2023

Evaluators' Perceptions on the Quality of Training For Conducting Classroom Observations

Mark Williams
Walden University

Walden University

College of Education and Human Sciences

This is to certify that the doctoral dissertation by

Mark Allen Williams

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. John Harrison, Committee Chairperson, Education Faculty
Dr. Darci Harland, Committee Member, Education Faculty
Dr. David Moffett, University Reviewer, Education Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2023

Abstract

Evaluators' Perceptions on the Quality of Training for Conducting Classroom Observations

by

Mark Allen Williams

MA, Saginaw Valley State University, 2001

BS, Central Michigan University, 1993

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Walden University

February 2023

Abstract

There is a problem in the state of Michigan's evaluation system regarding the training received by those conducting classroom teacher observations. A performance gap between teachers and students and provides training solutions for classroom observers was identified. The purpose of this basic qualitative study was to explore the perceptions of school administrators regarding their preparedness and ability to conduct effective observations of teachers. Social cognitive theory and its six constructs was the conceptual framework used for this research study. Research questions addressed the perceptions of school administrators about the training they have received and their need for additional training. A basic qualitative design with interview questions via Zoom was used for this study. Thirteen participants represented high school, junior high, and middle school levels. Through the coding process, the data collected revealed two major themes, future training, and training needs, with subthemes for each theme. Results showed themes around future training needs, networking with colleagues, practical training and debriefing with colleagues, and the need for refresher training. Recommendations included continued ongoing training and networking with colleagues through an online platform. Training practices identified in this study may create a positive school building, district, and community culture leading to a positive social change in the relationships of all school and community stakeholders.

Evaluators' Perceptions on the Quality of Training for Conducting Classroom Observations

by

Mark Allen Williams

MA, Saginaw Valley State University, 2001

BS, Central Michigan University, 1993

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Walden University

February 2023

Acknowledgments

I would like to thank Walden University and the faculty members that have worked so hard to assist me through this journey. All were there to support, guide and encourage me when I experienced some of the most difficult times in my life. Thank you, Dr. Norma Clark, Dr. Michael Burke, Dr. Janet Strickland, and Dr. Rochel Michael. Since this journey has taken forever there are many who have come in and got me through the finish line including Dr. John Harrison my chairperson and team leader, Dr. Barbara Schirmer my methodologist and positive support, my URR committee member Dr. David Moffett, and Dr. Mark Clauburg, program administrator. Without their guidance, I would have never completed this journey. I also want to thank my family. My wife Melissa. She is the foundation in my life and has been with me when I reached bottom in my life and will be forever. I want to thank my 3 children Brandon, Bryce, Shay, and 2 stepchildren Brian, and Aubre. You impact my life in a positive way each and every day. During this process, time and attention were taken away from all of them; I hope that my love and admiration for each of them makes up for the long hours apart. This journey is proof that when you set your sights on lofty goals, they can be achieved. I would also like to thank my work colleagues. Without them and their understanding, this journey would have never been completed. Finally, I am who I am and what I am today because of my parents Joan and the late Bill Williams. Their love has always been unconditional. Writing this dissertation has been the most challenging task of my academic career, and until recently, I never thought I would finish.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Teachers and administrators in the education profession across the United States have experienced a change in the way teaching and teachers are evaluated (Walsh et al., 2017). A significant research study funded by the Bill and Melinda Gates Foundation was conducted by the Measures of Effective Teaching (MET) Project and produced a report titled *The Widget Effect* (Weisberg et al., 2009). This study involved a research partnership of academics, teachers, and educational organizations committed to investigating better ways to identify and develop effective teaching. The study focused on effective teacher observations and evaluation ratings over a 2-year period starting in 2009. The studies reviewed showed that there were bigger differences in teaching quality among teachers within schools than there are between schools. This means that in the same school, a child taught by a less effective teacher can receive an education of vastly different quality than a student in another class who is taught by a more effective teacher. The study also pointed out the way evaluations are currently conducted does not provide a teacher who is struggling with a road map to improve (Weisberg et al., 2009).

According to Public Act (PA) 102 of, 2011, as of September 1, 2011, all Michigan school boards were mandated to adopt and implement a rigorous, transparent, and fair performance evaluation system for all teachers and administrators. The law amended Section 1249 (MCL.380.1249) and added Sections 1248 and 1249a, which addressed a statewide evaluation system for teacher evaluation applicable to all teachers, central office personnel, and school building principals in the state of Michigan. According to PA 102, individuals performing classroom teacher observations are

required to perform additional duties related to evaluations such as providing timely feedback, developing specific performance goals, and offering related training for teachers' professional development.

Since the passage of PA 102, teachers in the state of Michigan are currently evaluated on a four-category rating system of highly effective, effective, minimally effective, and ineffective. This is a change from the binary rating system of satisfactory or unsatisfactory (PA 173, Section 380.1249 (1)(c) of 2015). According to the Michigan Common Law 380.1248(1)(b)(c), which passed in 2011, the final teacher effectiveness ratings are determined by the following criteria (listed in order of significance): individual performance of the classroom teacher; student growth; classroom management skills; pedagogical skills; teacher attendance; teacher discipline record; significant accomplishments or contributions; and relevant special training outside the classroom.

In this study, I addressed the gap in the literature that addresses the quality and effectiveness of the observer training for teacher evaluation. The perceptions of administrators regarding the quality of their observer training were a focus of this study, specifically as they relate to the practices of observers who were using the 5 Dimensions Plus Teacher Evaluation Rubric (University of Washington Center for Educational Leadership, 2012). Understanding the quality of training, the self-efficacy and preparedness of observers was the gap this study addressed. Michigan legislation mandated that starting with the 2016–17 school year, all school districts are required to ensure that training is provided to all observers according to Public Act 173 of 2015, Michigan Compiled Laws, § 380.1249 (2) (n) (2015). This training may be provided by

an individual who has expertise in the evaluation tool or an individual who has been trained to train others.

According to the law, for each individual teacher, effective evaluators use classroom teacher evaluation data gathered from multiple classroom observations to inform decisions concerning teacher effectiveness and teacher promotion; grant tenure or teacher retention; provide relevant coaching to teachers; provide additional instructional support and professional development for teachers in need of support; and remove ineffective tenured and nontenured teachers (MCL 380.1249). Classroom observations are conducted multiple times per year for each classroom teacher depending upon their previous evaluation ratings.

According to Loewus (2017), in 2011, Michigan ranked 23rd in student achievement test scores and the state received a grade of C+. Six years later, in 2017, Michigan ranked 34th and had an average grade of D. This decrease in ranking and achievement calls attention to how teachers are evaluated. Despite continued poor student performance ranking, over 97% of teachers' ratings continue to be "effective" or "highly effective" (Michigan Department of Education, n.d.) There is a problem that needs to be researched in the state of Michigan's education system regarding the ability of observers to identify or report the true levels of effectiveness of teachers when conducting classroom teacher observations and final evaluations as stated by Kraft and Gilmour, (2017), Norman (2010), Steinberg and Sartain (2015), and Weisberg et al. (2009). Specifically, the problem stems from the effectiveness of the training and the levels of

training received by those individuals conducting teacher observations and final evaluations (Derrington, 2014).

For this study, I gathered data that allowed me to examine observers' perceived preparedness to implement the entire cycle of classroom teacher evaluation process, including the final rating given to each classroom teacher. my qualitative study includes interviews of individuals who currently perform classroom teacher observations and teacher final evaluations using the 5 Dimensions Plus Teacher Evaluation Rubric approved by the state of Michigan. The results can be used by observers to improve the quality of teacher observation training received, leading to an increase in the capacity of observers to impart strategies for increasing teacher effectiveness, leading to increased student achievement (see Marzano, 2003).

In the following sections of this chapter, I will include a brief overview of the teacher evaluation process, identify the existing challenges driving the research, I have identified the research questions and purpose of the research study, and the importance and potential social change implications of the study. Additionally, I will provide a description of the conceptual framework and the nature of the study.

Background

Teacher evaluation has become a significant part of teaching in Michigan since 2001 with the passage of the No Child Left Behind Act of 2001. This act has since been replaced with the Every Student Succeeds Act of 2015, which also emphasizes the importance of quality teachers and the evaluation of all teachers (114th Congress: Every Student Succeeds Act, 2015). In the state of Michigan, Public Act 173 of 2015

emphasizes the teacher evaluation process and its significance for continued employment and improved performance. The literature and the state laws and policies identified a need for ongoing, quality training for those responsible for implementing any teacher evaluation system.

Weisberg et al. (2009) provided the foundation for working on better teacher evaluation practices, bringing attention to the fact that over 94% of teachers were found to be rated as satisfactory. In 2017, the Center for Educational Performance and Information (CEPI) in the state of Michigan found over 97% of teachers were rated as either highly effective or effective over a span of 6 school years ending that year. During this same period, evaluator training was not required by law in Michigan, and observer participation in training was voluntary. According to CEPI, this lack of training in the past may be linked to the high percentage of teachers being ranked effective or highly effective while at the same time student achievement in Michigan has been declining compared to other states. Findings from the research literature indicate the need for additional ongoing quality training for observers conducting classroom teacher observations to implement any teacher evaluation system (Cosner et al., 2015; Dodson, 2015b; Smylie, 2014).

The 5 Dimensions Plus Teacher Evaluation Rubric has been used and recognized in the state of Michigan since the 2012–13 school year as an approved teacher evaluation instrument (The Michigan Council of Educator Effectiveness, 2013), The Michigan Council of Educator Effectiveness (MCEE), created by PA 102, was created to develop a fair, transparent, and feasible evaluation system for teachers and school administrators.

The main goals of MCEE are to enhance instruction, improve student achievement, and support ongoing professional learning (MCEE, 2013). To achieve these goals, the MCEE piloted the 5 Dimensions Plus Teacher Evaluation Rubric (University of Washington Center for Educational Leadership, 2012) along with Danielson's the framework for teaching (Danielson, 2013), Marzano's teacher evaluation model (Marzano et al., 2011), and the thoughtful classroom teacher effectiveness framework (Gargani, J., & Strong 2014. The state of Michigan followed the council's recommendations, one of which was allowing all local school districts to decide which of the four evaluation tools they wish to use.

In complying with Michigan law, particularly PA 173, of 2015, the 5 Dimensions Plus Teacher Evaluation Rubric consists of two sections. The first section consists of observation and evaluation of traditional classroom performance traits. The observed areas may be classroom management, student assessment, student engagement, classroom culture, curriculum, and pedagogy, and learning targets. Subsections A, B, C, and D of Section 1248 of the Michigan revised school code describe in detail the areas in which a teacher will be evaluated. Teacher observer training was explained in detail in Section 1248 of the Michigan Law; training requirements are in addition to the professional development or continuing education that is currently required by the state of Michigan for all educators. The second section of the observation consists of a student growth component made up of state and local assessments. Currently the student growth component of the observation is a minimum of 40% of the total evaluation. (PA 173 of 2015).

While observation and evaluation legislation in the state of Michigan has required a much more elaborate teacher evaluation process, including observer training, teacher evaluation ratings have remained relatively the same since 2009 (Michigan Department of Education, n.d.). Observers who have the responsibility of final evaluation ratings for teachers must be properly trained to fully implement the teacher evaluation process while differentiating teacher effectiveness.

Findings from the research literature support the need for additional training for observers who are using a variety of teacher evaluation rubrics and conducting classroom teacher evaluations (Cosner et al., 2013; Dodson, 2015b; Donaldson et al., 2014; Michigan Council of Educator Effectiveness, 2013; Semmelroth & Johnson, 2014; Smylie, 2014). Dodson (2015b) (year) conducted a survey of Kentucky principals' perceptions of the state's new teacher evaluation system and found more training was needed from the state on how to use the new evaluation system. Dodson also found over 70% of respondents said the observer preparation program they completed did not prepare them well for the implementation of the evaluation system, and over half said the state department of education did not provide adequate training for proper implementation. Cosner et al. (2015) conducted a study of evaluation practices in three states—Connecticut, Tennessee, and Ohio—and found that observers needed extensive training and ongoing support to develop necessary understanding and skills to implement the new teacher evaluation systems. The researchers concluded that the quality of the training is much more important than the quantity.

In a phenomenological qualitative research study which sought to explore, describe, and analyze the meaning of individualized teachers' experiences with the new teacher evaluation system currently in place in the state of Michigan, Stewart (2016) found that teachers adapted their professional practice to meet new teacher evaluation requirements. Although this study addressed teacher perceptions of the teacher evaluation system in the state of Michigan, there are possible links to the experiences of observers.

There is a gap in the literature that addresses the quality and effectiveness of the observer training for teacher evaluation. In this research study, I gathered data pertaining to the perceptions of administrators about the quality of their training. The findings may be used to improve the effectiveness of training for classroom observers.

Problem Statement

There is a problem across the United States, including in Michigan's education system, regarding the ability of observers to identify and report the true levels of effectiveness of teachers when conducting classroom teacher observations and final evaluations (Kraft & Gilmour, 2017; Norman, 2010; Steinberg & Sartain, 2015). Specifically, the problem stems from the effectiveness of the training and the levels of training received by those individuals conducting classroom teacher observations and final evaluations. Research supports that the quality of a student's teacher is the most important element in a school to increase student achievement (Lasswell et al., 2008; Steinberg, & Sartain, 2015; Weisberg et al., 2009). Therefore, increasing the effectiveness of the observer may lead to an increase in the effectiveness of the teacher

while at the same time effective observers are better prepared to identify ineffective teachers for dismissal.

Beginning in 2011, and through the Every Student Succeeds Act (2015), Michigan legislation has placed an emphasis on classroom teacher performance and teacher quality using an elaborate teacher evaluation system. These legislative actions that place an emphasis on classroom teacher evaluation bring to the forefront the quality or effectiveness of the training received by observers conducting the observations and final evaluation ratings. If the ultimate goal of the teacher evaluation system is to improve student achievement through teaching, observer training may be the foundation needed.

Strunk et al. (2014) stated that rater leniency among observers is also a factor in teacher final ratings, while Cosner et al. (2015) along with Kraft and Gilmour (2017) found that increased workloads, along with time demands, cause observers to struggle to complete every element on the assessment and implement all the required observations. Data regarding observers' training to address such leniency and time demands were gathered during the interview process for this study. Archer et al. (2016) found there was no magic number of hours of training needed to ensure accuracy; however, it was noted that continued training is necessary to become an effective observer.

Educational research has shown that there is a lack of training for observers to identify or differentiate different levels of teacher effectiveness. As early as 2003, research has shown that when students are in a classroom with an effective or highly effective teacher, their student achievement performance will increase more than the

performance of a student who has a minimally effective or ineffective teacher (Marzano, 2003). Student achievement will compound and increase every year for students in classrooms with effective or highly effective teachers compared to students who have minimally effective or ineffective teachers (Kane et al., 2011a; Lasswell et al., 2008; Michigan Council for Educator Effectiveness, 2013; Steinberg & Sartain, 2015).

The quality of training observers receive can directly impact teachers' performance and, indirectly, students' performance. It was the data gathered and analyzed in this research study that led to targeted professional development for classroom observers to improve the effectiveness of teacher performance which may lead to the increase of student achievement. Jacob (2012) stated that school leaders have long been aware of the potential impact of an excellent teacher on student achievement outcomes. In a study done in the Chicago Public Schools, 93% of teachers were rated as superior or excellent, the top two categories in the rating system, yet two-thirds of the schools in the district failed to meet state proficiency standards under Illinois's accountability system (Steinberg & Sartain, 2015). The study found the performance gap was a result of a lack of training, quality, and level of training for individuals conducting classroom teacher observations and final evaluations.

Similarly, in a study conducted by the University of Connecticut concerning the implementation of the System for Educator Evaluation and Development (SEED), researchers found that principals called for more professional development to learn how to use the evaluation tool. Teacher survey and interview data from that study indicated that there was substantial variability in the perceived skill level of the evaluators

(Donaldson et al., 2014). On average, less than 1% of teachers were rated as below effective, and 4% were rated as developing in the pilot districts. Seventy-three percent were rated as proficient and 23% were rated as exemplary (Donaldson et al., 2014). Recommendations from the University of Connecticut study include building the skills of the observers by offering specialized training for observers. Upon completion of such professional development, it was found that only 60% of teachers and administrators felt that their summative ratings under SEED program were accurate (Donaldson et al., 2014).

The current evaluation data in education, particularly teacher evaluation, focus the attention on final teacher ratings and the evaluation process used by administrators when conducting classroom teacher observations and final evaluations. Consistency among these observers must be established to validate the evidence required to ensure any teacher observation evaluation system is appropriate for evaluating teacher performance (Semmelroth & Johnson, 2014). Observer reliability encompasses many aspects such as rubric knowledge, knowledge of proven research-based instructional strategies, and knowledge of content standards and benchmarks, along with extensive knowledge of the teacher evaluation tool that comes with extensive ongoing professional development and training.

The results of my study may contribute to the body of knowledge needed to address the quality of training by identifying specific targeted professional development training practices that will raise the capacity of observers to differentiate effective levels in teaching, leading to increased student achievement. In this study, I examined the

experiences of observers pertaining to the quality of the training acquired to properly observe and evaluate teachers. Current research has not identified the best practices for effective ongoing observations, but it has addressed the quality of training provided to classroom observers, by recommending additional training (Smylie, 2014). However, the quality, effectiveness, and the duration of the training were absent from any research. Using the research questions identified, this study addressed the gap in the literature that exists on addressing the quality and effectiveness of the observer training for teacher evaluation in Michigan, specifically as it relates to the practices of observers who are using the 5 Dimensions Plus Teacher Evaluation Rubric. The data gathered in the interview process brings a better understanding of the training received by observers, as required by Michigan PA 173 of 2015, as well as the level of training and preparedness. This study is relevant to the field of education because it gathers data and evidence regarding the training of observers, leading to their ability to assist teachers to improve.

Purpose of the Study

The purpose of this basic qualitative research study was to explore the perceptions of school administrators regarding their preparedness and ability to conduct effective observations of teachers. Data analysis led to identifying trends, patterns, and strategies currently in place for observer training that are perceived to be effective and those that are not effective. The analysis of the data identified observer perceptions of training practices currently in use in observer training that are effective or ineffective. In this study, I collected data surrounding the skills and perceptions of school administrators and the practices used to implement the 5 Dimensions Plus Teacher Evaluation Rubric to

actively observe, evaluate, and rate teachers. The past and current research data regarding teacher effectiveness and student achievement in the state of Michigan and across the country requires a closer look into the quality and effectiveness of training and skill set needed to differentiate good teaching from bad, including the identification of those teachers in need of extra support (Rowna et al., 2013; Weisberg et al., 2009). Understanding the quality of training, the self-efficacy and preparedness of observers was the gap that this study addresses. Steinberg and Sartain (2015) examined the Excellence in Teaching Project used in Chicago Public Schools, and found the efficacy of the project depended on the observers' capacity to provide targeted instructional guidance. This means identifying and differentiating a teacher's instructional needs. Research shows the level of teacher effectiveness is related to the student's academic effectiveness (Michigan Council for Educator Effectiveness, 2013; Steinberg & Sartain, 2015; Weisberg et al., 2009). If this were true for the state of Michigan, the students in the state of Michigan would be performing higher than a D as measured by state assessments.

In this qualitative research study, I explored and analyzed observer experience while conducting classroom observations, the time commitment needed outside the classroom to complete the observation including feedback and any documentation required, the amount of training received working with the 5 Dimensions Plus Teacher Evaluation Rubric, and the practices of participants and their perceptions of preparedness for conducting classroom teacher observations and final evaluations. The data collected represents the perceptions of the administrative observers on the specific training practices.

Research Questions

RQ1. What are the perceptions of school administrators who are responsible for evaluating teachers about the additional training they need for performing effective classroom observations?

RQ2. What are the perceptions of school administrators who are responsible for evaluating teachers about the training they have received for performing classroom observations?

Theoretical Framework

For my study, I viewed the observation process through the lens of social cognitive theory (SCT). Understanding SCT and its practices enabled me to view observer training and practices currently used by teacher observers through a lens that emphasizes the environment in which the observers work along with their capacity to change their behavior, and the self-confidence to make that change. The framework used for this study, originally titled social learning theory in the 1960s by Bandura (LaMorte, 2018), has developed into what is now known as SCT. SCT has a framework that suggests learning occurs in a social context with a dynamic and reciprocal interaction of the person, environment, and behavior (LaMorte, 2018). The unique feature of SCT is the emphasis on social influence, as well as external and internal social reinforcement. The theory considers the way in which individuals acquire and maintain behavior, while also considering the social environment in which individuals perform the behavior (LaMorte, 2018). I touched each of the six constructs of SCT in various depths while focusing on observer training: reciprocal determinism, behavioral capability, observational learning,

reinforcements, expectations, and self-efficacy. The six constructs can be briefly defined as follows. Reciprocal determinism is the dynamic and reciprocal interaction of a person with a set of learned experiences (LaMorte, 2018). Behavioral capability is a person's ability to perform a behavior (LaMorte, 2018). Observational learning is where people can witness and observe a behavior conducted by others (LaMorte, 2018). Reinforcements are the internal or external responses to a person's behavior (LaMorte, 2018).. Expectations refers to the anticipated consequences of a person's behavior. Self-efficacy refers to the level of a person's confidence in his or her ability to successfully perform a behavior (LaMorte, 2018).

I used each of these constructs to view the responses of the participant's interview. Using reciprocal determinism as a lens to determine the actions and interactions observers have with teachers and the capability of the observer to change their own observation behavior to become more effective is one of the benefits of SCT. I gathered data regarding the observational learning of observers by asking questions regarding training and their experiences with watching videos of teachers and practicing those traits. I also engaged in conversation regarding observer expectations and their self-efficacy to perform those expectations. Using the six constructs of SCT for this research study, I was able to gather data focusing on the training and the strategies used to implement that training through the lens of each construct. Through the research questions and expanding interview questions I gathered data regarding the interactions of training, the capability of making a behavioral change for observers, learning capability through observation, the reinforcement of the new learned behavior to implement

processes, the consequences of not meeting expectations, and the self-confidence that a new behavior can be learned and implemented. More detail regarding specific strategies concerning SCT will be provided in Chapter 2.

In this qualitative study I used semi structured interviews as the data collection method. I used the data to consider both the unique way in which individuals acquire and maintain a particular behavior in the social environment and the way individuals perform that behavior. By using the lens of behavioral capability, observational learning, expectations, and self-efficacy, analysis of participant responses was more accurate. The focus was on current observation practices to identify levels of effectiveness in teacher performance along with practices that need to be changed or are not beneficial to the observation. One of the goals of SCT is to explain how people regulate their behavior through control and reinforcement to achieve goal-directed behavior that can be maintained over time (LaMorte, 2018). Through the interview process, questions were posed to participants stemming from the research questions regarding their emotional frame of mind as well as the related experiences and behaviors they may display during the training received and while completing observations. Understanding the experiences, related behaviors, and self-efficacy during training was beneficial in identifying practices and strategies observers are comfortable and uncomfortable implementing. Identified effective practices and strategies can then be incorporated into future and ongoing observer training to increase the effectiveness of the training delivered to and received by observers.

Nature of the Study

This basic qualitative study was focused on the teacher evaluation process and the training of individuals who conduct classroom observations and final evaluations of teachers in the state of Michigan. Focusing on the training opportunities available, the strategies used in training, and the perception and preparedness of observers, data collected and analyzed led to results that could be used to improve the training procedures and processes currently used in observer training. For the purpose of this research study, individuals who are conducting classroom teacher observations and final evaluations were referred to as observers. A qualitative research study enables a researcher to gather direct information from participants regarding their experiences (Creswell, 2009). In this study, I gathered information into the participant's experiences with observer training. In Chapter 3, I provide a more detail on the qualitative approach I used and why it was selected.

