

Concluding results showed that while all three measures can be used to show teacher effectiveness, when all three are used in combination with one another, the measures show reliability to potentially differentiate among teachers. Because observations and the surveys administered require multiple raters and are subjective to the individual scoring the educators, the test scores are not as reliable as the value-added evaluations. “The study’s findings suggest that, taken together, RISE, 7Cs, and VAM ratings are useful and complementary measures of teacher effectiveness” (Chaplin et al., 2014, p. 3).

By using all three measures together, results reflected internal consistency and complementary between results. Overall, teachers with high RISE results also had high 7Cs ratings and value-added measures as well.

CHAPTER III: DISCUSSION AND CONCLUSION

Summary of Literature

Research has reflected that student learning is most impacted by teacher quality and efficacy (Berg, 2010; Stronge, et al., 2011; Weisburg, et al., 2009). Therefore, evaluation tools should be used to consistently measure teacher growth and provide opportunities for improvement (Marzano, 2012). However, teacher evaluation has been perceived as ineffective and inadequate, often being criticized for not adequately differentiating strong educators from the weak and failing to develop the practices of skilled teachers (Toch, 2008; Kimball & Milanowski, 2009; Marzano, 2012). Because of the history of inconsistent feedback to educators and perceptions of inconsistent and confusing information to many teachers, multiple forms of evaluation have emerged to provide teachers with guidelines for effective teaching as well as tools for growth.

As a summary of the literature review, teacher evaluation is categorized in three ways: summative, formative, and emergent (Boyland, et al., 2014). Summative evaluation includes standards-based evaluation, such as the commonly used tool of Danielson's *Framework for Teaching* (2011) and Marzano's *Marzano Teacher Evaluation Model* (2012). Both of these systems include rating scales of teacher ability and comprehensive range of skills. According to the research of Sartain et al. (2011), Kettler and Reddy (2017), Kane et al. (2011), Danielson's *Framework for Teaching* have provided a differentiated tool that can evaluate teacher ability in the classroom as well as providing resources for future growth. Marzano's form of evaluation, Marzano Teacher Evaluation Model (MTEM), has also shown positive correlation between teacher evaluation results and student test scores (Basileo & Toth, 2019). These tools both

encourage celebrations of success for teachers as they improve over time or achieve specific learning goals.

When the phrase “formative evaluation” is used, often collaboration will come to mind, as this evaluation method encourages groups working together, such as principals and teachers, school leaders and staff, and teachers and learners. Professional Learning Communities (PLCs) encourage space for teachers to collaborate, share ideas, and challenge one another (DuFour & Mattos, 2013; Green, 2017). Research has shown that positive correlations exist between teacher collaboration and student achievement (Leana & Pil, 2009). Another practice of formative evaluation includes walkthrough observations, which also encourage collaboration and may allow for authentic evaluation of educators (Garza et al., 2016; Zepeda, 2005).

Lastly, emergent evaluation includes incentives for a “job well done,” encouraging teachers to do a good job to earn compensation above their normal pay or other tangible rewards for performance. Of the research studied, results of emergent evaluation did not provide sufficient data to encourage student growth as a result of teacher reward (Chiang et al., 2018; Boyland et al., 2014; DuFour & Mattos, 2013). Ultimately, the hope of any of these forms of teacher evaluation is that the tool is differentiated between educators and provides opportunities for maximized student learning.

Professional Application

The information in this literature review allows for application in multiple forms, specifically for those in education positions and leadership in elementary, middle, and high schools across the United States. Possible application pieces include using multiple forms of

evaluation to determine teacher ability and differentiate between teachers, utilizing teamwork to encourage teacher evaluation and growth, and providing opportunities for growth among school leadership in conducting effective evaluations as well as celebrating school success. Ultimately, depending on the school and leadership, the forms of evaluation identified in the literature review have the potential to be effective given the appropriate context and resources (Boyland et al., 2014).

There are multiple forms of evaluation that can be used to evaluate teaching practices and promote professional growth. Because there are a wide variety of opportunities for school leaders to evaluate teachers, each school or district can use one or a variety of tools to evaluate educators. Gardner (1999) advocated for differentiated instruction according to student individual needs, and this form of differentiation can be advocated for educators as well. School leaders can choose evaluation tools according to the personalities, needs, and opportunities of the educators in the building.

Lastly, each of the summative, formative, and/or emergent evaluation forms can be adapted to allow for conversation between educators and school leaders to improve teacher ability and promote growth. By promoting conversation between school leaders and educators, an environment of collaboration can be created. According to Sinek (2014), strong leadership includes empathy, relationships, and servanthood. School leaders can use evaluation as a gateway to connect with educators, build relationships, and show compassion. In summary, teamwork is a critical point of any business or organization, including schools. As a form of evaluation, working together as a team can better the school climate and encourage student growth and progress (Danielson, 2007; DuFour & Mattos, 2013; Garza et al., 2016).