Data were gathered from participants which included middle school, junior high, and high school public school administrators in Grades 7-12 who perform classroom observations and final evaluations and are currently using the 5 Dimensions Plus Teacher Evaluation Rubric. Participants represented rural, urban, and suburban school districts who had a variety of observation experience. By conducting interviews, data were gathered regarding the perceived quality of training received and the effectiveness of training received by administrators. An interview process is an art form that requires an intentional focus on the part of the researcher to promote a trusting and open relationship. I obtained honest and reliable data from the participants, benefiting the overall teacher

evaluation process, and resulting in a positive social change among teachers and observers. Developing a positive rapport with all participants was vital to building a trusting relationship. All participants were assured that identities would remain confidential, along with their school and district name. The interview process involved a virtual face-to-face meeting with each participant at a location that was convenient for the participant. However, face-to-face interviews are the most flexible form of data collection, with advantages including the ability of the interviewer to read the body language of the interviewee and to use follow-up and probing questions in order to motivate respondents (de Leeuw, 2008). Phone interviews were an option if a meeting place and time could not be established or if the participant felt uncomfortable working with the virtual face-to-face platform. None of the participants needed to interview by phone. The time and location of the interviews were decided by the participants to ensure confidentiality and convenience for the participant. Ensuring participants confidentiality allowed participants to feel comfortable to respond to all questions in an honest and trusting environment (Moustakas, 1994). All interviews were recorded using a recording device with the participant's approval.

The participants were a purposeful sample set that intentionally included a group of individuals that can best inform the study about the effectiveness of observer training (see Creswell, 2013). The six constructs presented earlier in this chapter were used as a basis for creating the questions and follow-up questions used during all interviews. Questions regarding how observers currently infuse the observation training to have interactions among teachers about teaching strategies, the capacity of each observer to

conduct observations, observer ability to provide feedback to teachers with expectations and consequences, and most importantly the belief in oneself to complete and implement all phases of teacher evaluation were asked of each participant. Data were collected from school administrators who were currently conducting classroom teacher observations and assigning final evaluation ratings. To start the process, four to five participants were selected from each type of school district: rural, urban, and suburban. Participants who were interviewed had a wide range of teaching and administrative experience. Interviews were conducted with 13 participants who are currently using the 5 Dimensions Plus Teacher Evaluation Rubric. Once the minimum number was met, I continued to interview participants as long as each additional interviewee presented more refined or somewhat different perspectives, thus reaching the saturation point on the quality of observer training (Rubin & Rubin, 2012). Interviews were conducted via a web application called Zoom.

Data were analyzed inductively using applied thematic analysis through several coding cycles using the NVivo12 software. Approaches to assure the trustworthiness of data analysis and interpretations were also applied. Details regarding data collection and analysis are provided in Chapter 3.

Definitions

Evaluation: The term evaluation means a process of gathering and analyzing collected data to reach a final effectiveness rating given to a classroom teacher for their performance throughout the school year. The current rating system for Michigan public

school teachers is a four-tier system of ineffective, minimally effective, effective, and highly effective (PA 102, of 2011).

5 Dimensions Plus Teacher Evaluation Rubric: This teacher evaluation rubric, which was created at the Center for Educational Leadership at the University of Washington, is based on the 5 Dimensions of Teaching and Learning (5D) instructional framework. The core dimensions include purpose, student engagement, curriculum and pedagogy, assessment for student learning, and classroom environment and culture (University of Washington Center for Educational Leadership, 2012). The 5 Dimensions Plus rubric also includes professional collaboration and communication, which is based on activities and relationships that teachers engage in outside of classroom instruction.

Formative observations: Formative observations are meant to provide teachers with feedback on how to improve performance and what types of professional development opportunities will enhance their practice (Mathers et al., 2008).

Notices and wonderings: These two terms are used when observers notice an act or instructional strategy used by teachers while being observed. A notice may be the observer documenting something they noticed during the observation while a wondering is where the observer will document an act or instructional strategy that brings up a question during the observation (University of Washington Center for Educational Leadership, 2012).

Observation: For the purpose of this study, the term observation means the activity of a building administrator physically entering into a classroom, remaining in that classroom for a period of not less than 5 minutes and no longer than a class period,

observing teacher conduct and traditional and nontraditional teaching activities during a class session (see MCL.380(1)(a)).

Observer: Observers must filter a dynamic and unpredictable scene in the classroom to find the most important indicators of performance, make an accurate record of them, and then apply a set of criteria as intended (Archer et al., 2016). For the purpose of this study, the term observer was used to describe a school employee or, in a select few school districts, a designated school employee or contractor who physically visits teacher classrooms for the purpose of conducting classroom observations. In most cases this will be the building principal or assistant principal.

Student growth: For this study, student growth is measured using multiple measures that may include student learning objectives, annual state student assessments, and local assessments chosen by the school district (see MCL380.1249(1)(c)). Student growth is used when determining final evaluation ratings for teachers. As of school year 2018–19, 40% of the teacher evaluation is made up of student growth data (PA 173, 2015).

Summative observations: Sometimes called final evaluations, summative observations are used to make a final decision on factors such as salary, tenure, personnel assignments, retention, and dismissal (Mathers et al., 2008).

Assumptions

Through participating in this study, participants identified areas of observation training which benefited them in determining teacher effectiveness and those areas where more training is needed, or the quality of training needs to improve. As a junior/senior

high school principal with over 20 years of experience, including elementary principal, it was assumed my experiences were very relatable to participants in the study and allowed me to develop a trusting relationship. I assumed that my knowledge of the evaluation terminology was sufficient for this study. II also assumed that because all participants knew they would remain anonymous, they would be transparent in their responses to all questions and follow-up questions.

Michigan legislators and other educational professionals embrace the assumption, that student achievement will increase when students are placed in a classroom with an effective teacher versus being placed in a classroom with a teacher who is minimally effective or ineffective (Weisberg et al., 2009). Placing a student in an effective or highly effective teacher classroom did not address the issue of teacher effectiveness and its relationship with student growth. However, determining teacher effectiveness is the responsibility of the observer and their training, which was the focus of this study. It was important for this study to determine perceptions of participant's quality of training, the ability to differentiate teacher effectiveness, and the willingness to rate teachers ineffective or minimally effective.

Scope and Delimitations

Recent legislation in the state of Michigan has ignited an intentional focus on teacher evaluation and its connection with student growth. In this research study, I focused on evaluator training and its perceived effectiveness for those individuals conducting classroom teacher observations and final evaluations. The methodology of the study incorporates a qualitative research design with interviewing as the data collection

strategy that yielded the most accurate and reliable data. The research data of teacher effectiveness reported currently in the state of Michigan are very similar to the data that were reported in 2009 in *The Widget Effect* where 97% of teachers were rated as satisfactory (Loewus, 2017). Evaluators must be able to differentiate teachers' classroom performance and address those individual teachers in need of support. The focus of this study was to gather data regarding the perceptions of observers regarding training, which was identified as the gap in the literature for individuals conducting classroom teacher observations and final evaluations, cumulating in the analysis of the data to determine if there is a need of additional support and effective training to be successful observers. Teacher performance and their final evaluation ratings and the achievement of their students will not be a part of this research study. The scope of this research study did not include a comparison between teacher ratings and student achievement.

Teacher evaluation is continually evolving and involves many moving parts that make up a successful teacher observation and a successful evaluation. I did not consider the effectiveness of carrying out each element in the teacher observation process in this study, nor did I consider the quantifiable evidence of student growth or student achievement and its effect on final teacher evaluations. In addition, rater reliability of each teacher evaluation was not a focus in this study because of the many variables that comprise rater reliability, such as multiple observers for each final evaluation effectiveness rating. This research study did not address the final evaluation ratings for each participant; therefore, data was not collected on classroom teacher final effectiveness ratings for each interviewee.

The scope of this study was focused only on middle school, junior high school, and high school public education administrators responsible for doing teacher observations. I limited the scope to observer training and its perceived effectiveness, as described by the participants being interviewed who are currently using the 5 Dimensions Plus Teacher Evaluation Rubric. The study was conducted through the lens of SCT, focusing on the training experiences of observers, as well as the external and internal social reinforcements that impact their selection of 5 Dimensions Plus Teacher Evaluation Rubric practices during observations. The participants were active school administrators who conduct observations leading to a final evaluation of teachers in a public-school system in Grades 7-12 in the state of Michigan. There were 13 participants in this study selected from rural, suburban, and urban school districts throughout the state of Michigan. The final number of participants was determined when the data collected and analyzed reached a point of saturation and when no additional themes or information that was extremely different than previously collected was presented. Schools or contracted personnel who conduct classroom teacher observations and final evaluations and use Danielson's evaluation rubric, Marzano's evaluation rubric, or any other professionally created or a locally developed state of Michigan-approved teacher observation rubric were excluded from this study.

There have been several studies conducted surrounding classroom teacher evaluation using a variety of conceptual frameworks as a lens to analyze the data collected. Another potential lens could have been generalizability theory (G theory). G theory is a statistical theory for evaluating the dependability of behavioral measurement.

G theory was used by Semmelroth and Johnson (2014) to measure rater reliability on the Recognizing Effective Special Education Teachers (RESET) observation tool designed to evaluate special education teacher effectiveness. Praetorius et al. (2012) used G theory to focus on the causes and amounts of rater bias for a study. Again, like Semmelroth and Johnson, I analyzed ratings while using multiple sources of error. The Institute of Research at the University of Michigan (2013) used G theory to examine how different errors in measurement affect measurement reliability. Donaldson (2013) used G theory to analyze scoring of ratings as well. Kraft and Gilmour (2017) used Lipsky's street-level theory in the public sector, which states employees cannot do the job according to the ideal conceptions of the practice because of limiting work structure. G theory and Lipsky's street-level theory do not fit for this study based on the type of data collected and the proposed analysis process. SCT and its constructs align with this qualitative research study, enabling researchers to attempt to gain insight into a universal feeling or experience of observers and to interview participants and collect data regarding the perception of observers.

Limitations

A qualitative research design using the interviewing process for gathering data encompasses some limitations. To realize the potential of a face-to-face interview, a well-tested questionnaire and a well-trained interviewer are needed to lend credibility to the study (de Leeuw, 2008). Currently, there are over 150 school districts in the state of Michigan that are using the 5 Dimensions Plus Teacher Evaluation Rubric. Selecting participants that represented the demographics was a limitation for this study and was

addressed by purposefully selecting a participant pool that fits the criteria established ensuring dependability. Dependability refers to whether the research process is consistent and can be carried out with careful attention to the rules and conventions of qualitative methodology (Guest et al., 2012) Dependability is evident whether other researchers could repeat the study and would arrive at the same findings with the same data. Selection of participants came from an extensive administrator network I have developed over the last 18 years. I am a member of the Michigan Association of Secondary School Principals (MASSP) and used my professional contacts with this organization to access participants. However, having an awareness of existing contacts could introduce bias in the selection of participants, so I ensured rural, urban, and suburban schools were represented along with observers with varied experience in observing teachers. Replicate studies could take place, enhancing transferability by selecting a similar participant pool using a different observation/evaluation rubric. Participants did not have to be members of the MASSP to participate in this research study.

When the interviewer has limited experience in the interview procedure itself, the following activities could be a challenge: using recording devices, knowing the number of questions to ask, or deciding when to dig deeper with follow-up and probing questions (de Leeuw, 2008). Interviewers might also not know how to present the research purpose to the interviewee or how to deal with responses from the interviewees that differ from their actual practice when conducting classroom teacher observations or final evaluation ratings. To address these limitations, I conducted two field test practice interviews with observers using the 5 Dimensions Plus Teacher Evaluation Rubric to rehearse the

interview protocol and make any adjustments necessary. Focusing questions on experience and behaviors, opinions and values, knowledge, and background/demographic questions can be limiting. All information gathered was the most relevant and direct information obtainable.

Observing and evaluating teachers entails many variables that need to be accounted for when determining an effectiveness rating. School culture and climate, student demographics, school leadership, and staff experience all play a significant role in teacher observations; therefore, another limitation in the data collection process was participant and researcher bias. As a principal with over 18 years of experience, I have some tendencies and work procedures that I was consciously aware of during the interview process and used the interview protocol to control these potential biases. Participant bias could occur in how they view the effectiveness of their use of the 5 Dimensions Plus Teacher Evaluation Rubric and their comfort level with the evaluation process. Researcher bias can be a limiting factor during the interview process and data analysis. As a building leader with experience at all levels, I possess a level of bias regarding teacher evaluation. The training I have received and the knowledge I possess regarding the evaluation process has its bias. How I perceive the teacher evaluation process and how the process should be implemented comes from a single perspective based on the training I received. Some observers may have a narrow experience observing teachers based on the number of teachers they evaluate because working in a rural district that has fewer teachers may present a limitation towards teacher observation. The biases I brought to the study were used to inform and enhance the interview by

gathering all important relevant data regarding the observation and evaluation process currently conducted by the observer and to document their experiences accurately. I used a value-neutral approach, which was very important in the social environment because values, emotions, and personal experiences can set a new path, hindering accurate observation and discourse. From a researcher point of view, putting personal biases aside was a requirement in this study. To address and assure objectivity, I had a critical friend/colleague review all findings from this research study. During the interview I was journaling and making notes regarding decisions made and answers given for each interview question. Dealing with participant bias as a limiting factor was addressed by developing a positive, trusting rapport with the interviewee. To achieve this rapport, the interviewee must believe that their confidentiality and anonymity are guaranteed. Also, when participant answers sound inconsistent, which may indicate underlying bias or issues that need to be explored, carefully crafted follow-up questions were used to reveal any bias that may exist in the participant's answers (Rubin & Rubin, 2012). Rubin and Rubin (2012) suggested using four questions which request interviewees to act as guides walking the interviewer through their work while pointing out what they think is important in their work. Documents disclosing the details of the interview process, the use of data, and the use of participant information were reviewed and signed by all participants.

Significance

Following *The Widget Effect* (Weisberg et al., 2009), not much had changed in the final ratings assigned to all teachers as of the 2017–18 school year. This report called

attention to teacher evaluation and the ability of the observer to distinguish different levels of classroom teacher effectiveness. According to Michigan data for the 2017–18 school year, most teachers were still rated in the effective or highly effective categories in their final evaluation. There is a problem in the education system across the country regarding the ability of observers to identify or report the true levels of effectiveness of teachers when conducting classroom teacher observations and final evaluations (Kraft & Gilmour, 2017; Norman, 2010; Steinberg & Sartain, 2015; Weisberg et al., 2009). This problem stems from the effectiveness of the training and the levels of training received by those individuals conducting classroom teacher observations and final evaluations. According to the National Council on Teacher Quality report by Doherty and Jacobs (2015), there is a pattern emerging across states showing most teachers, over 97%, are being identified as effective or highly effective. While teacher performance continues to be highly rated, student academic performance in the state of Michigan continues to fall compared to other states. The significance of this problem is that while teacher evaluation ratings stay the same, which is at an effective or highly effective level, the performance level of students is dropping in the state of Michigan. Gathering data surrounding training practices that observers perceive to be the most effective and training practices that they perceive to be not as effective in identifying levels of effective teaching was the primary focus of this study.

Teacher evaluation and the classroom observation process have evolved into a much more detailed rating system, requiring increased time and effort by observers. Cosner et al. (2015) noted that principals are the key to the evaluation process and require

extensive systemic training to develop the necessary skills and knowledge to fully understand and implement teacher evaluation practices with fidelity. It was evident there was a need to learn more about how principals make observation judgments about teachers' performance, which was the focus of a study conducted by Kimball and Milanowski (2009) confirming that there is a need for more research on the topic of quality observer training.

The training received by potential observers must meet the needs of observers, be systematic, and be high quality. Personnel decisions, such as promotion and dismissal, are based upon teacher ratings; therefore, it is important to ensure that observers are trained and prepared to effectively rate the effectiveness of teachers. The implementation process for the 5 Dimensions Plus Teacher Evaluation Rubric is extensive and requires specific rubric knowledge, observation, and final evaluation rating skill capacity; self-management; and follow-up with documentation and discussion. Fully implementing the rubric requires a time commitment that is significantly greater than previously used rubrics with a binary rating system. Typically, training with the 5 Dimensions Plus Teacher Evaluation Rubric lasts 3 days, covering each dimension and the dimension indicators.

It is vital for observers to have the capacity to differentiate teacher effectiveness and become proficient in classroom teacher observation by attending professional development sessions provided by a certified, designated person trained by personnel from the Center for Educational Leadership at the University of Washington. Evaluator training is important and necessary if the level of classroom instruction is expected to

increase and teachers who are in need of extra support are to be identified. Effective teachers have an academic impact on students along with a social impact on the school, school district, and the community by increasing the social, educational, and economic opportunities for each student who graduates. Observers can create a community of colleagues via zoom and conduct mini professional development sessions focused on specific areas of the 5 Dimensions Plus Teacher Evaluation Rubric. This was a specific response by one of the participants and has interest among other participants. Creating these sessions for observers presents an opportunity to build relationships among colleagues creating an open and positive social change among observers.

The results of this study may bring awareness the perceived training effectiveness of observers to principals and training providers. Careful data analysis and discovering patterns and themes can potentially lead to a change in training practices, bringing about positive social change. The contributions of this study brought to the forefront the overall perceptions of the effectiveness of the training received by building principals and vice principals at the junior high and high school levels. Working collaboratively with colleagues surrounding teacher evaluation strategies were evident in the findings. The results of this study can be used to enhance future development of training practices offered by the Center for Educational Leadership or any approved contracted educational organization, such as the MASSP, which currently provides the 5 Dimensions Plus Teacher Evaluation Rubric training for those conducting classroom teacher observations in Michigan. Michigan lawmakers may consider the current training requirements and

realize additional training may be needed to improve observation practices, resulting in a change in current law based on the results of this study.

Summary

Classroom teacher observation in the state of Michigan has had some significant changes over the past decade. New state and federal legislation has been passed and put into law, including the Every Student Succeeds Act of 2015 (ESSA). This legislation has affected teacher evaluation by placing it as a high priority topic for school districts in the state of Michigan.

The purpose of this study was to study the training strategies used for classroom observers and the perceived outcomes of those strategies to conduct effective observations for those who are currently conducting classroom teacher observations and final evaluations and are using the 5 Dimensions Plus Teacher Evaluation Rubric. I used a qualitative research approach to analyze the data gathered through an interview process. The data was analyzed and used to report the amounts and quality of training received by participants and their self-proclaimed preparedness to implement the evaluation process as it is recommended by the Center of Educational Leadership at the University of Washington (University of Washington Center for Educational Leadership, 2012).

The literature review for this research study identified a gap in the area of training and preparedness of potential classroom teacher observers who use the 5 Dimensions Plus Teacher Evaluation Rubric in Michigan school districts. Since the publication of *The Widget Effect* (Weisberg et al., 2009), a closer focus on teacher evaluation has taken place in the state of Michigan and numerous other states across the United States. Other

publications with significant data regarding teacher observation include the Measures of Effective Teaching project (Weisberg et al., 2009) founded by the Bill and Melinda Gates Foundation, which emphasized the lack of focus on the effectiveness of teachers. Despite recent legislation, state publications such as (Michigan Department of Education, n.d., 2013–14) in Michigan pointed out that more than 97% of teachers were rated highly effective or effective, which means that while the evaluation process has changed, the teacher rating results have remained relatively the same in Michigan while student achievement continues to decline.

Two research questions were developed to serve as the basis for my interview questions on the perceived quality of training available to those individuals who conduct classroom teacher observations and final evaluations and are using the 5 Dimensions Plus Teacher Evaluation Rubric. The results of this study identified significant challenges in implementing the observation and evaluation process with fidelity, as well as potential best practices in observer training. Interview questions produced data that can be used for modifying and redesigning training. By modifying training practices, challenges can be overcome for all involved in conducting classroom teacher observations and final evaluations. When researchers attempt to gain insight into a universal feeling or experience among a group of individuals, often a qualitative approach is used. This type of approach also lends itself to a smaller sample size, which then results in long, in-depth interviews with participants providing rich data. Although a smaller sample size can be a limitation, building rapport with participants assists in the interview process. However, interviewing itself can be limiting in that participants are not being articulate about their

answers or sometimes the interpretation of the responses of the participants may not be what was meant. These limitations and others are addressed in more detail in Chapter 3.

In Chapter 2 of this dissertation, I reviewed an extensive list of relevant and recent literature such as journal articles, books, dissertations, and reports that show the legislation enacted surrounding teacher evaluation, the current status of teacher evaluation and observation ratings, and the framework for quality observer training needed for effective classroom teacher observation and final evaluation. The following literature review provides (a) a historical perspective of the legislated reforms, (b) promising practices that observers have used to improve teacher effectiveness, and (c) information about Michigan's teacher evaluation components. Chapter 2 also contains subsections on other studies conducted surrounding teacher evaluation effectiveness, SCT, and how other researchers used SCT in their evaluations on similar topics.

Chapter 2: Literature Review

There is a problem across the country as well in Michigan's education system regarding the ability of observers to identify and report true levels of teacher effectiveness when conducting classroom observations and final evaluations (Kraft & Gilmour, 2017; Norman, 2010; Steinberg & Sartain, 2015; Weisberg et al., 2009). This problem is compounded by a lack of agreement on how best to identify and measure effective teaching (Kane et al., 2011b). Moreover, the severity of this problem is affected by the quality of training received by classroom observers (Doherty & Jacobs, 2015; Michigan Council for Educator Effectiveness, 2013). This represents a significant area of need in Michigan schools, as improving classroom observations by focusing on effective training may result in greater accuracy of teacher evaluations, leading to an overall increase in teacher effectiveness. Accurate teacher evaluations also identify and prescribe professional development needs (Doherty & Jacobs, 2013), as well as identify areas of needed ongoing coaching, support, and feedback, making the accuracy of teacher evaluations a critical issue within the field of education today.

Teacher evaluation systems were originally designed to differentiate teachers who improve student learning from those who do not (Steinberg & Sartain, 2015). Currently, there is a discrepancy between teacher effectiveness ratings and student performance in Michigan schools. Through the 1980s, Michigan's educational system was ranked in the top half of the nation, according to the National Assessment of Educational Progress (National Assessment of Educational Progress, 2009). Since then, its rankings have declined steadily. Yet, as recently as the 2013–14 school year, 97% of all public teachers

in Michigan had an evaluation rating of “effective” or “highly effective.” Michigan teachers continue to earn superior ratings, while Michigan students fall behind academically compared to other states, according to CEPI (Michigan Department of Education, 2017).

The discrepancy in teacher evaluation ratings and student performance found in Michigan reflects a broad pattern emerging across the United States wherein most teachers are identified as effective or highly effective (Doherty & Jacobs, 2015). In a study of Chicago Public Schools, it was found that nearly 93% of teachers received ratings of “superior” or “excellent” within their four-tier teacher rating system (Steinberg & Sartain, 2015). At the same time, Chicago remained the lowest performing district with two thirds of the schools not making state proficiency standards (Steinberg & Sartain, 2015). These results are like those reported in *The Widget Effect* (Weisberg et al., 2009) study, wherein the same percentage of effective and highly effective teachers were found. This study surveyed over 1,300 school administrators and over 15,000 active teachers over a 2-year period using an online survey strategy, with findings that resulted in a call to action to address our national failure to recognize indifferences in teacher effectiveness once and for all (Weisberg et al., 2009).

Using current teacher evaluation data, I explored the formal training of observers and the quality and effectiveness of training that is made available to administrators in Michigan schools. The purpose of this qualitative study was to gather data regarding perceptions of the types of training available to classroom observers and evaluators, as well as perceptions regarding the effectiveness of this training. In this study, I gathered

information specific to the skill set and practices school administrators use to implement their evaluation processes and narrowed the focus to school administrators who are currently using the 5 Dimensions Plus Teacher Evaluation Rubric for a Michigan public school system that is actively observing, evaluating, and rating junior/senior high teachers.

This chapter provides a review of the research and literature related to the purpose of this study. The review begins with a brief description of the search strategy and the key terms and phrases used to obtain the literature contained herein. Next, I describe the theoretical framework used to organize this study: SCT, its previous uses among researchers, and my rationale for selecting SCT as the theoretical framework for this qualitative research study.

After discussing the theoretical framework of this study, I then present key variables addressing teacher evaluation legislation, teacher evaluation and performance ratings, observer proficiency, and criteria for differentiating effective teachers. In the final sections of the literature review, I present a summary of themes currently trending in teacher evaluation and observer training in the state of Michigan and discussed how the identified gap in the literature was addressed in this qualitative research study.

Literature Search Strategy

The overall literature search strategy for this research began broadly, using terms that would capture a large number of sources applicable to classroom teacher observer training. Narrowing the search was accomplished with search terms that were more specific to the topic of observer training. Search terms included *teacher evaluation,*

*evaluator training, observer training, teacher observation, classroom teacher observer training, evaluation training, and administrator training, evaluation, observation,*and *training.* Overall, the search process for this literature review was extensive and I continually revised it to capture the most current literature. A detailed explanation of the search process and the terms used to conduct this literature review is provided in the paragraphs below.

Using Google Chrome as my search engine, I accessed the Walden University library for most of the search process. From Walden's library, using Thoreau multi database search, I was able to access several education-related databases that included EBSCO, Education Source, ERIC, and Education Source Combined Search. Related subject databases such as PsycINFO and SocINDEX were also used, but with little success using terms including evaluation, education, and evaluation or observer training. I also sought out educational journals to gather literature on classroom teacher evaluation or classroom evaluator training. Although numerous articles pertaining to classroom teacher evaluation were found, none of the articles were peer-reviewed and none were used in this study. As an alternative to finding limited resources related to evaluator training, I proceeded to search for other sources related to the search topic such as reports, white papers, and books as sources of information for the literature review. Also, to ensure the academic rigor of the literature search, only articles in full text that have been peer-reviewed were used in this study. Reports that were considered and used for this literature review were also in full text and peer reviewed. Finally, I also used the

multidisciplinary database and the SAGE Journals, the ProQuest Central database, and Academic Search Complete to identify resources.

The articles used in this research study met the following criteria. Each article was peer-reviewed, only articles that were the basis of primary research were used for this dissertation, and articles that have been published within the last 5 years from the inception of this study. The total number of articles collected for this study was 128. However, not all articles were used in this study.

Theoretical Framework

Since The New Teacher Project's publication of *The Widget Effect* (Weisberg et al., 2009), much has changed in the way observers approach teacher effectiveness including how classroom observations are conducted, the way observers determine final evaluation ratings, and how teachers are retained, promoted, compensated, or fired. Specific observer practices that have changed in the classroom teacher evaluation process include the evaluation instrument, additional classroom visits, and the nature of conversations regarding teacher performance in consideration of mandates by Michigan legislation. Approaching the change for improvement in teacher observation and the ratings of teachers can be viewed through the lens of SCT, which was originally developed as social learning theory in the 1960s by Bandura. SCT affords a clear understanding of the cognitive skills needed by an observer to effectively observe their teachers. The unique feature of SCT, which lent itself to this study, was that SCT has an emphasis on social influence and its impact on the external and internal social environment (see LaMorte, 2018). Self-efficacy is a central aspect of the SCT's

emphasis. Observer self-efficacy is vital to the implementation of an effective evaluation system and its demands to implement specific strategies regarding classroom observation. Therefore, the research questions that were used in this study directly relate to self-efficacy and the identification of observation practices observers feel confident implementing during classroom observations, as well as those practices observers are not confident in implementing and those for which they believe they need additional training.

The overall purpose of SCT explains how people can maintain and regulate their behavior through control and reinforcement that can be sustained over time (LaMorte, 2018). SCT embraces the notion that through forethought and self-reflection, people can substantially influence their own outcomes and the environments in which they live and work (James, 2014). Individuals learn by observing mentors whose behavior they choose to emulate, all of which is mediated by their self-efficacy beliefs, which is generally defined as one's belief in their ability to succeed in a specific task or situation (James, 2014). There are six constructs that make up SCT, according to LaMorte (2018): reciprocal determinism, behavioral capability, observational learning, reinforcements, expectations, and self-efficacy.

James (2014) found that self-efficacy and goal setting drive SCT because if an individual does not believe they can be successful at a particular task, a positive outcome only occurs in rare instances. A research study conducted by Pisciotta (2014) identified the relationship between teacher self-efficacy as it relates to job performance, in this case, teacher observation, student achievement, and teacher perceptions on the evaluation process. Pisciotta found that when observers lack self-efficacy and are not implementing

quality performance observations, the lack of feedback on the observation may lead to low teacher self-efficacy, affecting their ability to fulfill the requirements by the state of Michigan. Perceptions of self-efficacy influence one's choices and self-beliefs, including the goals he or she chooses to pursue and the effort he or she puts into the goals (Vinney, 2019). These beliefs can impact personal growth and social change. Belief in one's self-efficacy can be the difference between whether an individual even considers making a positive change in their lives and in the lives of others (Vinney, 2019). Having strong self-efficacy and a positive attitude of improvement, observers may be more prepared when observing teachers, thereby impacting teacher effectiveness by increasing their own effectiveness.

SCT provided the lens for me to analyze observer responses obtained during the interview process regarding classroom teacher observation instrumentation, training, and goal setting, while focusing on observers their own self-efficacy to improve teaching effectiveness and increase student achievement. In a study conducted by Bell et al. (2014) of 42 administrators in the Los Angeles Unified School District, it was determined that in order to effectively train observers, it is critical to understand two of the major tasks of classroom observation: creating accurate rating scores and formulating conversations around those ratings that support instructional improvement. Bell et al. used the lens of SCT to improve observer training and teacher ratings, adding to the validity of the use of SCT for the current study.

Another major component of SCT is observational learning. This is the process of learning desirable and undesirable behaviors by observing others, then reproducing

learned behaviors in order to maximize results (Vinney, 2019). This observational learning occurs through a sequence of four processes: (a) attention processes, (b) retention processes, (c) production processes, and (d) motivational processes (Vinney, 2019). A basic premise of SCT is that people learn not only through their own experiences, but also by observing the actions of others and the results of those actions (Vinney, 2019). By working with other observers and collaborating with them regarding training videos, training practices, and training techniques, the observation of other observers at work may lead to positive learned behaviors.

Current final teacher evaluation ratings in Michigan indicate that teachers are performing at a high level overall. However, the data on student achievement does not match this research. The quality and level of training received by individuals conducting classroom observations and final evaluation ratings are the factors that were looked at in this qualitative study. In this qualitative research study, I explored the quality of the training and preparedness of individuals conducting classroom teacher observations and determining final evaluation ratings. Using SCT as the lens for this study, like Bell et al.'s (2014) mixed-method study regarding observer training, this qualitative study interpreted the data obtained using the six constructs of SCT.

Pisciotta (2014) used SCT to analyze teachers' responses to the teacher evaluation instrument as it relates to their own self-efficacy towards improving teaching effectiveness and student achievement. In another study using the SCT perspective, Bell et al. (2014) viewed the scoring of observations of classroom interactions to be a complex socio-cognitive process that must be understood in order to improve observer training,

and ultimately, score quality. The study found that in order to effectively train observers, it is critical to understand how observers learn to complete two major tasks when observing; learning to score accurately and learning to have conversations around the scores to support instructional improvement (Bell et al., 2014). Bell et al. found that observers were not only thinking about creating scores based on the observation; they often thought about how they were going to have the post observation feedback session or how they would improve the teachers' practice while observing them. Walker and Posner (2003) used SCT in their study to test and determine if self-efficacy is used to influence people's behavior intentions directly and indirectly through effects on outcome expectancy. Data on self-efficacy and outcome expectancy were collected from 115 college students to jog two consecutive miles. It was found the more efficacious people were, the more positive the outcomes they associated with an activity and the surer they were they would perform the particular activity (Walker & Posner, 2003).

Current final teacher evaluation ratings indicate that teachers are performing at a high level. However, in the state of Michigan, the data on student achievement does not support these high teacher performance ratings. The quality and level of training received by individuals conducting classroom observations and final evaluation ratings were the factors looked at in this qualitative study. I used SCT as the lens for this study, similar to Bell et al. (2014) in their mixed-methods study regarding observer training. Using SCT as the framework for this study created a path that focuses on positive social influence and its emphasis on external and internal positive social reinforcements. Using the six key constructs of SCT to identify the current perceptions of observers, new training strategies

were identified and used to customize future training, moving closer to meeting the needs of all observers. Data gathered from participants surrounding specific observation practices that are embraced and implemented and those practices that are not implemented in combination with observation practices that will enhance the effectiveness of the training through the lens of SCT will benefit observers using the 5 Dimensions Plus Teacher Evaluation Rubric.

Literature Review Key Variables and/or Concepts

In this section, I address concepts related to the purpose of this study, its research questions, and the methodology and strategies used to collect appropriate and relevant data. Based on the literature, previous studies and their approaches to the teacher evaluation topic, and the theoretical frameworks used as a lens to view the data, I gathered data about what was known to advance this research study focusing on perceptions of the quality of training of the 5 Dimensions Plus Teacher Evaluation Rubric in the state of Michigan. There are two research questions that provided the foundation for this study and focus on training of observers and their practices. These research questions are as follows:

RQ1. What are the perceptions of school administrators who are responsible for evaluating teachers about the additional training they need for performing effective classroom observations?

RQ2. What are the perceptions of school administrators who are responsible for evaluating teachers about the training they have received for performing classroom observations?

Legislation and Teacher Evaluations

In this subsection of the literature review, I present the background of Michigan's teacher evaluation process, as it became the focus in education in 2011 with the passing of PA 100-103 and subsequent Michigan legislation. To this end, I review the reforms that have been legislated regarding teacher evaluation and the changes that have taken place regarding classroom teacher evaluation.

Early legislation in the field of education started with the federal Elementary and Secondary Education Act (ESEA) of 1965 (White, 2018). However, this law did not specifically address teacher evaluation. Michigan's Public Act (PA) 25 of 1990 forced a significant shift of educational focus by establishing an accountability system for public schools through a school improvement process to include parent and community involvement (Michigan Legislature, n.d.). This legislation, however, also failed to specifically address teacher evaluation. Ten years after the passage of PA 25, the No Child Left Behind Act of 2001 replaced the ESEA, which was then followed by the Race to the Top competition in 2009, and the Elementary and Secondary Education Act Flexibility program of 2011 (Michigan Legislature, n.d.). These initiatives emphasized state development of systems tying teacher performance evaluation and tenure decisions to student achievement (White, 2018).

Since September 2011, Michigan teacher evaluation laws have mandated a change in the way school districts and school building administrators evaluate and rate teachers as stated in the Michigan Revised School Code Act 451 (1976). PA 102 of 2011 provided a statewide system of educator evaluation that would have an extensive impact

on all public-school teachers (Michigan Department of Education, 2019). This law also established the MCEE as a temporary state commission to advise the governor, state board of education, and state legislature on a number of issues relating to the implementation of PA 102. Aside from Michigan legislation, federal legislation was also on the forefront of teacher improvement and evaluation. The No Child Left Behind Act (NCLB) was reauthorized nine years later with the passage of the ESSA of 2015.

In 2013, a pilot study performed by the Institute for Social Research at the University of Michigan found that most schools in the study lacked a fully developed teacher observation system, or well-structured policies and procedures and detailed documents to guide the process directed by PA 102 (Rowna, et al., 2013). As of 2015, PA 173 provides language for the state of Michigan to now provide training for observers and teachers about the teacher evaluation rubric used in their individual school districts, and to assist in learning, identifying, and assessing teaching practices and strategies (White, 2018).

Prior to ESSA's passage, Michigan legislation surrounding education placed an emphasis on teacher performance and teacher effectiveness through an elaborate teacher evaluation system (White, 2018). Since ESSA's (2015) passage, teacher evaluation has returned the authority of educational issues and programs back to the states, rendering current teacher evaluation less restrictive than previous legislation. Currently, 42 states and the District of Columbia now have teacher evaluation policies written into state law or regulations (White, 2018).

Teacher Evaluation and Performance Ratings

The Widget Effect (Weisberg et al., 2009), a report that examined teacher effectiveness, brought attention to the failure of school districts across the United States to recognize and respond to the differences in teacher effectiveness, specifically those teachers who are not effective. The qualitative study surveyed over 1,300 school administrators and over 15,000 teachers, resulting in a wide-ranging report that studies teacher evaluation and teacher dismissal in four states and 12 diverse school districts. The title refers to the practice of school district observers to assume teacher effectiveness is the same for most teachers. When the study was conducted, researchers found that in a binary system of rating such as satisfactory and unsatisfactory, 99% of teachers were rated satisfactory, professional development was inadequate for observers, and nearly 75% of teachers did not receive any specific feedback on improving their performance in their last observation. Poorly performing teachers were going unaddressed, along with new teachers being neglected during the observation process. Moreover, when multiple rating categories were used, less than one percent of teachers were left out of the top two categories, which were effective and highly effective. Overall, Weisberg et al. (2009) recognized teachers were not receiving the proper feedback for improvement, development, and support from those who were performing classroom observations. This report recommended that administrators receive rigorous training followed with ongoing support, conduct consistent observations, make fair assessments of classroom teacher performance, and provide constructive feedback along with differentiated support to all teachers.

In another study, researchers examined school administrators' perceptions of how well the Iowa Evaluator Approval Training Program (IAETP) made them feel prepared to conduct teacher evaluations using the Iowa Teaching Standards and Criteria (Lasswell et al., 2008). In this study, the researchers developed a survey instrument featuring short answer and Likert scale responses, which was administered to a randomly selected group of Iowa public school administrators. Over 65% of administrators who responded reported that the training provided by the IAETP had adequately prepared them for conducting teacher evaluations; however, this shows the inconsistency in training provided to observers in education, as surveys from Iowa research found there are differences in the way teacher evaluations are conducted. This finding was consistent with research conducted by Archer et al. (2016), which suggests principals are likely to need extensive training and ongoing support to enact teacher evaluation practices. Overall, classroom observation today may be better than in the past, when it was based on simple checklists, but the quality of implementation clearly remains uneven. The first attempts at observation training probably won't produce the accurate and meaningful feedback that can be expect from later attempts. Observation is too challenging to expect the first attempt to be fully successful.

A review of studies conducted regarding the effectiveness ratings of teachers shows little has changed in their performance ratings since the passage of federal and state legislation pertaining to teacher evaluation. Never in the history of education has teacher evaluation had such state and national attention. Recent teacher evaluation legislation brings a high stakes policy, which could possibly mean the dismissal of

teachers from the profession in a relatively easier manner than in the past no matter how long the teacher has been teaching. Since the publication of a federal white paper in 1983 called *A Nation at Risk,* a focus on public education and its effectiveness has risen. A publication titled *The Blueprint for Reform*, published by the U.S. Department of Education in 2010, focused on teacher evaluation as a means to improve teacher quality. It was then followed by the Obama administration offering to fund schools through the Race to the Top initiative's school improvement grants. Waivers were offered as relief from the requirements for Race to the Top funding if states used test scores in teacher evaluations to judge teacher quality (Hazi, 2014). Prior to this, teacher evaluations were mainly focused on final or summative evaluations, and most policies did not include any requirements for establishing teacher performance standards and evaluator training (Mathers et al., 2008).

MCEE was formed in 2011 to study school reform strategies, including teacher evaluation strategies that have been at the forefront of politicians' efforts to improve public schools (Michigan Council for Educator Effectiveness, 2013). In its process of gathering input from stakeholders, the commission collected papers commissioned from experts on a variety of educational issues and heard testimony from administrators, teachers, students, representatives of professional and public groups, parents, business leaders, public officials, and scholars. All groups were represented and testified at eight meetings in front of the whole commission, as well as at six public hearings, two panel discussions, a symposium, and a series of meetings organized by the Department of Education. This extensive data gathering process took over two years to produce a final

report. In the report it was found that observer feedback is vital and that targeted professional development opportunities should be made available to teachers (Michigan Council for Educator Effectiveness, 2013).

The purpose of teacher evaluations is to determine classroom teacher effectiveness while ensuring that opportunities for improvement are provided through coaching, professional development, and support (Bell et al., 2014; Michigan Department of Education, 2016). Bell and colleagues compiled data from a handful of large-scale, nationwide studies, interviews of 42 focus observers making up a large sample of 998 administrators found there is much to be learned about how to accurately score observations. Observers were shown 10-minute videos, with results indicating that the master raters used the observation rubric more consistently to score lessons compared to the trained raters, who more often used strategies other than the rubric, including reasoning from memorable training, calibration videos, or use of their own internal criteria. The results of my study support the idea that ongoing training in classroom observation is necessary for accurate teacher evaluation.

Classroom Observer Proficiency

Effective teachers have a direct impact on overall student achievement (Kane et al., 2011a; Kraft & Gilmour, 2017; Lasswell et al., 2008; Michigan Council for Educator Effectiveness, 2013; Putman et al., 2018; Steinberg & Sartain, 2015). Supporting effective teaching with professional development requires an effective observer. Researchers posit that effective observers will lead to an increase in teacher effectiveness and the ability to differentiate and identify teachers in need of support and to prescribe

the support needed. Observation results often are the only window a school system has into the state of the teaching in its classrooms. Without accuracy, it is impossible to know what support teachers need or if those supports are working (Archer et al., 2016).

In a study conducted by Kane et al. (2011a), the Cincinnati Public Schools' Teacher Evaluation System (TES) used peer evaluators selected based on their own TES score to serve as full-time observers for 3 years before returning back to the classroom. In TES, both peer and administrator observers complete an intensive training course and must accurately score videotaped teaching examples to check interrater reliability (Kane et al., 2011a). As mentioned earlier it is challenging for observers to differentiate and identify effective teachers and teaching practices. Kane et al. (2011a) found the teachers' classroom practices as measured by TES scores do predict differences in student achievement.

According to Gargani and Strong (2014), changes need to be made to the current teacher evaluation system. Multiple reports suggest not only are the ratings of teachers inflated, but observers are also asked to do too much, making the observation procedures too burdensome while leaving too little focus on feedback. Gargani and Strong (2014) conducted experimental studies of the Rapid Assessment of Teacher Effectiveness and found that by using shorter segments of instruction, fewer observations, less training, and a smaller number of simpler scoring criteria than in previously evaluated instruments, they were able to generate scores that were more consistent, more reliable, predictive, and less expensive. In other studies, it was also discovered that principals were struggling to complete the necessary number of observations in a given school year (Donaldson et al.,

2014), and that fidelity was a question in the implementation process due to time (Derrington, 2014).

Understanding of the evaluation rubrics used by observers to conduct classroom observations is another factor affecting teacher evaluation. Cosner et al. (2015) conducted case studies in three states, exploring two questions: (1) Are principals able to develop the necessary conceptual understandings to support the robust evaluation practices required? (2) Are they also able to develop productive approaches for addressing the new work demands generated by the new evaluation systems? Findings suggested that principals in each state found that administrators realized increased workloads and time demands with the new system, suggesting that principals are likely to struggle to complete every element of the system and may have challenges implementing all the required observations (Cosner et al., 2015). Donaldson et al. (2014) also found that principals were struggling to meet the required number of observations throughout the school year, averaging four observations per year and not meeting the recommended six observations per teacher.

The studies discussed in this literature review thus far have used either qualitative or mixed methods. A variety of strategies were used for gathering data including interviews, surveys, and the use of a Likert scale for responses. Reports and dissertations were also used in this literature review that combined the use of case studies, interviews, and surveys. Mixed methods studies used data such as classroom teacher evaluation ratings, the percentage of teachers rated in each category, and teacher evaluation ratings including student growth numbers (Bell et al., 2014; Dodson, 2015b; Donaldson, 2013;

Kimball & Milanowski, 2009; Kraft & Gilmour, 2017; Loewus, 2017; Michigan Department of Education, n.d.; Norris, et al., 2017). To date, there exists little evidence about the degree to which teacher evaluation reforms have fundamentally changed the distribution of teacher performance ratings (Kraft & Gilmour, 2017). Currently, research suggests that the percentage of teachers rated as unsatisfactory has not changed in the majority of states that have adopted new teacher evaluation systems (Kraft & Gilmour, 2017). At the same time, there is considerable variation across states in the percentage of teachers rated in the category just below proficient, as well as above (Kraft & Gilmour, 2017).

Identified Effective Observer Skills

The key to getting the most out of teacher evaluation is figuring out how to implement it in a way that challenges, supports, and motivates teachers (Donaldson, 2016). Donaldson also noted that teacher evaluation requires robust evaluation instruments, skilled and conscientious district and school leaders, and teachers who are willing to take risks, self-evaluate, and learn. Donaldson (2013) stated that effective teacher evaluation requires endurance on the part of the observer. The success evaluation systems depend on the will, skill, and capacity of school principals who have historically been tasked with evaluating teachers (Cosner et al., 2015).

According to Praetorius et al. (2012), in order for the quality of instruction to improve, it is important for those individuals conducting classroom teacher observations and final evaluations to identify the strengths and weaknesses of teachers concerning their teaching performance. Strengths and weaknesses are generally identified by the

rating and feedback teachers receive from observations or final evaluations. If these ratings are in question because of rater bias or experience, the intent to improve a teacher's instructional delivery may focus on the wrong area or not focus on any area at all. Having an honest and transparent observer–teacher relationship is critical to the ongoing professional development and improvement of each teacher even if the feedback is sometimes critical (Praetorius et al., 2012).

On average, the observers who participated in a survey conducted by Kraft and Gilmour (2017) estimated that 27% of all teachers in their schools were performing at a level below proficient. Moreover, 81% of administrators and 57% of teachers could identify a teacher in their school who was ineffective. This estimate is more than four times the percentage of teachers who were rated below proficient (Kraft & Gilmour, 2017). Loewus (2017) found similar results in a survey of 200 principals in a large urban district in the Northeast. The study found evaluators identified far more teachers as weak in a confidential survey than they did on the formal district evaluations. During interviews with principals, findings suggested that the large differences in the distributions of formative versus summative ratings are primarily the result of higher stakes attached to the summative ratings (Loewus, 2017). A cause of these higher rankings for teachers is the need for positive relationships with their staff, according to Loewus (2017). There are many other possibilities surrounding the classroom teacher observation process from the observer perspective that can lead to inflated final teacher evaluation ratings. When principals are asked their opinions of teachers in confidence and with no stakes attached, they are much more likely to give harsh ratings (Loewus, 2017).

Additionally, upon rating a teacher with a minimally effective or ineffective rating, intensive amounts of time are required to document their performance and to provide support for their professional growth. Similarly, Kraft & Gilmour (2017) found principals questioned the time to collect evidence in a few observations to mark a teacher below proficient. When observing and rating new teachers, principals in their study believed new teachers were still learning and it was unfair to rate new teachers as below proficient if they were working to improve. Kraft & Gilmour (2017) also found that principals believed giving a low rating to a potentially good teacher would be counterproductive to the teacher's development. As a result, these researchers found that even teachers who are truly inadequate in their practice may be given scores on their observations that do not reflect their true instructional capacities. Strunk et al. (2014) believed these inflated ratings may reflect a form of rater leniency as well as raters not being trained how to differentiate between inadequate and average levels of practice.

Classroom Observer Limitations and Barriers to Effective Evaluation

Historically, teacher evaluations were mainly focused on final or summative evaluations, and most policies did not include any requirements for establishing teacher performance standards and evaluator training (Mathers et al., 2008). Research conducted by Mathers et al. (2009) found that fewer than one out of 10 teacher evaluation policies required evaluator training. Similarly, another study conducted by the Regional Educational Laboratory Midwest (2007) also found that fewer than one out of 10 district policies required evaluator training. This may affect an evaluator's self-efficacy, as research conducted by Dodson (2015b) indicated that 72% of the respondents surveyed

said the principal training they completed did not prepare them well for the implementation of the evaluation system. Over half of the respondents stated their state department of education did not provide adequate training to them for the implementation of the new state-mandated instrument used for teacher evaluation (Dodson, 2015b).