Lastly, school leaders must be well-equipped to adopt any form of teacher evaluation, seeking out opportunities for training and then provide resources for educators before conducting evaluations. As evident by the research of Sartain et al. (2011) and Garza et al. (2016), training of school leaders to provide effective feedback to educators is important in delivering effective feedback and therefore promoting teacher growth. Teacher evaluation is an important aspect of education because teachers directly affect the learning and opportunities for their students.

Limitations of Research

A common critique of teacher evaluation is inconsistency of evaluators and/or biases of evaluators. As previously addressed, training of school leaders can be utilized to attempt to lessen the impact of inconsistencies, but unknown biases may still exist or impact scoring of many forms of evaluation, including summative evaluation. Formative and emergent feedback can be addressed by a variety of individuals instead of just one school leader or principal. This limitation of teacher evaluation was not addressed in the literature conducted in this research.

Another limitation of the research is the use of value-added measures and how this form of evaluation can be skewed because the data is affected by unknown or unaccounted for variables. Value-added measures are used to evaluate a teacher's effectiveness based on the scores of his/her students in previous years. However, this has often been criticized because the comparisons between years of learning do not account for the subtle ways of how students are different and experiences that affect learning over the course of multiple years.

Thirdly, the research conducted on the Marzano Teacher Evaluation Model (MTEM) involved schools in Oklahoma and Florida, two states that are experiencing challenges relating

to high need of resources, funding, and teacher development. Though these schools reported satisfaction with MTEM, they are experiencing challenges as a state in providing rigorous and consistent education to students.

Lastly, the keywords used in the literature review included the terms “teacher evaluation,” “summative evaluation,” “formative evaluation,” “emergent evaluation,” and “professional development.” The information of this thesis was limited to peer-evaluated journals and articles that had been published between 2000 and 2020, as they provided the most relevant information on the practices of teacher evaluation. The information of the literature review did not include teacher evaluation from countries other than the United States; this was because I wanted to provide information that was current to American practices. Data was also collected from a variety of states and regions in the United States with schools that had varying socioeconomic status and diversity. In summary, the research of this thesis was conducted with the hope to provide relevant and important information on teacher evaluation.

Implications for Future Research

In most professions and possibly all professions, individuals want to succeed and impress others on their performance or knowledge. In education, there are some teachers who seek this. Future areas of research could include the frequency with which educators take advantage of evaluation to improve pay or receive undeserved recognition.

Another form of research to be conducted is greater depth of summative, formative, and emergent evaluation forms. Areas not addressed in this literature review include the

formative evaluations of goal setting and peer coaching; emergent practices of portfolio evaluation, and greater depth of teacher evaluations from students and/or parents.

Conclusion

In summation, all of these forms of evaluation can be effective given the appropriate context and educators (Boyland, et al., 2014). Summative, formative, and emergent evaluation methods have varied strengths and weaknesses, and therefore research recommends using a combination of evaluations in addition to emergent evaluation to evaluate teacher effectiveness (Boyland, et al., 2014; Chaplin et al., 2014). However, one piece of information remains consistent through each of these forms of evaluation, feedback must also be given in adequate time following an observation, whether formal or informal (Danielson, 2011; Garza et al., 2016). With timely and effective feedback, the ultimate hope is that teacher evaluation serves as a way for teachers to apply the word *sankofa* and not only measure their efficacy as teachers, but also improve their work.

References

- Basileo, L. D., & Toth, M. (2019). A state level analysis of the Marzano teacher evaluation model: Predicting teacher value-added measures with observation scores. *Practical Assessment, Research & Evaluation, 24*(6).
- Berg, J. (2010). Constructing a clear path to accomplished teaching. *Theory Into Practice, 49*(3), 193–202. doi: 10.1080/00405841.2010.487754
- Boylard, L., Harvey, M., Quick, M., & Choi, Y. (2014). Teacher evaluation in Indiana: Principals' use and perception of field-based practices. *International Journal of Educational Reform, 23*(4), 270.
- Bradley-Levine, J., Romano, G., & Reichart, M. (2017). Teacher leaders' influence on teachers' perceptions of the teacher evaluation process. *International Studies in Educational Administration (Commonwealth Council for Educational Administration & Management (CCEAM)), 45*(3), 65–84. Retrieved from <https://search-ebscohost-com.ezproxy.bethel.edu/login.aspx?direct=true&db=aph&AN=128198212&site=ehost-live&scope=site>.
- Bushman, J. (2006). Teachers as walkthrough partners. *Educational Leadership, 63*(6), 58-61.
- Chaplin, D., Gill, B., Thompkins, A., Miller, H., & Regional Educational Laboratory Mid-Atlantic (ED). (2014). Professional practice, student surveys, and value-added: Multiple measures of teacher effectiveness in the Pittsburgh public schools. REL 2014-024. *Regional Educational Laboratory Mid-Atlantic*. Regional Educational Laboratory Mid-Atlantic.
- Chiang, H., Speroni, C., Herrmann, M., Hallgren, K., Burkander, P., & Wellington, A. (2017).

Evaluation of the teacher incentive fund: Final report on implementation and impacts of pay-for-performance across four years: Executive summary. NCEE 2018-4005. *National Center for Education Evaluation and Regional Assistance*.