In a study conducted by the University of Michigan's Institute for Social Research (Rowna et al., 2013), when observers received their initial training, many principals lacked confidence in their ability to use classroom observation tools with fidelity. This study found that only 60% of principals felt confident to conduct teacher observations after their initial training, while 52% felt confident to conduct pre- and post-observation conferences with teachers. Among the participants, 39% of principals were confident that their scoring of lessons was in line with the scoring of others (Rowna et al., 2013). Also, 50% of principals agreed or strongly agreed that they needed more information about the protocol they used for evaluation (Rowna, et al., 2013). According to Archer et al. (2016), however, there is no magic number of how many hours of training are needed to ensure accuracy and meaningful feedback. It is safe to say observer training will take school leaders away from their myriad other duties for dozens of hours over the course of a year (at least when they are first trained) and that some observers will take more time and more reteaching to get through it. There have been previous studies that have explored observer training and its relationship to teacher effectiveness.

Principals' perception of teacher evaluation is that the evaluation process serves two purposes: (1) to improve instruction, and (2) to identify teachers who are not effective for intervention and possibly dismissal (Donaldson, 2013). However, some of

the limitations concerning evaluation cited by the principals interviewed in a study conducted by Donaldson (2013) found that burdensome paperwork and an evaluation tool that is obsolete were barriers to completing and conducting teacher observations. The Donaldson study began with semi structured interviews with 30 principals in a charter or public school system in two Northeastern states. Some principals in the Donaldson study stated that the culture of the building can sometimes dictate the number of teacher observations a principal conducts and can also discourage them from critiquing teachers through the evaluation process. This is consistent with research conducted by Halverson et al. (2004), who found that most principals viewed evaluation as a time management challenge, with increased meeting and paperwork requirements. According to Halverson et al., (2004) some principals adjust these requirements by streamlining their evaluation approach or cutting back on the amount and types of evaluation evidence. Others make changes to build in more time at school for evaluation activities.

In a study conducted by Dodson (2015b), surveys were sent to school principals in seven states with a response rate of 17%. The survey used Likert scale attitude measures as well as forced-choice and open-ended questions, leading to the discovery of some disturbing results regarding the attitudes and beliefs of the principals surveyed. Over half of the respondents surveyed stated they might leave their profession earlier than planned because of the task of implementing the new evaluation system and instrument. The early departure from the profession was due to the increase in the number of teacher observations principals must conduct to meet the requirements of the instrument. Some, not quite half, were also against the inclusion of students' test scores

as part of the teacher evaluation (Dodson, 2015a). Principals were concerned with the number of teacher observations that are required with the modern teacher evaluation systems and the amount of time required to fully complete the observation. This is similar to findings from a study conducted by Cosner et al. (2015), which also found that administrators had concerns regarding increased workloads and time demands with new evaluation systems.

According to Dodson's (2015a) research, most Kentucky principals were unhappy with the new teacher evaluation system, including the proficiency test they must take to implement the system. These principals required more training on how to use the new evaluation instrument being implemented (Dodson, 2015a). In a mixed methods study conducted by Norris et al. (2017), survey and interview data revealed a similar theme; that is, where's the training? This study used a survey sent to administrators in one urban school district, along with formal semi structured interviews of 10 of the original surveyed administrators. The findings in this study identified a weakness in the area of teacher evaluation training for observers. Administrators noticed within their training and certification process a shortcoming regarding how to conduct effective teacher evaluations. Administrators believed the mandatory certification process did not provide adequate training to conduct teacher evaluations (Norris et al., 2017). The interview strategy used by Norris aligned with the intention of my study. Building upon what is already known about the training of classroom observers, the data from this study revealed training practices observers perceive are effective and those that are not

effective. Data was collected regarding the systemic ongoing training practices that are beneficial to observers.

The Need for Quality Observer Training

The research reveals the need for additional training for observers, no matter the type of teacher evaluation rubrics being used to conduct classroom teacher evaluations (Cosner et al., 2015; Donaldson et al., 2014; Michigan Council of Educator Effectiveness, 2013; Semmelroth & Johnson, 2014; Smylie, 2014). Classroom teacher observation is a complex process and observers must be trained in the use of their school districts' selected observation tool. To support fidelity, the vendor or a designated certified person must provide this training. The goal of the training should be to ensure a baseline level of interrater agreement/reliability (Michigan Council for Educator Effectiveness, 2013). The primary observer is responsible for providing the teacher with summative feedback from all observations and developing goals for improvement in consultation with the teacher being evaluated (Michigan Council for Educator Effectiveness, 2013). The success of a new teacher evaluation system implemented in any state depends on the capacity of school principals in many different areas (Archer et al., 2016; Cosner et al., 2015). As of 2015, nationally, approximately 20 states provided teacher evaluation training or are planning to provide this type of training leading to evaluator certification (Cosner et a1., 2015). Additionally, some states required evaluators to pass an assessment to obtain certification prior to conducting teacher evaluations. Training opportunities are listed on each state's website along with guidebooks and online training for evaluators to increase their evaluation skill sets.

Rowna et al. (2013) found that training should consist not only of the basic introductory days of training provided by the vendor, but also additional calibration training designed to improve observation scoring and reduce rater error. Similarly, Putman et al. (2018) found that through ongoing calibration training for observers, identifying proficiency at each grade level, content area, or school campus is made more reliable. Kentucky requires all observers to complete specific instrument training, which includes 2 days of training ending with an assessment lasting 6 hours before observing their staff. Teachers must also complete and pass a calibration assessment every 2 years after the principal has passed the initial assessment (Bell et al., 2014). This is consistent with implications from research conducted by the Center for Public Education, which found that using well-trained observers further ensures that teachers being evaluated receive similar scorers regardless of who conducts the observations (Hull, 2013).

Research by Smylie (2014) also addresses the concern of training. This study looked at whether the training focused on the rubric itself or if it includes conveying evaluation results in a way that leads to the support of poorly performing teachers through professional development, remediation plans, or coaching to beginning teachers or poorly performing veteran teachers to address a weakness identified in their evaluation. The consensus was that teacher evaluation that simply labeled teaching as ineffective would not improve student achievement. Productive, timely feedback must be present accompanied by opportunities to learn. Each observation should lead to new goals to work on with specific professional development (Smylie, 2014).

Trained observers have the capacity to provide feedback where opportunities are afforded to each teacher to gain information through formal scoring and feedback routines. The observation/evaluation should encourage teachers to be more self-reflective, while at the same time creating more opportunities for conversations with other teachers and administrators about effective practices (Taylor & Tyler, 2012). Ensuring that observers can provide accurate and meaningful feedback in rich conversations with teachers is essential for improving teaching and learning (Archer et al., 2016).

Quality observation takes a special set of knowledge and skills. It has been well reported that teacher evaluation has swept the country in the past 8 years regarding improving public education. Seven challenges have been identified for this new movement: limited student assessment data; untested subject areas; quality of observers; individual versus team-based accountability; working conditions; and engaging all stakeholders (Cosner et al., 2015). Race to the Top and No Child Left Behind waivers were used by the federal government to push, prod, and cajole districts and schools to overhaul the way they assess teachers (Donaldson, 2013). Cosner et al. (2015) found that principals are going to need extensive training and ongoing support to develop necessary understanding and skills to enact new teacher evaluation systems. The researchers further found that directors of preparation programs would be wise to learn about new state evaluation systems and actively build alignment between instructional leadership courses, principal residency experiences, and new systems to cultivate the deep understandings that are essential for more robust teacher evaluation systems (Cosner et al., 2015).

Training of observers has several important dimensions: (a) observers must be able to recognize examples of the evaluative criteria in action; (b) observers must interpret the evidence for some aspect of teaching against the evaluative criteria; and (c) observers must make a judgment about the teacher's performance, linking interpretations to the descriptions of levels of performance (Cosner, et al., 2015). As a result, to address these issues, in this study, observer bias was mitigated in the data collection process during the interview of participants. Questions, follow-up questions, and probing questions were posed regarding observer perceptions of teacher effectiveness before and after training.

Observer bias is a major concern when performing classroom observations that can be combated through evaluator training. Murphy et al. (2004) pointed out that cognitive limitations and a lack of skills or information needed to accurately evaluate classroom teacher performance are also concerns when conducting observations. Moreover, evaluator motivation is also likely to affect the strength of the rating criterion relationship and can affect the degree of leniency of the evaluator (Kimball & Milanowski, 2009). Evaluator skill in observing and processing information about employee behavior may furthermore influence the performance rating. The more skilled the evaluator, the more likely he or she will give ratings that accurately reflect how the teacher actually performs on the dimensions of the evaluation instrument (Kimball & Milanowski, 2009).

Observers need to be aware of the many biases they may subconsciously hold during formative and summative observations. Addressing such biases as evaluator

motivation, attitude towards the evaluation tool, the relationship the evaluator has with the teacher, the will and skill of the evaluator, and evaluator training is necessary to avoid there being any doubt in the validity and reliability of the observation (Kimball & Milanowski, 2009). Kimball & Milanowski found that the more an evaluator is able to control these and other biases, the more likely he or she is able to give evaluation ratings that reflect the teacher's performance. Learning to control influences that hinder an effective observation is done through high-quality, effective training; however, even after high-quality initial training is in place, a school system may find that as many as 40% of trainees still need additional support (Archer et al., 2016).

Criteria for Differentiating Effective Teachers

Differentiating teacher performance does not happen simply because states and districts have a new evaluation rubric. Some policymakers and reformers have naively assumed that because states and districts have adopted new evaluation procedures that put a much stronger emphasis on student achievement, evaluation results will inevitably look much different; however, that assumption has proven incorrect (Doherty & Jacobs, 2013).

The core driver of teacher development is not accurate scoring or rating, but skillful coaching; that is, working with teachers on one specific concrete action that will improve results (Michigan Department of Education, 2016; Smylie, 2014). The goal of teacher evaluation is to provide a more valid measure of teacher quality by distinguishing between teachers at different performance levels (Putman et al., 2018). The MET project demonstrated that it is possible to identify great teaching by combining three types of measures: (1) classroom observations, (2) student surveys, and (3) student achievement

gains (Weisberg et al., 2009). Such observation practices should not only identify great teaching, but also provide the feedback that teachers need to improve their practice and that serves as the basis for more targeted professional development.

The MET Project was the first large-scale study to demonstrate, using random assignment, that it is possible to identify great teaching. The recognition of strong teachers for retention, while simultaneously encouraging fewer effective teachers to leave the profession, is one of the many responsibilities of the observer. This is in addition to helping all teachers learn and grow, as well as increasing student gains in learning and other positive student outcomes (Putman et al., 2018). Evaluation systems must be able to differentiate among different teachers' effectiveness and apply that information in key personnel decisions in order to build the strongest possible teacher workforce (Putman et al., 2018).

Observation requires significant instructional expertise and the ability to put aside long-held pedagogical preferences in favor of a shared vision of instructional quality (Archer et al., 2016). Keep in mind that even after training is well developed, it takes time for those trained to fully master the art and science of observation. Data from the literature suggest new evaluation systems that include multiple rating categories have not necessarily resulted in more differentiated ratings for teachers (Kraft & Gilmour, 2017). Nevertheless, the importance of training is evident with the increased emphasis on high-stakes evaluation and its impact on employee retention, dismissal, and promotion. Consistency among observers is a concern, as is an emphasis on the importance of training for the evaluation process and its final ratings. Schmitt et al. (1986) noted that

differences among evaluators can stem from a variety of reasons such as different opportunities to observe, experience, and the cognitive makeup or perspective through which the observer approaches the evaluation process. Knowing the policies and procedures of observers allows for a comparison of how each observer uses the information available to them in making final rating decisions (Schmitt et al., 1986). Training sessions can then emphasize certain aspects of the evaluation process, striving for more consistency among evaluators and addressing evaluator bias.

Identifying teacher effectiveness in classrooms is challenging for many reasons, including the difficulty of deciding on the definition of effective teaching. Observers do not totally agree whether a certain lesson was good or bad. The results regarding the interaction between observers and instructional sequences provide evidence that observers have different implicit theories about good instruction (Praetorius et al., 2012). Classroom observation is a highly challenging task requiring the observer to filter a dynamic and unpredictable scene in the classroom to find the most important indicators of performance, make an accurate record of the indicators, and then apply a set of criteria as intended (Archer et al., 2016).

The National Council on Teacher Quality (Doherty & Jacobs, 2015) referred to a pattern emerging across states showing that most teachers, almost all, are identified as effective or highly effective. Another source of pressure to revise teacher evaluation practices was the Gates Foundation project titled Measures of Effective Teaching, published in 2013. A basic premise of teacher evaluation systems is that the more proficient the teacher, the higher the academic outcomes of the students in that

classroom. School leaders have long been aware of the potential impact of an excellent classroom teacher and the academic achievement of the student in the classroom (Jacob, 2012).

There is a subjective aspect that plays into the evaluation of teachers by administrators (Green and Oluwole, (2015). Principals may feel strongly about a teacher they have hired and supervised for many years, making it very difficult to be critical of their classroom practices. Teachers in the past may have been given a satisfactory rating but are now under new performance-based evaluation systems and changes are being made in the way evaluation ratings are awarded. Principals are also giving effective ratings for teachers that may need improvement, in order to avoid having difficult conversations with those individual teachers (Doherty & Jacobs, 2013).

Papay and Kraft (2016) stated that teachers can continue to improve over the course of their career under the condition that teacher evaluation allows for detailed, valid feedback on classroom practices, as well as support for improvement from knowledgeable and well-trained administrators or peers. These are just some of the dynamic characteristics for today's administrators performing classroom observations that need to be addressed while trying to overcome the barriers and challenges of the evaluation process.

The Current State of Teacher Evaluation in the State of Michigan

Beginning with the implementation of the federal Race to the Top policy and its agenda for increasing teacher quality and effectiveness, potentially leading to an increase in student achievement and learning, Michigan, like many states, has implemented new

and more complex teacher evaluation systems. This, in turn, means that observers need to have access to certain resources, develop new knowledge and procedures, and build new skill sets if these new evaluation systems are to be successful. Training needs to be provided systemically throughout the tenure of all principals if the skills needed for effective observations are to be developed and maintained. Kimball and Milanowski (2009) stated that it would take more than basic training for observers and other evaluators to successfully enact the new teacher evaluation systems. Research by Cosner et al. (2015) would concur that observers are likely to need extensive training and ongoing support to develop the necessary understandings and skills to successfully implement the new teacher evaluation practices. The United States Government cannot mandate state adoption of any evaluation system or evaluation reform initiatives. However, when federal dollars are attached to education policy and money for school districts is dependent upon the implementation of evaluation policies such as the Race to the Top initiative, states are obligated to adhere and follow federal guidelines of educational reform such as the new teacher evaluation initiatives that are now in place in most states (Croft et al., 2016). School districts in Michigan have been forced to provide systemic training for teachers and administrators on how to use their newly selected evaluation tool in order to receive federal funding.

As stated in the introduction to this chapter, there is a problem in the state of Michigan's education system regarding the ability of observers to identify or report the true levels of effectiveness of teachers when conducting classroom teacher observations and final evaluations (Kraft & Gilmour, 2017; Norman, 2010; Steinberg & Sartain, 2015;

Weisberg et al., 2009). Observers acquire the knowledge to provide accurate evaluations through high-quality, effective training that is systemic and ongoing and includes being able to provide feedback to teachers. Many observers, however, may have limited capabilities, time, and preparation to engage in post-conferencing and coaching feedback effectively, resulting in a poor teacher–evaluator observation experience for observers and teachers (Smylie, 2014).

The reinvention of teacher evaluation systems has stemmed from three forces, according to Smylie (2014): (1) a consensus that teacher evaluation at the current time is ineffective, (2) a strong influence from President Obama's Race to the Top legislation, which provided U.S. Department of Education waivers and funds for those revamping their evaluation systems, and (3) the idea that new teacher evaluation systems would increase teacher effectiveness, leading to an increase in student achievement (Smylie, 2014). During the research on teacher evaluation, Smylie found that training received by observers must cover a wide range of areas including focusing on the rubric itself, conveying evaluation results, helping poorly performing teachers develop professional development or remediation plans, or providing coaching to beginning teachers or to poorly performing veteran teachers to address a weakness identified in their evaluation. This includes rigorous training and certification of observers and observing multiple lessons by multiple observers (Ho & Kane, 2013).

Michigan Law and Teacher Evaluation

The state of Michigan has established observation tool requirements. One of these requirements is training for individuals conducting classroom teacher observations and

final evaluations. Also, Michigan requires at least two classroom observations per year, of which one must be unannounced. The administrator responsible for final evaluation must conduct one of the observations. Feedback must be provided to the classroom teacher within 30 days of the observation and must include a review of lesson plans along with classroom student engagement.

Michigan law governs educator evaluations for teachers and administrators, and when implemented with fidelity, will be a key strategy in the efforts to see Michigan become top 10 in education among states in 10 years (Michigan Department of Education, 2016). High-quality evaluations provide teachers with critical feedback on how they can improve their own practice to impact the lives of students. Districts can use evaluations to identify trends, develop data-driven strategies, and plan professional development to the needs of teachers (Michigan Department of Education, 2016).

In Michigan, PA 173 of 2015 requires all teacher observation tools used by school districts be research-based, with the authors and their qualifications identified; the tool providing evidence of reliability, validity, and efficacy; and a plan for developing evidence, evaluation frameworks, and rubrics (Michigan Department of Education, 2016). The school district must also have in place a process for observations and feedback to teachers, along with a plan for training of evaluators and observers (Michigan Department of Education, 2016). PA 173 Section 1249 (c) also mandates that student growth is used as a significant part of the final teacher evaluation. Including student growth in the evaluation process requires that the 5 Dimensions Plus Rubric must also contain a final scoring section allowing for student growth percentages. In addition,

according to PA 173 Section 1249 (2)(n) of 2015, beginning with the 2016–17 school year, a school district shall ensure training is provided to all observers who will be performing teacher observations. This law created a possible learning gap for observers of teachers regarding their formal training on how to perform and conduct teacher observations and final evaluations.

Taken as a whole, the principals who work with the new teacher evaluation systems are expected to develop new procedural understandings of the teacher evaluation process, while also increasing their skills in recognizing quality teacher classroom practices and pedagogy (Cosner et al., 2015). Specifically, Cosner et al. (2015) found principals conducting classroom observations are expected to acquire a new understanding of student learning objectives and broaden their abilities to collect and manage large volumes of evidence from numerous data sources, while at the same time enhancing their communication skills to provide meaningful, evidence-informed feedback to teachers after each classroom observation.

Ideally, observations should provide teachers feedback on all aspects of a lesson—both the quality of the content and skill in presenting it. This feedback is crucial to improvement in teacher effectiveness. In most Michigan schools, observers still struggle to provide specific, high quality and useful feedback on what they see in a teacher's classroom, much less provide feedback that can help teachers adapt their instruction to the common core state standards (The New Teacher Project, 2013).

Rater reliability, as referenced earlier (Kane et al., 2011b; Hill et al., 2012; Semmelroth & Johnson, 2014), encompasses many aspects such as rubric knowledge,

knowledge of proven research-based instructional strategies, and knowledge of content standards and benchmarks, along with quality systemic training to recognize quality pedagogy. While rater certification is not required of classroom observers in the state of Michigan, rater certification is available with the 5 Dimensions Plus evaluation rubric, which is mentioned in my study.

According to the *State of the States* report on evaluating teaching, leading, and learning, a troubling pattern has emerged across states that are implementing a new performance-based teacher evaluation system. It was found that a vast majority of teachers are identified as either effective or highly effective (Doherty & Jacobs, 2015). In the years 2012 through 2014 in the state of Michigan, it was found that just over 3% of teachers were rated minimally effective or ineffective on their final ratings.

Current Observer Training

To collect evidence, the observer needs to know what evidence is, and what kinds of evidence are relevant. To rate performance, observers need to understand the conditions under which each rating is merited. To provide feedback effectively, one needs to know how to coach. Training requires the most resources and the most know-how (Archer et al., 2016). Districts could invest in training for principals focused on how to navigate difficult conversations about substandard performance with individual teachers and their entire staff (Kraft & Gilmour, 2017). With deeper training about teaching practices, principals are also likely to need specific training on how to provide evaluation feedback (Cosner et al., 2015). Archer et al., (2016) would also agree with Kraft and Gilmour (2017) and Cosner et al., (2015) that ensuring observers can provide

accurate and meaningful feedback in rich conversations with teachers is essential for improving teaching and learning.

Observer training must also include the use of a web-based or a database application an evaluator can use to record or script teacher actions, behaviors, classroom notices, and wonderings. Training for observers does not guarantee a successful observation with reliable ratings for teachers. In the study conducted by Praetorius et al. (2012) involving trained and untrained raters, the focus was on identifying the differences between trained and untrained raters about their final ratings of teachers. Rather than focus on trained versus nontrained observers, this study conducted a thorough analysis of the data and explored the training strategies used for classroom observers and the perceived outcomes of those strategies to conduct effective observations.

The observer training that will be the subject of my study will be the 5 Dimensions Plus Techer Evaluation Rubric created by the Center for Educational Leadership at the University of Washington. The 5 Dimension Plus Teacher Evaluation Rubric has 6 dimensions, each containing indicators of performance. The Center for Educational Leadership (CEL) recommends 3 days of extensive training, two dimensions each day, for classroom observers using the 5 Dimensions Plus Teacher Evaluation Rubric. A 3-day rater reliability training is also available, but not required by the state of Michigan, for observers who would like to calibrate their rating skills with the staff from CEL regarding observer practices. Although it is mandated by legislation in the state of Michigan that all observers and teachers receive training, the quality of the training received was the topic of my research study. The 5 Dimensions Plus Teacher Evaluation

Rubric had been used in Michigan since the 2012–13 school year, when Michigan formed the MCEE. The MCEE was formed to pilot four potential observation rubrics in schools to determine which rubric best fit the needs of schools, leading to the selection of the observation rubric. The four evaluation rubrics included Charlotte Danielson's framework for feaching, Marzano's teacher evaluation model, the Thoughtful Classroom, and 5 Dimensions Plus Teacher Evaluation Rubric for Teaching and Learning. Michigan followed the Council's recommendations, which included allowing local school districts to select any one of the four evaluation tools to use. The Michigan legislature has since passed legislation that allows local school districts to select any rubric besides the four suggested, provided they can fulfill PA 173 sections 3 (a), (b), (c) and (d) requirements. In this study I addressed the training of observers in Michigan by gathering data related to the research questions during the interview process.

Ho and Kane (2013) also stressed the importance of well-trained observers who should be assessed for accuracy before being allowed to rate classroom teacher performance. It is not only important for quality training to be received by classroom teacher observers; it is just as important to sustain those skills acquired in training by calibrating skills throughout their experiences as a classroom teacher observer. Observers in Michigan do have an opportunity to calibrate their observation skills and become rater-certified using the 5 Dimensions Plus Teacher Evaluation Rubric through a certified trainer trained at the CEL on the University of Washington Campus.

There are many proven observation strategies that make up an effective observation such as observation certification that includes extensive knowledge of the

rubric being used. As mentioned earlier in this chapter, observers in the state of Michigan have the opportunity through the 5 Dimensions Plus Teacher Observation Rubric trainers to become rater-certified. Recent studies have found that the use of multiple observers, along with ensuring observers are certified, is a critical factor for achieving acceptable and reliable evaluation ratings for teachers (Culbertson, 2012; Semmelroth & Johnson, 2014). Observer training should also include extensive teaching, familiarization, and knowledge of the criterion-based assessment or evaluation rubric used to perform classroom observations. Observers who are trained and familiar specifically with one evaluation rubric will promote consistent and fair observations. Stitt et al. (2003) found that communication of teacher expectations by observers should be evident and explicitly explained to teachers. Moreover, observations results should encourage teachers to participate in their own learning by identifying what to work on towards for improvement.

Lack of Observer Training

Lack of training can threaten the reliability of evaluation and the objectivity of the results, which may lead to the misuse of the evaluation instrument itself and a reduction in the ability of the instrument to improve teacher classroom performance (Mathers et al., 2008). Evaluators need to understand quality teaching, the rubric being used, its characteristics, and the teacher behaviors it intends to measure (Mathers et al., 2008). Two of the greatest challenges facing classroom teacher observation are the inconsistencies with the application of teacher evaluation practices and a lack of trained and knowledgeable observers (Mathers et al., 2008). Observers need to understand and be

able to identify quality teaching, along with the extensive understanding of the teacher evaluation rubric. Mathers et al., (2008) also found that without teacher evaluation training, observers may be unaware of the potential bias they may bring to the classroom observation. Quinn (2014) confirmed in an interview with Robert Marzano that there is a pattern in teacher evaluations showing that individuals awarding final evaluation ratings are giving ratings that are much higher than deserved. Similarly, according to Kane et al. (2011b), a stumbling block in teacher observation practices is a lack of consensus on valid measures for recognizing and rewarding effective teaching. This may stem from a lack of training for those administrators conducting classroom teacher observations. Based on this information, the question arises whether classroom observations are hopelessly flawed in assessing teacher effectiveness (Kane et al., 2011b). Kane et al. (2011b) adhere to the thought that quality training can overcome these types of barriers. However, Kane et al. (2011a) found that in the Cincinnati Public School System's TES, when based on sound well executed classroom observations performed by trained evaluators, observers are able to identify effective teachers and their practices. However, Dodson (2015b), found that in some principal preparation programs, it was recommended that individuals spend significant time doing field experience along with traditional training strategies. In a multistate study conducted by Dodson, he found that principals believed that practical training helped them prepare for leading schools. Although not all states involved mentioned the benefits of gaining practical experience in observation as a top benefit, it was a significant experience for some principals in three of seven states and was thought to be beneficial in all states (Dodson, 2015b). Herlihy et al. (2014) also

found, while conducting a research study in 17 states, each receiving a grant from the Race to the Top grant monies, that enhanced training of observers is needed to meet standards. The majority of states, including Michigan, require principals or other observers to have the training to conduct teacher evaluations; however, the quality of the training provided remains in question (Doherty & Jacobs, 2015).

Starting in the 2016–17 school year, Michigan teachers and administrators were required to receive training on the teacher observation tool and how it is used (Michigan Department of Education, 2016). Those individuals who conduct the teacher observations and final evaluations must also receive expert-provided training on the implementation of the tool with fidelity (Michigan Department of Education, 2016). In a study conducted by the University of Michigan Institute for Social Research for the Michigan Council for Educator Effectiveness, it was found that school districts who used the 5 Dimensions Plus Teacher Evaluation Rubric tended to rate teachers lower than the other teacher evaluation rubrics (Rowna et al., 2013). Much can be speculated about this scoring difference compared to other rubrics used by districts in the study, such as rater bias, rater reliability, different areas observed, and teacher effectiveness. Researchers state the best guess about the differences in the observation rubrics used in the study might be differences in the emphasis of the 5 Dimension Teacher Evaluation Rubric in the weight given to assessing cognitively demanding instruction.

Since that time, no studies have been conducted surrounding the 5 Dimensions Plus Teacher Evaluation Rubric. Interview questions during the data collection phase of my study addressed some of the issues brought up in the University of Michigan study.

Keeping in mind that the focus of this study was administrators who have received the training provided by certified trainers for the 5 Dimensions Plus Teacher Evaluation Rubric, I addressed the gap in the literature that deals with the quality and effectiveness of the observer training for teacher evaluation in Michigan. Increasing the quality of observer training will enable users of the 5 Dimensions Plus Teacher Evaluation Rubric to be more effective in observing and coaching teachers, leading to more effective teachers, resulting in increased student achievement in the state of Michigan.

For Michigan teachers and administrators, it is the obligation of the school district to provide training on the observation tool and how it is used (Revised School Code Act 451 of 1976 MCL 380.1249). Individuals conducting classroom teacher observations and final observations having expert-provided training on the implementation of the observation tool with fidelity is a necessary requirement to assure reliability and validity (Michigan Department of Education, 2016).

Summary and Conclusion

The purpose of teacher evaluations is to determine educator effectiveness while ensuring ample opportunities for improvement are provided, including relevant coaching, support, and professional development (Michigan Department of Education, 2016). The ability to differentiate effectiveness among teachers when performing classroom teacher observations or final evaluations is critical to student success; however, it seems to be a skill not yet obtained by most observers (Kraft & Gilmour, 2017; Norman, 2010; Steinberg & Sartain, 2015; Weisberg et al., 2009). Norman (2010) states two of the most important challenges we face in improving teaching are distinguishing good teachers

from bad teachers and helping all teachers improve. Knowing what impacts student learning will directly result in evaluating teachers and improving the teacher evaluation process. Dowley and Kaplin (2014) stated the success of teacher evaluation depends entirely upon the rigor with which the systems are implemented in each individual school district.

High-quality evaluations provide teachers with feedback that is critical for the improvement of their own practice, which in turn impacts students' lives (Michigan Department of Education, 2016). When valid and reliable evaluation systems are put in place with fidelity, districts can use evaluations to identify trends and develop professional development strategies that meet the needs of educators (Michigan Department of Education, 2016). In this study I gathered data regarding the perceptions of observers who are using the 5 Dimension Plus Teacher Evaluation Rubric regarding the quality of training received.

The Dodson (2015a) study proposed several recommendations included training for observers, taking and passing an observer exam, continued calibration, and certification requirements for observers in order to conduct classroom observations. The 5 Dimensions Plus Teacher Evaluation Rubric also uses a collaboration and certification strategy by having observers watch and rate videos that have been rated by the experts at CEL. There are characteristics of all effective performance evaluations, starting with being research-based, clear, straightforward performance standards, and having multiple rating options that include frequent observations from multiple observers providing timely feedback. However, as The Widget Effect (Weisberg et al., 2009) study suggested,

there is no single correct model for performance evaluation when observing teachers, so it recommended that observers be well trained at conducting observations effectively.

To date, little research has been conducted regarding observer training and its correlation to teacher evaluation scores. Also, concerning the field of rating instructional quality, only a few investigations have looked at the thinking and reasoning behind the ratings of observers surrounding instructional quality (Praetorius et al., 2012). No research was found specifically addressing the quality and effectiveness of training provided by the 5 Dimensions Plus Teacher Evaluation method. my research addressed this gap in the literature. Cosner et al. (2015) recommends that state training designs have a combination of both in-person learning experiences facilitated by highly skilled trainers, along with targeted online training and resources such as videos, PowerPoint presentations, and written guidance including group and individual discussions of teacher pedagogy through classroom visits or similar district practices. In addition to these recommendations, Smylie (2014) suggested strengthening the link between professional development within the teacher evaluation system and practice, while increasing opportunities for professional development for teachers and evaluators. Increasing the capacity of school administration, particularly those conducting observations and final evaluations, by allocating resources is one of the most efficacious acts states and school districts can take to increase the effectiveness of the teacher evaluation system (Smylie, 2014). Training along with certification training is made available in the state of Michigan through the professional organization for junior high and senior high school

principals multiple times throughout the school year for the 5 Dimensions Plus Teacher Evaluation Rubric.

Using data obtained from a variety of school districts, I compared rural, urban, and suburban school district principals/observers on their perceptions of teacher evaluation preparedness, focusing on training for administrators. Implementing a new teacher evaluation system comes with many challenges for observers and school districts. Some challenges include observer time needed to conduct formative observations, using multiple observers when conducting formative observations, and an increased workload when implementing the evaluation system with fidelity. Data was gathered with the intention of providing information to increase the effectiveness of observers by identifying areas of improvement and strategies for professional growth and development for teachers in specific areas. Areas of observer training explored included initial training, field experience, strategies that are used and those that are not used, and systemic ongoing training for observers.

My study addressed gaps in the literature regarding school district training programs and their approaches for initial training and systemic support for districts using the 5 Dimensions Plus Teacher Evaluation Rubric. Data collected can lead to training practices that will generate stronger teacher evaluation practices for district observers, which may result in increased teacher effectiveness. I used a qualitative approach by interviewing selected individuals who conduct classroom teacher observations and final evaluation ratings. Questions during the data collection phase addressed the observation/evaluation practices that are working and those practices that are not working

and the areas participants feel additional training is needed to increase their capacity to differentiate effective levels of teaching. Chapter 3 will address the research design and rationale for the selection of a qualitative study, as well as the process of gathering data for analysis. The methodology of participant selection and criteria will also be addressed in Chapter 3, along with instrumentation, trustworthiness, data reliability and validity, and ethical issues and how they will be addressed.

Chapter 3: Research Method

As shown through the review of literature, more training is needed concerning teacher effectiveness when conducting classroom observations and final evaluations in the state of Michigan (see Cosner et al., 2015; Norris et al., 2017; Rowna et al., 2013; Weisberg et al., 2009). Currently, teacher observations and evaluations are connected to high-stakes accountability, as well as ongoing professional development and support measures. In an online survey of over 380 observers conducted by Dodson (2015b) it was found that a large majority of respondents, 72%, said the principal preparation program they completed did not prepare them well for the implementation of the evaluation system, and over half, 56%, said the state department of education did not provide adequate training to them for the implementation of the new instrument. Despite this study, little is known about the perceptions of the observers regarding the quality of training received to conduct classroom observations and evaluations. The purpose of this study was to gather data focusing on the quality of training received and its effectiveness by those individuals who conduct classroom teacher observations and final evaluations. In this study, I gathered data regarding the training strategies used for classroom observers and the perceived outcomes of those strategies to conduct effective observations. The research included participants who are currently observing, evaluating, and rating teachers using the 5 Dimensions Plus Teacher Evaluation Rubric in the Michigan public school system.

Chapter 3 provides a description of the research method that was used to conduct this study. The first section describes the study's research design and the rationale for using this approach. A diversified participant pool was gathered that provided rich insights regarding observer training which may lead to new findings and training strategies. Next, the study's research questions are restated and provide a foundation for the interview questions (Appendix A) that were used in this research. The role of the researcher and the data collection strategy are explained, followed by participant selection and instrumentation. Data collection and data analysis processes are then discussed in detail, along with the justification for the chosen strategies. Validity and credibility strategies to ensure trustworthiness throughout the research process are also discussed in detail and how those assurances were accomplished. Qualitative research design limitations were mitigated through such strategies as journaling, bracketing, and working with a critical friend to ensure reflexivity and objectivity. The chapter concludes with a discussion of ethical issues.

Research Design and Rationale

I used a basic qualitative research design for this study. Merriam (2009) described qualitative research as using rich descriptive words about a phenomenon while being cognizant of the changing conditions of the phenomenon. In the current study, the participants' perceptions about the quality and effectiveness of observer training were studied. A qualitative research design allowed the collection of data from the experiences of observers with an interview protocol. My prior work with classroom teacher

observations provided some experiences conducting instructional conversations with staff members and peers regarding observer training and enhanced this research design.

With this research, I looked at the perception of effectiveness of training for classroom observers who are currently working with the 5 Dimensions Plus Teacher Evaluation Rubric. More specifically, the phenomena of interest in this study arose when it was discovered that 97% of teachers in Michigan were rated on their final evaluations as either “effective” or “highly effective” since the new teacher evaluation rating system rating was implemented in 2011; meanwhile, student achievement scores continue to decline in the state, leading me to look further into the quality of observer training. The basic qualitative research design selected for this study was considerate of strategies traditionally used to gather information on the experiences and perceptions of those individuals responsible for observer training.

A quantitative research study was conducted examining the relationship of the number of classroom observations, observation characteristics, evaluator characteristics and the school districts letter score from the state of Indiana in the 2012–13 school year (Pies, 2017). Pies (2017) focused on quantity of observations along with observer and observation characteristics. This study did not consider the training of observers, which is a prelude to observation. In a qualitative phenomenological study by Stewart (2016), individual educators, particularly teachers, were asked how educators experience and make sense of the new teacher accountability evaluation system. This study focused on teachers and their experiences, not administrators. In 2014, a qualitative study was conducted regarding the Michigan teacher evaluation laws; it included surveys and

interviews with teachers and administrators in school districts in southeastern Michigan. The study relied on Gorte's framework for effective evaluation implementation, which is based upon several components (Mowrer, 2014). This study also did not explore the quality of training provided to observers. A qualitative study conducted by James (2014) examined the abilities required to be an effective evaluator of teachers and what skills, knowledge, and training are required to reach this desired ability level. This study also used SCT as a framework for the skills needed of an observer to effectively evaluate teachers. While the James study was more aligned with my research study, by identifying the skills and abilities, it did not examine the training techniques and quality needed for observers to reach a desired ability. The study which comes closest to my study was conducted by Pisciotta (2014); however, it also did not address the quality of observer training. For the above reasons, it was the purpose of this qualitative study to address the perceptions of participants regarding observer training, diving much deeper into the area of quality of strategies used during observer training.

In conclusion, the use of a qualitative approach offered several advantages for this research study that led to the gathering of data in real-world settings as interviews are conducted using open-ended questions under conditions that are comfortable and familiar to the participant (see Patton, 2015). Any component of qualitative research may need to be reconsidered or modified during the study in response to new developments or changes to other components (Maxwell, 2013). Qualitative research takes the idea of moving back and forth between the different components of the study, assessing the implementation at each phase (Maxwell, 2013). Therefore, while an interview protocol

was used to explore the perceptions of the participants of this study, the qualitative research design allowed for additional probes and member checking in order to fully answer the study's research questions, which are stated below.

Research Questions

The research questions listed guided the data collection process as a foundation for developing interview questions that generated rich data regarding observer training. However, interview questions may need to be modified or expanded as a result of new discoveries or new learning while conducting the research (Maxwell, 2013). The interview questions (Appendix A), stemmed from the research questions that generated data addressing the training received by principals as well as anticipated follow-up questions that dove deeper into practices and behaviors on how to improve observer training in the state of Michigan. It was anticipated that the data served to enhance a training protocol that addresses perceptions of any identified gaps in observer training.

RQ1. What are the perceptions of school administrators who are responsible for evaluating teachers about the additional training they need for performing effective classroom observations?

RQ2. What are the perceptions of school administrators who are responsible for evaluating teachers about the training they have received for performing classroom observations?

Interview questions served to elicit data that answer the research questions that were used in this research study and can be found in Appendix A. The research questions

addressed the observer's experiences in training and the implementation of the 5 Dimensions Plus Teacher Evaluation Rubric.

Role of Researcher

The qualitative approach requires the researcher to bracket or set aside their own experiences and biases in order to understand those participating in the study (Creswell, 2009). According to Patton (2015), emotions and feelings lead to caring, and caring is a primary source of bias. By using a value-neutral approach to this research study, I kept my personal bias in check during all phases of the research study. As a researcher, I brought over 18 years of administrator experience working with teachers, observing teachers, and conducting final teacher evaluation ratings. I have over 25 years in education working with elementary, junior high, and high school teachers, and have worked with the 5 Dimensions Plus Teacher Evaluation Rubric for over 7 years. In this qualitative research study, the role I took on as the researcher as defined by Creswell (2013), I became the key instrument of data collection as I conducted interviews, examined documents, and observed behavior.

In this study, I used an interview strategy to gather data that were focused on the experiences of observers. As the researcher, one of the roles I played was the role of being the key instrument in collecting and analyzing the data. During the interview process, data was collected not only through the interview questions, but also by being aware of body language and behavior during the interviewing process (Creswell, 2013). I also needed to have a two-way awareness where I played the role of the interviewer/researcher as I watched the interviewee. The same is true as the interviewee

is aware of my mannerisms during the interview (see Patton, 2015). Janesick (2011) pointed out that the interviewing process increases one's skill set and mindset as a qualitative researcher, building on existing observation skills. In this process, the quality of the information obtained during an interview is largely dependent on the interviewer (Patton, 2015). Another important role I assumed as the researcher was to develop rapport and establish a trusting relationship to develop the data needed for greater understanding. One way to establish a trusting relationship where participants felt comfortable was to emphasize the focus of the study on the quality of the training received to conduct classroom observations, not the observer's performance during such observations. The interview complexity requires that the interview accommodates contextual shifts and reflexivity, and rather than suppress a respondent's reflexivity, the active interviewer encourages these shifts and reflections (Holstein & Gubrium, 1995). I bracketed my biases during the interview process, allowing the interviewee to expand on any subject while continuing to probe in a direction with follow-up questions to generate rich, deep data.

During each interview, I asked open-ended interview questions and listened closely to the participants I interviewed. I refrained from assuming the role of an expert researcher with a rigid protocol regarding my questions and acted more flexible, allowing the interviewee to guide the flow of the interview. However, follow-up and probing questions were used to keep the interviewee on track. As the interviewer, I needed to understand the importance of recognizing the meanings, beliefs, and attitudes of the participants in the study (see Maxwell, 2013). Being able to become more refined during

the process of research to reflect on an increased understanding of the problem was critical during the interviews (see Creswell, 2013). This was accomplished using the Rubin and Rubin (2012) model for conducting interviews, which offers a more simplistic and flexible structure consisting of seven stages in the interview process. I used the following stages in designing my interview strategy for this study recommended by Creswell (2013):

- Design open-ended questions that are directly linked to the research questions stated in this chapter and in Chapter 1.
- Identify participants who meet the participant criteria.
- Decide on the type of interview to use.
- Use adequate recording devices for all interviews.
- Design the interview protocol.
- Continually refine interview questions if necessary.
- Arrange interviews convenient for participants and stick to all time and location commitments.

Professional Relationships with Participants

At the time of this research study participants were employed as public-school administrators who conduct classroom teacher observations and final evaluations using the 5 Dimensions Plus Teacher Evaluation Rubric. Participants were not from the school district where I was employed as an administrator; therefore, a supervisory relationship did not exist. All participants held the same position level within school districts; therefore, any type of power imbalance was minimized while conducting interviews.

Among the identified school districts currently using the 5 Dimensions Plus Teacher Evaluation Rubric, I selected participants based on their demographic data using the Michigan school data web site (www.mischooldata.org), which provides all demographics for public schools.

As a long-time member of MASSP, I have established an extensive network with colleagues from school districts across the state of Michigan. As a result, I used this network as a resource for potential participants. MASSP has established a list of school districts that are currently using the 5 Dimensions Plus Teacher Evaluation Rubric which I used to recruit potential participants. This list was used to identify school buildings and school districts that are using the 5 Dimensions Plus Teacher Evaluation Rubric. I accessed the Michigan Department of Education's website to identify rural, suburban, and urban school districts and cross reference these schools with the MASSP list of school buildings and school districts. Once building and districts were identified from the resources used, invitation letters, emails, and follow-up phone calls were used to contact and invite potential participants. All efforts to contact potential participants were cold calls, where no previous relationship has been established or any influence to participate was present.

Rubin and Rubin (2012) referred to the interview process as a relationship based on trust, where the interviewer will protect interviewees in situations where information could be embarrassing to the interviewee. However, even before this protection takes place, I established trust by demonstrating that we share a common background and have relevant job experiences that would make the interviewees more confident that I

understood their answers (Rubin & Rubin, 2012). As an interviewer, it was important to establish rapport with the participants, but the rapport must be established in a way that does not undermine the neutrality concerning how the person responds to the questions (Patton, 2015). As Patton (2015) pointed out, neutrality must be present at all times, which allows the interviewee to respond without judgment on the interviewer's part.

Management of Researcher Bias

As the interviewer, I stayed focused on participant meaning, allowing me to control interviewer bias. As Creswell (2013) suggested, during the entire qualitative research process I maintained a focus on learning the meaning the participants hold about observation training, not the meaning or beliefs that I brought to the research. This was accomplished throughout the interview process by developing and following a predetermined protocol that gathered data addressing all the research questions. Following the protocol helped in ensuring the integrity of the research and the validity of the data collection process addressing any bias the researcher may have (Maxwell, 2013). Questioning also controls researcher bias in that question design does not undermine the neutrality of the participant's response, allowing the participant to convey any information without engendering favor or disfavor regarding that response (Patton, 2015).

I identified all biases I brought to the study through the process of self-reflection. As a principal who observers teachers and have had the 5 Dimensions Plus training, I have been conscious of my personal experiences regarding training. I used a colleague in my own school district along with two practicing observers in nearby districts to act as practice participants to assist in identifying any bias I may have in my design and

delivery. As stated, I brought over 20 years of observer experience, along with completion of multiple professional development sessions, enabling me to establish a relationship with participants because of our common experiences while obtaining more accurate and valid findings. As the qualitative researcher in this study, I set out to understand the quality of training received by the observer while documenting the reality of their experiences, making no attempt to manipulate, control, or eliminate any data, and realizing the possible variables in responses. Using the bracketing strategy of setting aside personal experiences, biases, and preconceived notions about the research topic assisted in mitigating biases I may have had. Also, using a critical friend, in this case a colleague who has over 15 years of observer experience, as a resource during the research phase of the study provided another perspective and helped keep a neutral approach and mitigate my biases. As a researcher it was important to refrain from allowing my judgements to responses and remain open to possible new perspectives of the phenomena being studied. More information is provided in the methodology subsection of this chapter.

Other Ethical Issues

This research study did not take place within the school district where I am currently employed. I did not have any supervisory role with participants, nor am I currently sitting on any professional committees with participants. Participants may have strong bias towards the evaluation/observation process in the state of Michigan. I asked participants to expand on any such bias or any biases towards the observation requirements to fully understand their perception. As an interviewer I needed to keep an

awareness of perceived biases or strong beliefs towards teacher observation and evaluation from each interviewee. I used questions to redirect the conversation to a less stressful set of questions while still gathering rich data for the study.

In the initial introduction letter explaining the purpose of this research study, it was noted that all participants potentially have an impact on how future observer training is delivered and the strategies that may be emphasized during training sessions. It was written in the consent form signed by participant and researcher that all participants remain anonymous and the information they provide will remain confidential. All data collected remains stored on a data disk kept in a lockbox in the home office of the researcher. All data gathered and connected to this research study will be destroyed 3years after the study has been finalized and approved by Walden University.

Methodology

Participant Selection Logic

The population from which participants was selected for my research study were those individuals who conduct classroom teacher observations and final evaluations at the secondary, junior high, or middle school level in the state of Michigan. These individuals are currently employed in a public school system that was using the 5 Dimensions Plus Teacher Evaluation Rubric in rural, urban, and suburban school districts. Participants were selected from a list of identified school districts currently using the 5 Dimensions Plus Teacher Evaluation Rubric accessed from the Michigan Association of Secondary School Principals. The purposeful recruitment of participants generated a selection pool that had a varying number of years of experience conducting observations and final

evaluations. However, as the researcher I selected both participants that had limited years of observing and participants who had an extensive number of years of experience observing teachers.

Qualitative research typically focuses on small sample sizes. The participants in my my study were part of a purposeful sample that intentionally sampled a group of classroom observers that can best inform the study about the effectiveness of observer training (Creswell, 2013). The selection and the strategies that were used relied on the works of Patton (2015). The effectiveness of purposeful sampling lies in the selection of information-rich participants for an in-depth study in which one can learn a great deal about issues of central importance to the purpose of the study. In-depth information from a smaller sample size can be valuable, especially if the identified participants are very valuable and are information rich. What was critical to the selection process was that the sampling strategy and decision be fully described, explained, transparent, and justified so that information users, such as district and state educational administrators, have the appropriate context for judging the sample set.