Danielson, C. (2011). Evaluations that Help Teachers Learn. *Educational Leadership*, 68(4), 35-39.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2012). Observing classroom practice. *Educational Leadership*. 70(3), 32-27.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. New York, NY: Teachers College Press.

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

Retrieved from <https://search-ebSCOhost-com.ezproxy.bethel.edu/login.aspx?direct=true&db=eric&AN=EJ1023870&site=ehost-live&scope=site>

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

Ebmeier, H. (2003). How supervision influences teacher efficacy and commitment: An investigation of a path model. *Journal of Curriculum and Supervision*, 18(2), 110-141.

Gardner, H. (1999) *Intelligence reframed: Multiple intelligences for the 21st century*. New York, NY : Basic Books.

Garza, R., Ovando, M., & O'Doherty, A. (2016). Aspiring school leaders' perceptions of the

- walkthrough observations. *International Journal of Educational Leadership Preparation*, 11(1).
- Granada, J., & Vriesenga, M. (2008). Web-based walkthroughs. *Principal Leadership*, 8(7), 24-27.
- Green, R.L. (2017). *Practicing the art of leadership: A problem-based approach to implementing the professional standards for educational leaders*. Boston: Pearson.
- Griffin, L. (2013). Charlotte Danielson on teacher evaluation and quality. *School Administrator*, 70(1), 27-31.
- Hager, K. D. (2018). Teachers' use of video self-monitoring to improve delivery of effective teaching practices. *Teaching Exceptional Children*, 50(5), 283–290. doi: 10.1177/0040059918765749
- Kane, T. J., Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation.
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, 46(3), 587–613.
- Kettler, R. J., & Reddy, L. A. (2019). Using observational assessment to inform professional development decisions: Alternative scoring for the Danielson framework for teaching. *Assessment for Effective Intervention*, 44(2), 69–80.
- Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34–70. doi: 10.1177/0013161X08327549

- Leana, C. & Pil, F. (2006). Social capital and organizational performance: Evidence from urban public schools. *Organization Science*, 17(3), 353-366.
- Leana, C. & Pil, F. (2009). Applying organizational research to public school reform. *The Academy of Management Journal*, 52(1), 101-124.
- Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership*, 70(3), 14-19.
- Marzano, R. (2017). *The new art and science of teaching*. (Revised and expanded ed.). Bloomington, IN: Solution Tree Press.
- Marzano Research Laboratory. (2011). What works in Oklahoma schools: A comprehensive needs assessment of Oklahoma schools: Phase 2 report. Retrieved from <http://sde.ok.gov/sde/sites/ok.gov.sde/files/SIPhaseIIStateReport.pdf>.
- Maslow, V. J., & Kelley, C. J. (2012). Does evaluation advance teaching practice? The effects of performance evaluation on teaching quality and system change in large diverse high schools. *Journal of School Leadership*, 22(3), 600-632.
- Northwest Evaluation Association . (2011, January). Technical manual for measure of academic progress and measure of academic progress for primary grades. Portland, OR: Author.
- Quinn, A. E. (2014). Looking at the bigger picture with Dr. Robert Marzano: Teacher evaluation and development for improved student learning. *Delta Kappa Gamma Bulletin*, 81(1), 12–18. Retrieved from <https://search-ebSCOhost-com.ezproxy.bethel.edu/login.aspx?direct=true&db=aph&AN=98474481&site=ehost-live&scope=site>

Roegman, R., Goodwin, A. L., Reed, R., & Scott-McLaughlin, R. M., II. (2016).

Unpacking the data: An analysis of the use of Danielson's (2007) "framework for professional practice" in a Teaching Residency Program. *Educational Assessment, Evaluation and Accountability*, 28(2), 111-137.

Schiff, D., Herzog, L., Farley-Ripple, E., & Iannuccilli, L. T. (2015). Teacher networks in Philadelphia: Landscape, engagement, and value. *Penn GSE Perspectives on Urban Education*, 12(1).

Sinek, S. (2014). *Leaders eat last: Why some teams pull together and others don't*. New York, NY: Penguin Group.

Stronge, J., Ward, T., & Grant, L. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, 62(4), 339-355.

Struyve, C., Meredith, C., & Gielen, S. (2014). Who am I and where do I belong? The perception and evaluation of teacher leaders concerning teacher leadership practices and micro politics in schools. *Journal of Educational Change*, 15(2), 203-230.

Toch, T. (2008). Fixing Teacher Evaluation. *Educational Leadership*, 66(2), 32–37. Retrieved from <https://search-ebscohost-com.ezproxy.bethel.edu/login.aspx?direct=true&db=eric&AN=EJ814363&site=ehost-live&scope=site>

Viadero, D. (2007). New center asks: Does merit pay work? *Education Week*, 26(28). Retrieved

from <https://search-ebSCOhost-com.ezproxy.bethel.edu/login.aspx?direct=true&db=eric&AN=EJ759599&site=ehost-live&scope=site>.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *Education Digest: Essential Readings Condensed for Quick Review*, 75(2), 31-35.