Participant Selection Criteria

Participant selection criteria were based upon the participants being an active classroom observer and final evaluator and currently using the 5 Dimensions Plus Teacher Evaluation Rubric. At the time of this research study there were over 150 school districts using the 5 Dimensions Plus Teacher Evaluation Rubric, according to the MASSP website. Using the MASSP list and the Michigan Department of Education website, which identifies rural, urban, and suburban school districts, to identify potential

participants, invitations were sent to the participants inviting them to participate in the research study (Appendix B) via email, followed up with a phone call to potential participants as the recruitment strategy for this study. School district websites along with the Michigan Public and Private School Directory were also used to access names and contact information for potential participants. Participants may have had varied amounts of experience of conducting classroom observations.

Number of Participants

As Patton (2015) pointed out, the focus in a qualitative study is on a relatively small sample size to provide in-depth information that is valuable and information-rich. The intent of my qualitative research study was not to study a high number of participants; rather, this study involved fewer participants, enabling more extensive in-depth data to dive into details and information from each participant (Creswell, 2009). I decided upon a sample size of 13 participants to fit the time frame of this study and to gather a sufficient amount of data to reach saturation.

Patton (2015) suggested that the purposeful sample size is about judgment and negotiation. The recommendations are that qualitative researchers state a minimum number based upon reasonable coverage of the phenomenon and purpose of the study, keeping in mind the design should be flexible and emergent (Patton, 2015). However, if data collection and the coding process shows a potential pattern unlike a majority of the data collected, more participants may need to be selected to fully saturate the data to a point where no new trends or patterns are evident. Once the minimum number of participants was reached, I continued to interview participants as long as each interview

presented more refined or a different perspective to the quality of observer training, thus reaching the saturation point (Rubin & Rubin, 2012). To avoid this, an additional selection of five to 10 participants were identified and put on a list to be contacted if necessary.

Specific Procedures for Identifying, Contacting, and Recruiting Participants

Participants were identified using the MASSP master list of schools currently using the 5 Dimensions Plus Teacher Evaluation Rubric. From the list, a purposeful sample of school districts was identified, along with the buildings and building administrators that have students in grades 7 through 12 in the same building or in separate buildings.

If more participants had been needed, emails would have been sent to those individuals who conduct classroom observations from the MASSP master list of schools, inviting them to participate in my qualitative research study. All participants who agreed to participate in my research study and committed to a specific interview date and time were sent a confirmation email or a confirmation phone call (Appendix B). The original number of invitations to potential participants was more than four times the number of those individuals who did participate in the research study. Initial invitations to individuals were extended via personal contact, email, or a phone conversation asking for a face-to-face interview, or an interview using an online application such as GoToMeeting or Zoom. An introduction of my research study and its purpose and benefits was discussed before any invitation was extended to potential participants. In the event of too few participants willing to participate in the research study, the recruitment

process would have started again with the initial list of over 150 school districts using the 5 Dimensions Plus Teacher Evaluation Rubric. Initial contact was made via email introducing myself and the research study I was conducting.

Saturation came when participant interviews continued to generate trends in the data that were very similar. Saturation became evident when analysis of the data collected showed evidence of nothing new being learned and therefore no additional themes are discovered (Patton, 2015). Patton (2015) stated that saturation implies that data collection and data analysis are happening simultaneously, and this analysis informs subsequent data collection decisions. To avoid prematurely reaching saturation, Patton suggests avoiding having a narrow sampling frame that has a skewed or limited perspective from participants, the data collection method is ineffective and does not follow a process, or the researcher is inexperienced and unable to get beyond the surface with participants. To avoid these shortcomings of sample size, I invited five to 10 additional participants under the understanding that they may not be needed to participate in the study. Bracketing my biases during the interview avoided having a limited perspective and conducting rehearsal interviews with a close colleague improved my interview skills, enabling the opportunity to collect more rich data.

During the recruitment process, the purpose and problem statement were described in detail, as well as how participating in my study would benefit all individuals who conduct classroom teacher observations and final evaluations. It was assumed that the participants would see the value of the study and use the results and conclusions of

this research study as an awareness of the beliefs and attitudes of colleagues towards teacher observation and the quality of training received in Michigan.

Instrumentation

The data collection process for my study was based upon methodology proposed by Patton (2015) for using interviews in a qualitative research study. Patton pointed out that skilled interviewing requires planning different kinds of questions, such as descriptive questions versus questions that require interpretations and judgments. When designing interview questions, I conducted a small number of practice interviews to measure and assess the interview process by monitoring the validity and reliability of the questions asked and the data gathered, ensuring the proper questions were being asked and the data was valid. I invited a small group of colleagues who I work with, but not in the same building or district, to assist me in rehearsing the questions. The process was assessed for its value concerning the perceived quality of observer training with the 5 Dimensions Plus Teacher Evaluation Rubric. I identified questions that may require a follow-up question to clarify or gather additional information. Also, listening for key words in responses that may have led the direction for the follow-up questions is an important part of the interview. In addition, the process involved distinguishing interviewee responses into categories such as behavioral, attitudinal, or knowledge-based (Patton, 2015). The interview is an interaction and a relationship wherein skills and experience affect the quality of responses (Patton, 2015). By conducting two practice interviews I gained experience at refining questions and digging deeper into conversations while gathering data. Competencies such as listening, using clear and

open-ended questions, using follow-up questions where appropriate, observing, knowing the types of questions to be asked, being prepared for the unexpected, and being present throughout the interview are essential characteristics according to Patton. It was agreed upon by the participant that the interview would be recorded to ensure the highest accuracy when transcribing the interview for coding purposes.

Unlike the informal conversational interview Patton (2015) describes, I used a semi structured interview approach, which assisted in keeping the interviewee on course. Using an interview protocol with specific open-ended questions offers the advantages of exact wording for each participant. All participants answered the same questions, increasing the comparability factor among participants, minimizing interviewer bias, and making the organization of data gathered much easier in the data analysis phase of the study (Patton, 2015). However, follow-up questions may have varied for participants, in order to reach a more in-depth response and gather rich data stemming from the original response. Once the exchange of dialog had ceased regarding the follow-up question or questions, I referred to the interview protocol and questions as planned.

During interviews, I observed gestures and body language of the interviewees, and used the memoing technique to scribe notes and information that I believed were important for the study. As mentioned, face-to-face interviews were not possible due to the pandemic. Observing body language was challenging since all interviews were conducted remotely online, limiting the view of the participants. Audio recordings were also used in all interviews to capture the narrative of each interview, which was later transcribed and stored for analysis. As stated earlier, Creswell (2009) suggested that the

researcher is the key instrument in a qualitative study. Using the interview process as the primary source, steps must be taken to assure the richest data was collected.

Researcher-Developed Instrument

Merriam (2009) suggested that when developing an interview protocol, it is important to avoid questions that may have multiple responses, leading questions, and closed-ended questions that limit the information gathered. The interview protocol and interview questions are listed in Appendix A and C of this research study. To this end, I created and used an interview protocol for asking questions and recording answers during the interviews. The protocol contained the date, location, interviewee name, employment position, school district type, years of observer experience, and instructions for the interviewer to follow so that standard procedure was used for each interview as recommended by Creswell (2009). The interview protocol was designed as a semi-structured interview that allows flexibility, which can lead to the possibility of gaining a richer insight into the training of observers while allowing for follow-up probing questions to gather a deeper perspective on the training (Merriam, 2009). Specifically, strategies for developing the protocol suggested by Merriam (2009) included a greeting; the purpose of the interview; confidentially statement; request from the participant to the interviewee allowing the interview to be recorded; an explanation that the participant could end the interview at any time if he or she chooses to do so; and avoiding questions that have multiple responses, leading questions, and closed-ended questions that tend to lead to yes-or-no responses. Open-ended questions were developed and modified as

needed to gather a sufficient amount of valid data. If follow-up questions were needed, those questions and answers were also scribed and recorded.

Questions for the interview instrument used in this my were developed using published dissertations that were similar to this research study, literature sources, studies, and published books. Many of the questions contained in the instrument were similar to those developed by Pisicotta (2014), who used an online survey with teachers in her dissertation regarding the teacher evaluation system, its effectiveness to increase teachers' effectiveness, and self-efficacy. I used those questions that related to observer training. Question foundations or stems were also used from James (2014), who used a combination of interview questions and survey questions for participants to gather data. Questions related to training and observer perceptions from the James study were used as a basis for developing selected questions. I ensured qualitative validity was established and maintained throughout this research study by continually checking for the accuracy of the findings by employing the following procedures: member checking to determine accuracy by taking back participant responses for review by participants to ensure accuracy, reflecting on any bias brought to the study by utilizing a critical friend/colleague to review and check the narratives as a result of the data collected, and presenting any negative or discrepant information that may not align with themes (Creswell, 2009).

I conducted two practice interviews to assess the interview process by monitoring the trustworthiness of the data gathered and ensuring the questions elicited data that could answer the research questions. I invited a small group of colleagues who I work with, not

in the same building, to participate in the rehearsal. The data were assessed for value concerning observer training with the 5 Dimensions Plus rubric. However, the data were not included in data analysis. The purpose of these interviews was to gain interviewing experience, assess the data gathered for appropriateness, practice the use of recording devices, and monitor the interview process as a working, flowing process.

Procedures for Recruitment, Participation, and Data Collection

The general procedures for recruitment and data collection were described in detail earlier in this chapter. Participants were identified using a list of school districts that are currently using the 5 Dimensions Plus Teacher Evaluation Rubric. A series of contacts with school principals or assistant principals who conduct teacher observations. The original research design was to conduct face-to-face interviews or utilize an online platform such as GoToMeeting. However, during the introductions and rapport-building phase, I discovered that all the participants were more familiar with the online platform Zoom. Therefore, I used Zoom to conduct the interviews and collect data during the months of June, July, and August of 2020. I followed up with participants by emailing them a copy of the transcript of their interview to review for accuracy regarding responses to the questions about observation training while also bringing closure to the participant's responsibility.

Each interview took between 30 and 40 minutes. Consent forms, purpose of the research study, and any other information was previously sent to all participants. It was much more important to be thorough rather than vague and non-direct when first starting the interview process to ensure a positive relationship was developed and rapport was

established. Interviewer neutrality was critical in how the interviewee responded and determined my ability to act in favor or disfavor of any responses displaying a behavior of trust and understanding (Patton, 2015). There was a possibility that a conversation may occur regarding the introduction and purpose of the study prior to the initial interview, which would decrease the interview time needed. This conversation would take place possibly over the phone during the initial planning of the interview meeting. This conversation would only take place after institutional review board approval has taken place.

Recording of Data

Data were recorded in audio and written formats. It was explained to each participant at the start of the interview that along with the recording devices used, notes would also be taken during the interview as a backup document in case of technology failure. An interview recording protocol was established for recording all interview conversations (Appendix C). This detailed protocol was created and followed for each interview conducted. Specific instructions were included in the protocol to avoid the interview getting off topic or on a subtopic tangent. Subquestions, probing questions, comments, and if necessary, follow-up questions were used, requiring participants to provide more detail and specifics regarding observer training. A paper-pencil template was created to manually record details on each question and follow-up question, with plenty of space to record notes. Written notes, emails, and recorded interviews were transcribed after each interview and stored in two separate areas. A flash drive was used to store interview data that will be in a lockbox in a personal office. Data is now stored

on my personal computer where all files are password protected and is secured in the researcher's home office.

Communications With Participants

A larger participant pool was established with the possibility of participants not being used in the study. Participant recruitment letters explaining the purpose of the research study were delivered to over 70 potential participants to ensure the desired sample size would be available to participate in the research study. This information was relayed to each possible participant at the outset of the recruitment process.

Communication with participants also included information regarding their decision to withdraw from the research study. Specifically, it was made known to each participant during the consent process that there would not be any negative consequences if they decided not to continue with the study. According to Creswell (2013), it is important to thank participants for their time and effort for participating in the research study. Exiting the study would come in the form of a letter sent via email to each participant acknowledging their participation in the study and notifying them that the data collection process has officially ended, and their participation has ended. Also included in the letter would be the notification of official completion date. At that time, they would be furnished with a copy of the dissertation.

Follow-Up Procedures and Debriefing

Upon completion of each interview, I briefly summarized each response to ensure what was recorded and documented was truly the intent of participants' responses (Creswell, 2013). A transcript of the interview was sent to each interviewee via email to

assure accurate recordings of each response to avoid misinterpreted answers. Upon the completion of the research study, participants received a final copy of the research study once published so as to not feel abandonment after participating in the study (Creswell, 2013).

Data Analysis Plan

All data were collected through the interview process for my qualitative research study. Except for demographic data, all interview questions, sub questions, or follow-up questions during the interview process connected to one of the two research questions stated in this study related to the perceptions of participants regarding classroom observer training, additional training, and their experiences conducting classroom observations. Throughout the data collection phase, the process of making sense out of the data, preparing the data for analysis, and moving deeper and deeper into understanding the data was iterative (Creswell, 2009).

Upon completion of each interview, I began the transcription process using an application which downloaded all recordings into text form. When considering data analysis and the potential for ethical issues to arise, Creswell (2009) suggested that researchers consider protecting the anonymity of the participants, the proper care of data and its storage or elimination, and the accurate account of the data collected from participants. To achieve this, I stored all data on a data disk and password-protected laptop device in a lock box in my home. Participants were assigned an ID number to ensure their anonymity and confidentiality of their responses.

Interview data were recorded using a remote recording device and downloaded, transcribed, and analyzed. During each interview, I engaged in handwritten memoing in which I noted gestures, topics that the participant seemed to be passionate about, and areas for further discussion and questioning. During this phase, I set aside and bracketed any personal bias that might have hindered any deeper dialog regarding the quality of training received or experiences of the participants.

Issues of Trustworthiness

The test for credibility of the findings occurs throughout data collection and data analysis. Procedures were set in place for gathering, storing, coding, interpreting, and reporting findings for the study. Creswell (2009) recommended using one or more strategies to check the accuracy of the findings. Using triangulation as an internal validity check was established by having multiple participants provide an abundance of information surrounding observer training throughout the data-gathering process and data analysis. By analyzing each participant response measured among other participant responses, I was able to build a coherent justification for themes and patterns among the data.

Using member checks, I also established credibility by electronically sending the findings along with a transcript of their interview, having participants review and determine the accuracy of the report, looking at themes and patterns (Creswell, 2009). I continually exercised self-conscious analytical scrutiny of my introspection to ensure maintaining reflexivity throughout this entire research study as prescribed by Patton (2015). Self-awareness, even a degree of self-analysis, has become a requirement for

qualitative research studies (Patton, 2015). Possessing a degree of reflexivity forces an awareness of potential biases, values, and experiences I brought to my research study (Creswell, 2013). Creswell also pointed out the reflexivity has two parts, each dealing with the personal experiences of the phenomenon as explored and how past experiences have shaped the interpretations of the phenomenon. During the data gathering process, I used a colleague to review the interview process, the data collected, and the analysis of the data for any irregularities or discrepancies in the data themes. This colleague was a well-respected administrator in our school district who was often sought out for district educational decisions. Rubin and Rubin (2012) suggested that when confirming data gathered from interviewees, the researcher should repeat or summarize what they thought they heard and the interviewee either agrees with the statement or modifies the statement. This strategy is important when the interviewee has said something that is different than the perspective of the interviewer or when the interviewer needs to be assured, they have understood correctly (Rubin & Rubin, 2012).

External transferability threats arise when researchers draw incorrect inferences from the sample data to other persons, other settings, or past or future situations (Creswell, 2009). Having a purposeful representative participant sample of urban, rural, and suburban school districts as well as participants with varying observer experience ensures transferability. Ensuring external validity, a participant sample set was established that was diverse in demographics and years of observer experience that led to the transferability of the results from this study. Creswell (2009) suggested researchers identify the potential threats that may arise in the study, define the exact type of threat

and the potential issue to the research study, and discuss plans of how to address the threat. I kept a daily research journal along with memoing throughout the research process that tracked all decisions and processes, providing the documentation necessary for transparency and dependability.

Ethical Procedures

Ethical issues can emerge throughout the entirety of a qualitative study and researchers need to consider what ethical issues might arise during a study and develop a plan to address these issues (Creswell, 2013). Having a frame of mind that everything was not going to move along and having an alternative plan for issues that arise helped provide the flexibility needed to proceed when issues did arise. Prior to conducting any part of the study, I sought approval from the Walden University Institutional Review Board. Once approval was granted (approval numbers 05-04-20-0242427), all participants had full disclosure of the purpose of the study and signed off on their commitment and participation in the study. A consent form was created for participants to sign before they participated in the research. Informational consent forms had all pertinent information regarding withdrawing from the study, protection of confidentiality of participants, known risks of participating in the study, and signatures of participants and the researcher along with the acknowledgement that participants' rights were not violated during the data collection process as proposed by Creswell (2009). The consent form had the critical elements of the selection process, identification of researcher, identification of the purpose and benefits of the study, level of involvement by participants, identification of any risks and withdraw procedures, and contact information

if the participant has any questions throughout the study. All participants were given a cover letter containing information about the research study (Appendix B).

I took into consideration the ethical obligations of the role of the researcher throughout the study. All recruitment documents mentioned above were provided and were used during the initial stages of recruiting. All recruitment procedures were documented, continuing the process audit trail to ensure all ethical standards were met. As with all data collected, recruitment information is password protected and kept on my computer in my personal home office.

When collecting data, I respected all protocols of each school building and school district. I sought not to disrupt the flow of the day by recognizing and respecting the time commitments of participants (Creswell, 2009). Interview questions (Appendix A) and an interview recording protocol (Appendix C) were used with every participant to ensure a consistent ethical procedure was followed. As mentioned above, a consent form also provided all details for participating and if necessary, withdrawing from the study. A process of full disclosure was used with all participants, answering all questions regarding data and procedures in my research study as recommended by Creswell (2013). The data analysis and reporting started with full disclosure of results while respecting the privacy of the participants (Creswell, 2009). Interview and all other data, including notes, journal entries, memos, and digital recordings, have been kept in locked cabinets in a locked office at my residence. All electronic transcripts are being kept in password-protected files, to which only I have access.

In concluding this section on ethical considerations, it was important to emphasize that I needed to protect their research participants and develop trust while promoting the integrity of the research study (Creswell, 2009). Ethical practices involve much more than following a set of published guidelines. I needed to anticipate ethical issues and actively address them in the research plan (Creswell, 2009). To this end, I conveyed the purpose of the study to all participants to avoid any deception issues (Creswell, 2009) regarding the observer training and the 5 Dimensions Plus Teacher Evaluation Rubric. During each interview I kept an awareness, if for some reason the interviewee appeared to be getting upset or emotional about a question or about the 5 Dimensions Plus rubric, and I would redirect the questioning to a more comfortable, less stressful set of questions while continuing to gather rich data. Another issue regarding the treatment of participants was that I assured that all participants' identities remained confidential. As Creswell (2009) pointed out, this allows the participants to retain ownership of their voices and may generate unfiltered data.

Summary

The purpose of this study was to gather data and report findings regarding available training and its effectiveness for those individuals conducting classroom teacher observations and final evaluations. I utilized a basic qualitative research design, using interviews as the data source. The design selection was based upon the need to gather information on the experiences and perceptions of those individuals conducting classroom observations and final evaluations surrounding observer training. In chapter 4, I presented and discussed the results of this study. I described the setting the study took

place along with the participant demographics. Tables and charts are used to describe and present the findings, specific results, the categories, and themes that resulted from participant responses surrounding training needs and future training needs of each participant.

Chapter 4: Results

The purpose of this qualitative research study was to explore the perceptions of administrators regarding their preparedness to evaluate and observe classroom teachers and the quality of their teacher observer training. All participants in this study were administrators of Grades 7-12 in public schools and used the state-mandated 5 Dimension's Plus Teacher Evaluation Rubric. Interviews were conducted via an online platform with participants, which resulted in identifying data and the final findings.

The interview questions (Appendix A) were designed to gather data that led to further insight into the perceptions of quality training provided to observers who are using the 5 Dimensions Plus Teacher Evaluation Rubric. As a researcher, I needed to have the patience and objectivity to engage in a complex process of data analysis through the ambitious task of sorting large amounts of data while reducing the data into themes or categories.

I gathered data that identified specific training practices that may potentially raise the capacity of observers to differentiate effective levels in teaching. The research questions focused on the observation process and the quality of training received by individuals who conduct classroom observations and final evaluations for teachers in the state of Michigan. Focusing on the training opportunities available, the strategies used in training, and the perception and preparedness of observers, data collected and analyzed led to results that could be used to improve the training procedures and processes currently used in observer training.

This chapter provides a brief description of the study participants' demographics along with an extensive review of each interview and the themes that evolved from the interview transcripts of all 13 participants. Following sections in this chapter describe the processes that was taken, starting with the initial practice sessions to conduct each interview, and finishing with the analysis of the data and producing the final results. Because face-to-face interviews were not possible due to COVID restrictions during the pandemic, I used Zoom, a video-conferencing platform (Zoom Technologies, 2011). The electronic platform Zoom was used based on the amount of experience and familiarity I had working the Zoom application while working with colleagues during professional meetings versus the other platforms. A description of the interactive, open-ended interview method applied in this study that was used due the changing circumstances surrounding the worldwide pandemic follows. The process of collecting, recording, and interpreting the data was the result of numerous immersions into the data and a thorough analysis as described by Moustakas (1994). The final sections in this chapter provide a description of the data analysis processes, including the trustworthiness used to ensure the level of validity and credibility exercised during this research study.

Field Test

While designing and planning the interview and the questions, I conducted two practice interviews with colleagues in a face-to-face situation to fine-tune the interview process and monitor the trustworthiness of the interview questions. The practice interviews took place in the offices of each principal, where a recording device was used as well as written notes taken from the interviews. Communications were sent to the two

field test participants via email regarding the description of the study along with the consent form that needed to be signed and returned. I asked the original 20 questions that were answered by the participants. The field test was a valuable source for making slight adjustments in the questioning techniques including the delivery of questions, selection of probing question, and question design. A follow-up question was modified slightly to enhance the likelihood of eliciting answers about any changes participants might suggest for future training. Also, a modification was made to a question regarding the most effective training received by participants. Data collected in the field test were not included in the results of the study.

Setting

The original plan for this qualitative research study was to conduct face-to-face interviews in order to collect data. In March of 2020, the state of Michigan and the United States were in the beginning stages of a pandemic. In the following weeks, the state of Michigan education system followed government orders to shut down schools and businesses to prevent the spread of the COVID-19 virus. The COVID-19 virus and the declared government responses to the pandemic had a major impact on daily lives, including the strenuous process of completing a dissertation. The pandemic changed the way citizens conducted their daily business and altered the way I would be able to conduct and complete all the interviews and data collection process for this study. Restrictions and safety protocols were placed on communities and school systems in the state of Michigan during the time the interviews were scheduled to take place. Citizens

were encouraged to stay at home to avoid spreading or contracting the COVID-19 virus and school buildings were shut down while teachers delivered instruction virtually.

Adhering to the COVID-19 restrictions placed on communities and schools, an alternative plan was created and implemented to conduct the interviews to complete the study. A small number of learning platforms were considered to conduct the virtual face-to-face live interviews with participants. Zoom allowed for a virtual face-to-face interview to take place that could also be recorded, which then could be used to generate a hard copy of each interview transcript. Invitation letters for each Zoom interview were sent via email as previously planned, schedules were followed, and procedures were followed with the intention that all interviews would be scheduled and taking place virtually using the Zoom electronic application. Participant responses were favorable, and interviews were scheduled. Since participants were no longer required to attend to a full day of working in their schools and, in some cases, ultimately working from home, all participants were familiar with the Zoom application and agreed to an interview through this online platform. I believe this made securing a 35- to 45-minute interview much easier because of the convenience of the interview process. The population from which participants were purposively selected were those individuals who were conducting classroom teacher observations and final evaluations. These individuals were currently employed in a public school system that was using the 5 Dimensions Plus Teacher Evaluation Rubric in Grades 7-12. The participants were selected because they best informed the study about the effectiveness of observer training (see Creswell, 2013). The primary source for data gathering was semi structured interviews with the participants.

Demographics

I used purposive sampling to identify potential participants. The criteria for participation included being a current principal or assistant principal of Grades 7- 12, working in a public school district and using the 5 Dimensions Plus Teacher Evaluation Rubric.

As shown in Table 1, the 13 participants came from rural, urban, and urban/suburban school districts from across the state of Michigan, with four participants from a rural, four from an urban/suburban, and five from an urban district.

Table 1

Participant Demographics

Participants	Years' Experience Observing	Grade Level	Setting
Participant 1	6	7–8	Rural
Participant 2	2	7–8	Urban
Participant 3	10	9–12	Urban
Participant 4	3	9–12	Rural
Participant 5	8	9–12	Urban
Participant 6	8	9–12	Urban
Participant 7	4	9–12	Urban
Participant 8	6	9–12	Urban/Suburban
Participant 9	10	7–12	Rural
Participant 10	6	7–8	Rural
Participant 11	7.5	7–8	Urban/Suburban
Participant 12	7	7–8	Urban/Suburban
Participant 13	2	7–8	Urban/Suburban

Note: An **urban** area, or built-up area, is a human settlement with a high population density and infrastructure of built environment. R**ural** to include all people, housing, and territory that are not within an urban area. **Suburban** areas are lower density **areas** that separate residential and commercial **areas** from one another.

The participants ranged from 2 to 10 years of observer experience. The average years of observing teachers among the participants was 6 years. Schools or contracted personnel who conduct classroom teacher observations and final evaluations and use Danielson's evaluation rubric, Marzano's evaluation rubric, or any other professionally created or a locally developed state of Michigan-approved teacher observation rubric were excluded from this study.

Data Collection

The data collection process included setting the boundaries for the study, collecting the information through semi-structured interviews, and establishing a protocol for recording the information gathered (see Creswell, 2009). As discussed earlier, I used a purposive sampling to identify potential participants who met criteria to ensure that the participants could best answer the research questions (see Creswell, 2013). Using multiple lists from Michigan Department of Education, the Center for Educational Performance and Information, and school demographics for the state of Michigan, I created master lists for rural, urban, and urban/suburban school districts in the state of Michigan and matched these with published lists of schools and school districts that were using the 5 Dimensions Plus Teacher Evaluation Rubric to generate a list of school districts that were currently using the rubric. I identified 70 potential participants and sent invitations via email to all of them, as previously planned.

The data collection process required alternate collection methods to be put in place due to COVID-19 and the restrictions and stay-at-home orders put in place by the Michigan Department of Health and Human Services and the governor in the state of

Michigan for businesses, schools, and the citizens. It was evident that the concept of meeting with participants in a face-to-face format was impossible and at best would make both the interviewer and interviewee feel uncomfortable. All interview schedules were then altered and changed to use an online Zoom platform. Using an online platform limited the interview atmosphere, which in turn presented difficulties in building a trusting relationship while also making it difficult to read body language.

Original interviews were scheduled to take place at a location of the interviewee's choosing for their convenience. However, as a result of the ever-changing environment brought on by the COVID-19 world pandemic, all interviews needed to be changed and took place in an alternate setting other than a true face-to-face meeting. The interviews for this research study took place over 4 weeks from May through June of 2020. Since face-to-face interviews were no longer possible, interviews were scheduled at the participant's convenience with the determination of the location and time of interview that best fit their schedules. The researcher conducted the interviews at the participants convenience either at their home or school office. All interviews took place during the school day, with the exception of one interview which took place in the late afternoon. All participants were asked and agreed to the interview being recorded and that only the narrative of the interview would be used for the research study. Establishing rapport with the participants seemed more comfortable in the commonality of all educators facing new challenges to deliver the curriculum and meet the students' academic and social and emotional needs while conducting daily work responsibilities virtually.

During the interviews, I also took notes and memoed any narrative from the participants that may have been of interest throughout the interview. Upon completion of each interview, all recordings were sent to a transcriptionist for a final electronic and hard copy to be used in the analysis process. The transcription processing time was 1-2 days for each interview.

There were 13 participants in this study. There were 70 invitations sent to identify potential participants via email, resulting in an 18.5% response rate. In the initial email I introduced himself as a Walden University doctoral candidate conducting research and letting them know the letter was seeking participants to interview regarding observation training. Included in the email, I also attached the invitation letter and consent form. Participants were asked to contact the researcher if there were any questions regarding participating in the study. No phone conversations took place regarding questions of the study. Final participants returned an email to the researcher with a signed consent form.

Interviews were intended to be conducted with a minimum of 12–15 identified participants who are currently using the 5 Dimensions Plus Teacher Evaluation Rubric. To have the best representation of demographics, a target participant pool of four to five participants were planned to be selected from each type of school district, rural, urban, and suburban. The final 13 participants brought to the study a variety and range of administrative experience.

Two recording devices were used for all interviews, along with using pen and paper to take notes during each interview. A small, handheld, manually operated recorder was placed near the laptop speaker during all interviews. The second device was the

recording option included in the Zoom application. As the host of the Zoom meeting, I controlled the start and finish of the recordings by clicking the button on the screen to activate the recording option. During the interview a bracketing technique was used to control any researcher biases I may have, allowing the interviewee to expand on any subject while continuing to probe in a direction with follow-up questions.

I followed Rubin and Rubin's (2012) model for conducting interviews, which involved three stages in the interview process: observe the body language of participants through the monitor, use memoing to scribe notes and information important for the study, and use audio recordings that will be transcribed and stored for analysis.

To assure the accuracy of the transcriptions to reflect the perceptions of the participants, I used member checking. I sent a copy of each participant's transcript via email to the participants and asked them to check for accuracy and indicate if any additions or corrections were needed. Nine responded and none of them offered corrections or additions.

Data Analysis

Coding and Analysis

Data analysis was an ongoing process involving continual reflection about the data, and was conducted concurrently with gathering data, making interpretations, and writing reports (Creswell, 2009). I followed the six steps for analyzing data suggested by Creswell (2009), working through organizing and preparing the data by interpreting the meaning of themes and descriptions. The duration of the data collection process, analysis, and preparing the data for presentation, took up to 4-6 months. Steps 1 and 2 involved

organizing and preparing the data for analysis and making sure all data was read. This involved the transcription of all data, both written and recorded. I sent the interview recordings to a professional transcriber ensuring accuracy in the participants reporting. Once the interview data were transcribed, I uploaded the data into NVivo12 which generated the reports I used to create codes, categories, and themes. I reviewed the finished reports for similarities, commonalities, and patterns. I began to code each interview response according to the similarities or differences in the responses. During Steps 3 and 4, I continually reviewed the codes assigned to each response. I then started to identify potential categories and themes that would be based on the codes assigned to each response. Themes and categories were considered throughout the interview process to develop an understanding of patterns along with identifying outlying or different information during the data collection process. Steps 5 and 6 involved making decisions on how the descriptions and themes were presented and final interpretation of the data (Creswell, 2009). Themes emerged from the frequency of each word or phrase from the participants responses by placing like responses together and creating themes. Table 2 displays the categories that emerged from the assigned codes.

Table 2

Categories

Code Name	References
Demographics	15
Challenges	42
5 Dimensions Plus Rubric	158
Future training needs	165
Training	137
Feedback	75
Observations	163

Table 3 identifies the themes that emerged from the data analysis. Two themes were identified for final reporting. However, three groups were used to generate the theme preferred training.

Table 3

Themes

Codes	References
Future Training Needs	105
Networking with Colleagues	13
Practical training then debriefing with colleagues	7
Refresher course	12

The qualitative software NVivo12 enabled me to manipulate the data manually or use auto code for coding purposes. I used the visuals to continually analyze and question the data gathered to create the final analysis. Table 4 displays the connection between the research questions, interview questions, and the constructs of social cognitive theory, the theoretical framework for this qualitative research study.

Table 4

Research Question Connections to Interview Questions and Constructs of SCT

Research Questions	Interview Questions	SCT #1	SCT #2	SCT #3	SCT #4	SCT #5	SCT #6
#2	4, 7, 8, 9, 14, 15, 16, 18, 19, 20	X X X	X	X	X	X	X
#1	6, 7, 9, 17			X	X		X

Table 5 displays the connections between the interview questions and the SCT constructs that were used in this study.

Table 5

Interview Question Connection with Constructs of SCT

Interview Questions	SCT #1	SCT #2	SCT #3	SCT #4	SCT #5	SCT #6
	4, 6, 11, 12, 14, 15, 16, 18, 20	7, 10, 11, 13, 14, 15, 16, 18	11, 12, 13, 14, 15, 16, 17, 18, 20	8, 11, 12, 13, 14, 16, 17, 18, 20	9, 11, 12, 14, 20	10,11, 12, 13, 14, 15, 17, 18, 20

During the data analysis process, I established and maintained transparency by documenting all possible codes, categories, and themes. I created a codebook to use as documentation of possible codes throughout the coding process. I used memoing as a

technique to write down any emerging ideas throughout the data collection and analysis process as prescribed by Creswell (2013). As Maxwell (2013) suggested, memoing is one of the most important techniques a researcher has for developing ideas. Maxwell also suggests that memoing is a way to understand the topic by engaging in serious reflection and analysis, leading to analytic insights that can be kept chronologically throughout the study and made available for reporting. I ensured ethical procedures were taken, requiring all data be reported regardless if the data aligned with limited themes or matched with no other themes or trends (Creswell, 2013). To ensure transparency and accuracy with all participants, interview transcripts were made available to check all data was transcribed as originally gathered. All data gathered, data analysis, finding and conclusions which lend credibility and validity to this study making it more realistic due to the thoroughness of the data gathering and analysis (Creswell, 2009). Assigning an participant number to each participant (Creswell, 2013) assured that reports did not disclose information that would be damaging to any school district or building administrator.

If for any reason a participant decided to withdraw from the research study for conflicting interests or disagreement with the researcher, the situation was discussed and the process for withdrawing from the study was reviewed. All accommodations were exercised to ensure any participant who wished to withdraw from the study could do so with no expectations or limitations. I ensured any data gathered prior to a participant withdrawing from the research study remained confidential. All responses were considered, and no discrepant responses were noted. Participants had the opportunity to voice and document any disagreements or issues, and beliefs for leaving the research

study for any reason. There were no participants who wished to leave in this research study, and all participants started and completed their commitment to being a part of the study.

I organized and prepared the data for analysis. I created folders that were used to store all information during the data collection and analysis process for each participant. Upon receiving each interview transcript from the transcriptionists and printed hard copies for each transcript, I then read over each transcript along with listening to the recordings to identify those potential codes that quickly emerge more frequent than others. The initial reading was conducted with no memoing or annotations performed to avoid any bias. Included in step 1 was determining any obvious patterns or potential categories evident in the interview responses to direct and guide the coding process (Maxwell, 2013). During this time a codebook was created and used throughout the data analysis process. The codebook was used to record potential categories, patterns, and store short statements, words and phrases, themes, patterns, and potential codes that were ultimately used in the final data analysis process (Appendix D). I used the codebook as a reference throughout the analysis process, I also created a document containing each individual question response for each participant. This allowed for a deeper dive into participants' response to each question by identifying similarities and differences thus creating patterns and potential categories and themes.

After reading the transcripts, step two involved uploading the transcripts into a software program called NVivo12. NVivo12 software program facilitates analysis of qualitative data by creating queries, word frequency searches, and manipulation of the

data. Using the identified potential patterns, notes, annotations, words, highlights from the transcripts, possible themes, along with the codebook notes, were entered into NVivo12. Word and phrase frequency were identified from the participant responses to identify potential themes. Each transcript was read numerous times allowing for the identification of codes. Using the software each transcript was reviewed and codes were identified based on frequency use in the interview responses. For example, the code future training needs was developed by participant numbers eight and eleven stating the need for additional training as noted in Figure 1 and the expression of training needs during responses. Figure 2 lists the methods that training should take place as referenced by number of participants. Participant numbers 8, 5, and 11 all noted issues with the training such as the length of training, the complexity of the training, and the amount of material involved in the training. Participants one, 11, and five also referred to working with colleagues during training sessions as beneficial. Working with colleagues was the most impactful part of the training, as noted by participant number 8, who stated, “Being able to practice the process and then talk about it was probably the most impactful.” Participant number 11stated the need for networking by saying, “I like talking to other educators just to talk about what works, what they’ve tried, and what suggestions that they have or, stay away from this type of thing.” A desire to stay in communication with colleagues was evident in the findings. The code working with colleagues was created by the responses by participant numbers 8 and eleven stating the calibrating, like the watching the same clip and us comparing what we saw and what we coded, is always good, I think. ‘That it’s good practice.”

Patterns and relationships among categories were used to identify the themes by matching similar terms. The main categorizing strategy in qualitative research is coding (Maxwell, 2013). Over 80 codes (Appendix D) were originally identified during the data analysis. I identified several categories from which the two major themes emerged. Specific codes consisted of the 5 Dimensions Plus Teacher Evaluation Rubric, coding, scripting, training, future training, observations, and time. Categories identified throughout the analysis process included extensive training, extensive time for implementation, challenges of coding and scripting, engaging in difficult conversations with teachers, the 5 Dimensions Plus Teacher Evaluation Rubric as a growth tool, and strengths and weaknesses of the rubric.

Data analysis included all participant transcripts. During the reporting of the final results and findings it was found that all participants were very interested in participating in my research study and were open to follow-up participation if necessary; however, this was not necessary in this research study.

Evidence of Trustworthiness

To assure the trustworthiness of data collection, data analysis, and interpretations, I addressed issues of credibility, transferability, dependability, and confirmability (Korstjens & Moser, 2018).

Credibility refers to the plausibility of findings from the viewpoint of participants. Credibility was established using NVivo12, where the original transcripts were uploaded and thoroughly reviewed for patterns and themes, ensuring credibility. The findings represented plausible information drawn from the participants' data and are a correct

interpretation of the participants' original views (Korstjens & Moser, 2018). Reflexivity was evident through maintaining a high degree of self-awareness, continually having a conscious analysis of my own perspectives, and recording in a journal any thoughts that resonated with me in the interview process (Patton, 2015). Having a degree of reflexivity forces the awareness of potential biases, values, experiences, processes, and procedures within the research study are exercised ensuring credibility (Creswell, 2013).

I used follow-up questions to confirm and dive deeper into participants perspectives. By creating an atmosphere of respect and develop positive rapport with the participants credibility was established. This was done by having small talk surrounding our commonalities in our workload and profession. Credibility happens by having confidence in the truth of the findings including an accurate understanding of the context and is commonly used in qualitative research studies in place of the word validity (Guest et al., 2012). Credibility of the findings occurred throughout the data collection and the data analysis process. It was evident in the like responses from participants numbers eight and 11 that the training received and how it was delivered rose as a common theme. The need for additional training was a frequent response by participants for credibility. Frequency of common language, words, and phrases from the participants reached the point of saturation and also enhanced credibility. It was evident that saturation was occurring by the responses from the participants that surrounded the two themes. Participants were electronically emailed their transcripts to check for accuracy and add any additional information that was not recorded. Of the thirteen participants in this

study, I received feedback from eight of the participants confirming the accuracy of the responses.

Transferability refers to the degree to which findings can be claimed to be applicable to other participants and contexts. I selected participants that represented high school, and middle school levels, and those with a demographic of rural, suburban, and urban that allowed for a greater possibility for others to replicate this study.

External transferability threats arise when the degree to which findings can be claimed to be applicable to other contexts. Having a purposeful representative participant sample of urban, rural, and suburban school districts as well as participants with varying observer experience ensures transferability. However, transferability occurs based on those in other similar settings considered the findings to be applicable to theirs. A participant sample set was established that was diverse in demographics and number of years of observer experience, ensuring the transferability of the results. The high degree of transferability in my research study depended upon other contexts or settings and respondents and was a result of the participant sample and process followed (Korstjens and Moser, 2018). I kept a research journal memoing throughout the research process to track all decisions and processes, providing the documentation necessary for transparency and dependability to address any threats to transferability. A thorough description of the participants and the research process mentioned earlier ensures the findings are transferable to other settings and further research (Korstjens & Moser, 2018).

The protocols of data analysis were thoroughly followed for each participant, ensuring the stability of findings over time (Korstjens & Moser, 2018). Exercising

transparency by describing the data analysis steps taken from the start to the development and reporting of the findings (Korstjens & Moser, 2018) was evident and ensured dependability in this research study. The analysis process was in line with the accepted standards for this design, ensuring a high degree of dependability. Dependability is also evident when other researchers can repeat the study and arrive at the same findings with the same data. Dependability refers to whether the research process is consistent and can be carried out with careful attention to the rules and conventions of qualitative methodology (Guest et al., 2012) and the likelihood that other researchers could repeat the study and would arrive at the same findings with the same data.

Confirmability refers to the extent to which results can be corroborated. Guest et al. (2012) suggested that to confirm data gathered from interviewees, the researcher should repeat or summarize what they thought they heard and the interviewee either agrees with the statement or modifies the statement. This was evident in multiple readings of the transcripts where the participant responses were often repeated to confirm their answers to questions. Korstjens & Moser (2018) refer to confirmability as the neutrality of the research study. Confirmability was established by the corroboration of the findings.

Results

The two research questions in this study focused on the additional training observers need for performing effective classroom observations and the perceptions observers have about the training they have received for performing classroom observations. Each research question was addressed in the interview questions (Appendix

A). Social cognitive theory's constructs were used as a lens for the data analysis. Training involves the acquisition of new learning and behaviors. three of the six SCT constructs were used in the data analysis process specifically regarding training. Self-efficacy, behavioral capability, and observational learning all relate to a learned behavior and performing that behavior specifically to observation training. Four of the interview questions addressed the need for future training while 10 of the interview questions directly addressed the quality of training received. The findings from the data analysis discovered two themes, with each having multiple subthemes. The first theme identified many participants needing additional training in a number of areas. Subthemes identified desired areas and the methods by which the training should be delivered. Regarding observer training, participant number eight stated, "I think there is a need for it,". The second theme identified the quality of training received by participants and the need for training including the methods used to deliver the training. Subthemes surrounded the variety of training delivered.

Theme 1: Future Training

The first theme revealed that all participants desired future ongoing training and identified how that training should be delivered. Participant number 11 stated, "I think there is always a need for it," when referring to future training. The emergence of subthemes was evident when five areas of future training were identified as well as the methods in which that training should be delivered. Figure 1 displays the five areas identified for future training.

Figure 1

Future Training Needs

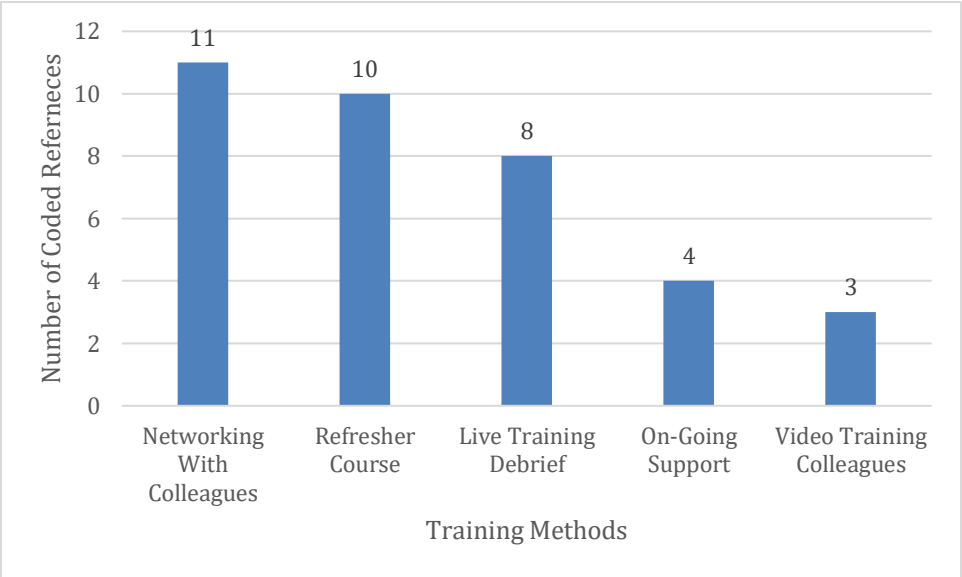


Figure 1 also displays the number of participant references regarding the types of future training and the methods of training delivery. Collaboration with colleagues, a continuous offering of refresher courses, and the opportunity to network with colleagues regarding the observation process were areas for future training. The future training areas can be distilled into two subthemes: the areas most desired by participants for observation training followed by ongoing training support in the form of refresher courses.

Training Perceptions

Although not displayed in table 3, was the perceptions of 8 of the participants describing how the 5-day initial training for the 5 Dimension Teacher Evaluation Rubric they received was either overwhelming, extensive, or intensive was. One participant described it as “very thorough. I mean, I learned a lot; it was really long, though.” It was also discovered the participants thought the many steps involved to learn the entire rubric

created an overwhelming feeling, especially for a new person. Regarding the training for the 5 Dimensions Plus Teacher Evaluation Rubric, participant number five stated, "5D is not something you can sit and get and understand it because you have to be able to work through it." Participant number thirteen described the training as "really good and really intensive." Participant number seven stated "I was overwhelmed" when describing their observation training. Also mentioned about the training received and was described by participant number eleven as, "There was way too much lead-up, I thought, versus the actually doing it and trying and see how we did." Working with colleagues was the most impactful part of the training, as noted by participant number 8, who stated, "Being able to practice the process and then talk about it was probably the most impactful." The constructs of behavioral capability and observational learning were the lens used to analyze the training perceptions data.

Professional Networking

A subtheme expressed by eight of the participants was that networking with colleagues regarding the observation process as a type of training that participants preferred. Communicating and sharing experiences and processes with other observers who have experience observing, was a method that participants chose for training. Looking through SCT's self-efficacy lens, participant number one referred to the interview process that took place for this research study where individuals were engaged in talking about observations as, "would be really, really powerful for evaluators in the field to have discussions about how they do the process." Participant number eleven stated the need for networking by saying, "I like talking to other educators just to talk

about what works, what they've tried, and what suggestions that they have or, stay away from this type of thing." A desire to stay in communication with colleagues was evident in the findings.

Collaborating and debriefing with colleagues regarding the coding process after the viewing of a short classroom teaching video was a desired training method by participants. Participant number nine stated, "The calibrating, like the watching the same clip and us comparing what we saw and what we coded, is always good, I think. That it's good practice." It was also suggested that with the emergence of live face-to-face technology platforms, individuals could watch videos and then, via a platform such as Zoom, collaborate regarding observations. Participant four suggested that multiple observers watch the same video clip from their own buildings and conduct the process of the observation and then compare how everyone coded the teacher in the video. Participant number 4 stated, "All of that can be done from I'm sitting right here in my office."

Also, conducting live classroom observations with colleagues or individually and then gathering as a group to compare and debrief about the coding process was referenced often by participants. Participant number five suggested that observers in groups of two or three go out and do an actual live observation and regather so that everyone can say, "Hey look, this is why I focused on this, or this is what I focused on." Participant number 9 stated, "The opportunity to be able to do a little bit of work with other administrators and other people in education, just be able to get some of their vantage points." Networking with others was emphasized with participant number eleven

who stated. “Well, I think I'm always looking for other ideas. I mean, like what you said, I like talking to other educators just to talk about what works, what they’ve tried and what suggestions that they have or, stay away from this type thing.” Participant number 5 stated “I think the more opportunities that you can get people to live within it and then, come back and have conversations about it and debrief about it and reflect on it, the more quality conversations, and the better ownership that people with have about 5D.” Participant number eleven stated, “Well, I think I’m always looking for other ideas

Live Debrief with Colleagues

Live debriefing with colleagues and training emerged as a subtheme. This involves actively observing with others in a classroom and then debriefing outside the classroom immediately following the observation. SCT’s 3 constructs of self-efficacy, behavioral capability, and observational learning were used as the lens to analyze participant responses regarding live debrief with colleagues. Conducting professional development on a platform such as Zoom or Google Meet was mentioned specifically as it can be done without physically leaving the building. This, of course, involves viewing a video of a classroom and then scripting and coding the video. Participant number seven suggested the idea of being placed in training cohorts of observers and utilizing the technology to connect and communicate with colleagues, stating, “Especially with the use of Zoom and Google Meet now, I think there should be some kind of mentorship.” Added participant number thirteen, “I guess even just a refresher even every year or every other year, even if it was optional, just even like a Zoom meeting or something, that would be beneficial.”

Utilizing an online platform to collaborate with colleagues was another form of professional development along with the variety of content that can be the focus of the meeting. It was suggested that community refreshers would be an idea. Participant number two stated, “We’re hosting this forum for admin to come in and just share.” Participant number four stated, “I guess coming out of the situation we are with all these lessons online, I mean, why couldn’t our training continue where, OK, here’s a short five-minute video clip.” Participant number four also stated, “Everybody scripts it, code it, let’s throw it out there, and we’ll see where we all land.”

Refresher Training

Another subtheme that emerged from the data analysis was the need for ongoing training or a refresher course for observation, as evidenced by participant number two stating, “I think refreshers would be huge.” “I think there’s always a need for it” was the message from participant number 8. Participant eight also referred to ongoing training as, “It’ll definitely be continuous or needs to be, and I’m sure of that.” The findings have shown that ongoing refresher training is wanted, and that training may come virtually via an online platform. “This would be difficult for districts to do, but if there was even a refresher once a year, on just how to do it,” stated participant number thirteen. “I do think you need to go every couple of years and kind of have a refresher to make sure you haven’t gotten off track and all of a sudden interpret it,” stated participant number 10. Participant 10 also stated “I think some refresher would go.”

When asked about refresher training, participant number nine stated, “I think it’s good going back to get some . . . refreshers and retraining.” Gaining a different

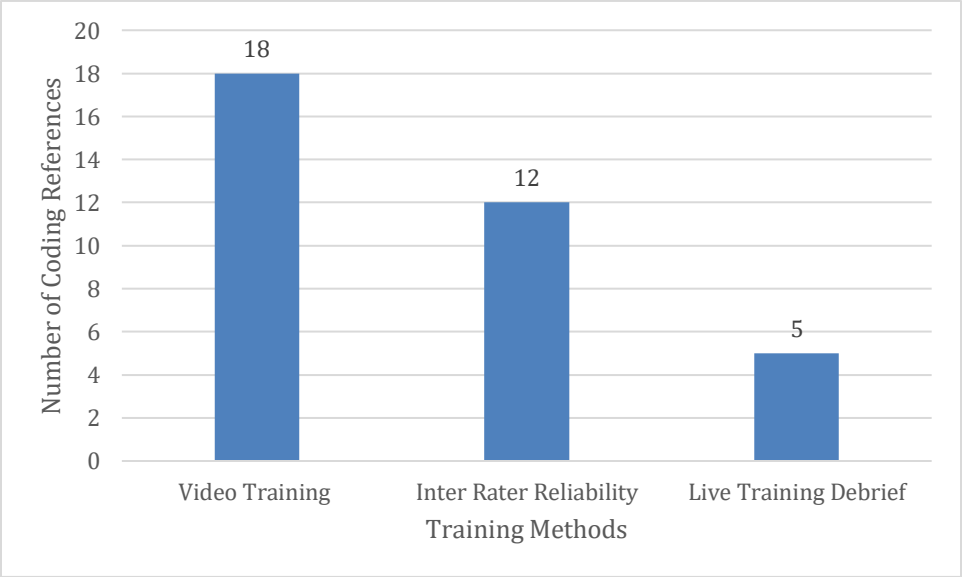
perspective on the observation process was the motive for participant number nine, who also stated, “I think there’s sometimes that it’s just a matter of me going through and doing a refresher or getting a different perspective on it that after five years things changed in my philosophy, and in some different interpretations and stuff like that.” All participant responses were analyzed through the SCT lens of behavioral capability and observational learning.

Theme 2: Training Needs

Figure 2 displays the three training methos identified through the data analysis of the participants’ responses as the preferred method of training delivery. Three subthemes emerged from the data analysis under training needs.

Figure 2

Methods of Training



Video Training

The main subtheme theme that emerged from the quality of training came through the lens of behavioral capability and observational learning of the SCT constructs, although only four of the 13 participants mentioned this type of training method, the training which involved watching videos of teachers teaching in their classrooms and practicing the observation process. This method had the most references. Participant number ten thought "it would be nice just to be able practice in a live classroom, but they had what was second best, which was viewing teaching video clips of teachers." Participant number eight stated, "I think the best parts of that training were, one, they would play a lesson and then as a group, we would kind of just practice it." Training as a group and watching the videos followed by comparing the observation practices such as scripting and coding among training participants was an effective practice and was desired for future training as well. Participant number four, recalled the trainers saying, "Hey, watch this video and we'll kind of jump in and we'll all script and code it and see where you're at and see where I'm at." Participants recognized the importance of viewing classroom teacher clips and practicing the observation process, which was evident in participant eleven stating, "When it finally came to the part where we were able to script and code and then we would swap with other people, we would all see the same thing. And then we would swap how we coded it." After viewing short video clips of classroom teachers, participants favored the collaboration among colleagues as a desired training practice. Participant number six stated, "Being able to practice the process and then talk about it was probably the most impactful."

Interrater Reliability

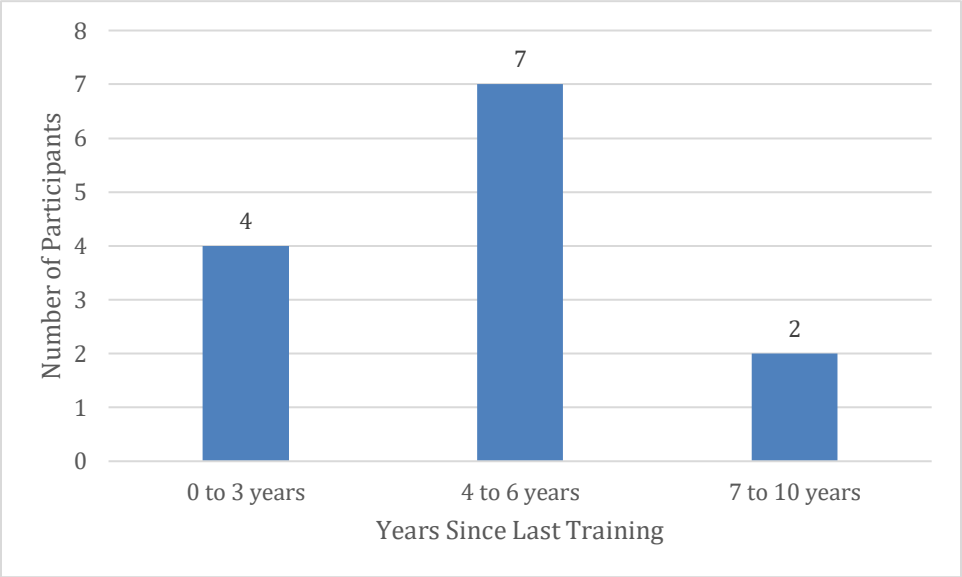
One of the subthemes was the rater reliability or the interrater reliability training process. This occurs when observers can consistently rate teachers and properly code teachers' actions in comparison to the University of Washington's Center for Educational Leadership's established ratings. During the video clip, trainees' script what they observe and then code what they observed. The codes are compared to the codes of the certified raters from the Center for Educational Leadership at the University of Washington. If the codes matched those of the raters from the Center for Educational Leadership, the observer was then considered to be rater certified. This task involves the three constructs of the SCT lens of self-efficacy, behavioral capability, and observational learning. This is a difficult process and requires "a tremendous amount of time, all doing the exact same thing," according to participant number five. Another participant stated that their district is still working on rater reliability but thinks it is the most effective practice: "It was validating for rater reliability, to be able to score well on those and what the researchers at the Center for Education and Learning at Washington are focused on," stated participant number 1. Participant number ten stated, "I think the reliability training that we went to is certainly helpful." Only four participants discussed rater reliability training; however, the data shows its value for an administrator and administrative team as noted by participant 3 who stated "I think the most effective, actually, what we're still working on is the rater, interrater reliability." Participant number ten described the reliability training as "kind of like a refresher, a quicker refresher."

Recent Training

Figure 3 displays the most recent training received by each participant.

Figure 3

Most Recent Training Received by Participants



Seven of the participants received their training four to six years ago and four participants were trained within 3 years. Two of the participants received their training over 6 years ago. Eleven of the participants received professional training for using the 5 Dimensions Plus Teacher Evaluation Rubric through their administrative organization and only one participant received their training from the creators of the 5 Dimensions Plus Rubric from the Center for Educational Leadership at the University of Washington. One participant could not recall who provided their training but did recognize that their training was led by individuals outside their school district.

Summary

Each participant was able to express their experiences through an interview process surrounding observer training for those using the 5 Dimensions Plus Teacher Evaluation Rubric. I found the following themes as they relate to the research questions. Participants described the 5-day initial training for the 5 Dimension Teacher Evaluation Rubric as either overwhelming, extensive, or intensive. The data revealed that participants thought the many steps involved to learn the entire rubric created an overwhelming feeling, especially for new observers using the rubric. Multiple training sessions are needed to fully grasp the steps required by the 5 Dimensions Plus Teacher Evaluation Rubric. Extensive training identified by the participants was in the form of watching videos of teachers, then debriefing with colleagues emerged as a theme in the findings. Interrater reliability training followed up with networking with colleagues was also discovered in the findings

All of the participants expressed perspectives on what they experienced during their observer training. Addressing Research question 1, future training, data analysis also revealed a need for future training and in which that training should take place. Through a deep dive into the data and an extensive analysis utilizing an electronic coding application a thorough review of the data took place. Research question 2 asked the participants' perspective on the quality of training received. Watching video clips of classroom teachers and practicing the observation scripting and coding process, followed by debriefing conversations with others in the training, was a theme that was consistent among a most of the participants. Through an in-depth mining of the data, I discovered

the following themes surrounding the two research questions. The thirteen participants reflected on the two research questions which consisted of watching video clips of teachers and then coding the observation followed with a debriefing process with colleagues. Another theme that emerged in the data analysis process was the participants' desire for ongoing training and support and an opportunity to network with colleagues regarding the observation process utilizing the 5 Dimensions Plus Teacher Evaluation Rubric.

In chapter 5, I reiterate the purpose and nature of the study and why it was conducted, as well as summarize the key findings and what ways the findings confirm or disconfirm the findings from the literature review in Chapter 2. An analysis and interpretation of the findings in the context of social cognitive theory was a part of this study. A description of the limitations to trustworthiness, recommendations for further research, and the potential impact for positive social change.

Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this qualitative research study was to understand the perceptions of observers regarding their preparedness and the quality of teacher observer training they receive. This qualitative research study contributed to the body of knowledge needed to address the research problem of exploring the training those observers have received, as well as the future training observers feel they require to be more effective observers. Data analysis led to identifying trends, patterns, and training methods currently in place for observer training that are perceived to be effective and what participants would prefer for future training.

In this qualitative study, the focus was on the training process of those individuals who conduct classroom observations and final evaluations for the teachers in the state of Michigan. Focusing on the training opportunities available, the strategies used in training, and the perception and preparedness of observers, specific training strategies were identified that can be used in training sessions. Data collected and analyzed led to results that can be used to improve the training procedures and processes currently in place for observer training.

Relevance of Study

The relevance of this study was shown in the findings where it was proven there is a need for additional training for observers conducting classroom teacher observations. Additionally, the findings revealed the type of training method preferred for initial and future training for observers such as watching short video clips followed by debriefing with colleagues and refresher courses via an online platform. Trainers and the

organizations providing the 5 Dimensions Plus Teacher Evaluation Rubric training can use the findings provided by this research study to plan and organize future training sessions.

Key Findings

The findings of this research study identified two overarching themes: training needs and quality of training received. Several subthemes, such as networking with colleagues, live training followed by debriefing with colleagues, refresher courses via video conferencing, video training, and interrater reliability, emerged during the data analysis phase. The themes and subthemes can help design the content and frequency of future training sessions delivered to observers. Using an online platform where observers can collaborate with colleagues after watching video clips of teachers and compare ratings and scripting would address the needs of participants. Training received in the form of watching video clips of classroom teachers, conducting a mock scripting, coding and then comparing and debriefing with colleagues was the participants' preferred form of training. Participant number 11 stated,

> When it finally came to the part where we all script and code and then we would swap with other people, we would all see the same thing. And then we would swap how we coded it. That was probably the most—the part that was helpful for me.

The findings also revealed the value of interrater reliability training for observers being offered more frequently through an online learning platform, which participants felt was valuable.

The second theme discovered during the data analysis was the quality of additional training surrounding the observation process, such as coding and scripting after watching teaching video clips followed by a debriefing with colleagues and ongoing support such as video conferencing with colleagues periodically throughout the school year, along with the specific methods of delivering that training.

Interpretation of the Findings

The findings from this research study echoed the literature regarding the need for additional training of evaluators. Also, the findings in this research study bring the awareness of knowledge in the quality of training delivered by identifying methods in which the additional training should take place. Cosner et al. (2015) recognized that observers are more than likely to need extensive training and ongoing support. Smylie (2014) also stated training should cover a wide range of areas, like those areas identified by the participants in this research study. Doherty and Jacobs (2015b) and Dodson (2015b) also stated more training is needed, reinforcing the findings from this study. The types of training, the areas training is needed, and how that training should be delivered, are areas of new knowledge discovered from this research study.

Closely connected to the research question regarding the perception of quality of training and confirming the literature in Chapter 2 is the ability of the observer to distinguish between the rating levels of effectiveness when rating teachers. Determining good teaching from bad was a concern of a participant, which aligns with other findings from Rowna et al. (2013) and Weisberg et al. (2009). Gargani and Strong (2014), Kraft and Gilmour (2017), and Loewus (2017) aligned with participant responses, along with

Strunk et al. (2014) and Donaldson (2013) who also found that ratings of teachers were inflated. Also confirming previous knowledge from the literature review, four of the participants identified the struggle of completing the steps required of the observation rubric and the number of observations required, echoing what Halverson et al. (2004) stated regarding time management and how it is needed to complete required observations.

SCT was the theoretical framework I used in this research study. It emphasizes the environment in which observers work, their capacity to change their behavior, and the self-confidence to make that change (LaMorte, 2018). SCT places the emphasis on social influence, as well as external and internal social reinforcement. Three of the six constructs in SCT were used as a lens for the data analysis. Training involves the acquisition of new learning and behaviors. Three of the six SCT constructs were used in the data analysis process, specifically regarding training. Self-efficacy, behavioral capability, and observational learning all relate to a learned behavior and performing that behavior specifically to observation training. A blend of all six constructs were used during the data analysis process by looking through the lens of each construct and its meaning and applying them to participants' responses during the data analysis.

The theory considers the way in which individuals acquire and maintain behavior while also considering the social environment in which individuals perform the behavior (LaMorte, 2018). Participants identified five areas of need for future training. The areas identified are networking with colleagues, refresher courses, live training followed by a debriefing with colleagues, ongoing support, and video training with colleagues.

Providing training in each of the identified areas may result in a change in behavior surrounding classroom teacher observation, leading to an improved social and work environment. The findings in this study identified the need for future training and the specific areas the training is needed in. Currently there is a lot of attention given to the content in curriculum and how specific teaching methods are utilized in the classroom. Using the preferred methods found in this study for training that were identified in the data analysis, a positive culture of training and collegiality may result among observers leading to a positive behavior change in observation practices. Addressing the observation needs and implementing the methods identified may lead to a positive relationship and behavior change in teachers and observers resulting in a positive social classroom environment.

Limitations of the Study

Conducting a research study that involves interviews with participants in a traditional face-to-face manner was not possible during the pandemic. Due to the pandemic an alternate plan was put in place and interviews were conducted using a virtual face-to-face video conferencing platform called Zoom. Two practice interviews were conducted to gain interviewing experiences and to make any necessary adjustments to the interview protocol and the flow of the interview. However, these interviews were conducted in the traditional direct face-to-face environment not using the ZOOM platform. Upon finalizing all interviews, the pandemic closures and executive orders of no travel, these restrictions did not permit a face-to-face meeting, thus forcing a change in direction for conducting interviews.

During the interview process developing a report with the participants was very pleasant and went much smoother than expected possibly due to the fact that everyone in education was experiencing a new working environment via the video conferencing platform. Having an interview protocol ensured all interviews went as planned, overcoming the limited experience of conducting interviews via an on-line platform.

Only selecting participants in Grades 7-12 presented a limited number of available participants along with focusing on only those participants using the 5 Dimensions Plus Teacher Evaluation Rubric. In addition, it was necessary to have rural, urban, and urban/suburban represented in the study which was accomplished successful by having an equal number of participants from junior high and high school level and district demographic. By creating and sending over seventy invitations to potential participants it was possible to have demographic representation from rural, urban, and suburban school districts.

My being the only researcher in this qualitative study presented a dependability issue that was addressed by using member checks via email. Researcher reflexivity also endured dependability by exercising an intentional focus on each step by reviewing the data analysis process multiple times.

Recommendations

As the data were collected and analyzed from the interview questions about observation and training, it was evident that there was a need for further research surrounding observer training. The following are recommendations to expand previous knowledge and use the data gathered in this dissertation as a basis for further research.

- Future research should investigate the effectiveness of other teacher evaluation rubrics and the training provided for those rubrics.
- Experimental studies on the effectiveness of training approaches that incorporate the recommendations of this study are indicated, such as watching videos and then debriefing with colleagues and interrater reliability training followed by networking with colleagues.
- Further research should include kindergarten through Grade 6 and settings that include independent schools.
- Investigations of training conducted in virtual format is a need.
- Future experimental studies that investigate the effectiveness of training approaches that incorporate the recommendations of the administrators in this study.

The findings of this study present strategies that can be used to increase the awareness effective observer by those who use the 5 Dimensions Plus Teacher Evaluation Rubric. Increasing the perceptions of observers regarding the quality of observer training can potentially lead to an increase in the effectiveness of observations conducted by principals, resulting in effective teaching practices and leading to an increase in student achievement. The findings from this research study can also be used to design future training, the methods used for delivering the training, and how the training should take place. Also, how often the training is delivered, the length of each training session, and the method of delivery should be considered. The findings from this study may be used to create a training program leading to a work environment that is designed to meet the

needs of the observers, thus resulting in a more effective observation experience. Using these findings in the education setting may result in a positive social change among the trainers and observers, leading to more effective teachers, resulting in an increased student achievement. Effective teachers have a lasting academic and social impact on students along with the school environment, school district, and the community by increasing the social, educational, and economic opportunities for students.

The social significance of this study will be realized when the findings of this study are implemented in the training sessions of those individuals and organizations who provide training for the 5 Dimensions Plus Teacher Evaluation Rubric. The Michigan Association of Secondary School Principals and the University of Washington's Center for Educational Leadership may also benefit from the findings in this research study by addressing the methods of training and the areas training is needed.

Implications

Michigan law requires that individuals who are conducting classroom observations receive training on how to conduct classroom teacher observations using their teacher evaluation rubric of choice. However, the law does not require observers to attend additional training, which was evident in the findings of my research study.

The findings of my study provide strategies that can be utilized to design and deliver training practices to observers leading to increase the effectiveness of observers who use the 5 Dimensions Plus Teacher Evaluation Rubric. High quality observer training can potentially lead to an increase in the effectiveness of observations conducted by principals resulting in effective teaching practices leading to an increase in student

achievement. Increasing student achievement levels can lead to more opportunities for students while in school and after graduation. The findings can also be used to design future training, the methods that training should use, the length of each training session, and the areas the training should occur that meets the identified needs of observers. These findings can be used to create a training program leading to a work environment that is designed to meet the needs of the observers resulting in a more effective observation experience. Using these findings in the education setting will result in a positive social change amongst the trainers and observers leading to more effective teachers resulting in a positive training and observer experience and an increase in student achievement. Effective teachers have a lasting academic and social impact on students along with the school environment, school district, and the community by increasing the social, educational, and economic opportunities for students.

In addition, my study is relevant to the field of education regarding the training of observers that lead trainers' ability to assist teacher's improvement directly resulting in an increase in student achievement. These students then will be provided more opportunities leading to social change within themselves, their families, and their communities. The social significance of this study will be realized when the findings of this study are implemented in the training sessions of those individuals and organizations who provide training for the 5 D Plus Teacher Evaluation Rubric. The Michigan Association of Secondary School Principals and the University of Washington's Center for Educational Leadership may also be benefited the findings in this research study.

Based on the findings of this research study, the following are recommendations for professional practice for classroom observer training using the 5 Dimensions Plus Teacher Evaluation Rubric.

- Use video clips of classroom teachers while conducting a mock classroom observation followed by a debriefing with colleagues locally, regionally, or statewide.
- Attend and participate in on-going training specifically utilizing an on-line platform to deliver the training.
- Network with colleagues both locally and statewide to debrief and engage in discussion regarding the observation process using the 5 Dimensions Plus Teacher Evaluation Rubric.
- Continue efforts by observers to reach out to other observers to collaborate regarding the observation process, teacher effectiveness, and student achievement resulting in a greater collective knowledge.

Conclusion

It is known that the more effective a classroom teacher is, the better the chance of high levels of learning in the classroom (Marzano, 2003). This study dove deeply into the perceptions of observers' training strategies that are currently and the perceived outcomes of those strategies to conduct effective observations for those currently conducting classroom teacher observations and final evaluations and using the 5 Dimensions Plus Teacher Evaluation Rubric. The data show the need for continued professional development regarding the evaluation of teaches and their effectiveness. It is evident in

the results that ongoing training and networking are requested and needed for the success of teacher observation. The data gathered in the interview process brought a better understanding of the training received by observers, as required by Michigan PA 173 of 2015, as well as the level of training and preparedness. Classroom observers are a significant part of the improvement in classroom teacher effectiveness and meeting the training needs of classroom observers is vital for this improvement to take place. my study identified the reported training needs and in what form that training could take place. This study used peer-reviewed scholarly literature that had referenced the topic of classroom observations and the necessary processes it involves. A semi structured interview format was used to gather the data for my study. The interviews consisted of 20 questions surrounding observation and training with follow-up questions asked of the participants when more information was needed. As a result of the pandemic, interviews were conducted using a video conferencing platform that allows for live video. Thirteen participants were selected and agreed to participate in this research study. Participants represented rural, urban, and suburban demographics. An interview protocol was used as a guide for conducting the interviews. Interview notes were taken during each interview and all interviews were recorded for transcription purposes.

I gathered specific training methods that participants feel would benefit their professional practice, leading to improvement in teaching practices and student achievement. Since the passage of Public Act 173 Section 1249 (2) (n) of 2015, beginning with the 2016–17 school year, a school district shall ensure that training is provided to all evaluators and observers who will be performing teacher classroom

observations. The preferred training by participants was in the areas of observing live teaching followed by a debriefing with colleagues. The other method of training included the viewing of short video clips of classroom teachers scripting and coding the video clip with an opportunity to debrief with colleagues. Networking with colleagues via an online platform such as Zoom was also preferred by participants for initial training. At the time of the publication of this dissertation the Michigan Association of Secondary School Principals does offer additional training via an online platform for observers to attend without gathering in a group. Although this type of delivery was a result of the worldwide pandemic, it has benefited school observers by allowing training to continue. As a result, I plan to put in place professional development that involves the Zoom platform. Observers can gather via Zoom from their own school buildings. These meetings can be topic-generated specific to observer needs and will be set up by an invitation via zoom with a link to join the meeting. This was specifically requested by a participant.

Finally, conducting this research study I was able to evaluate my experience as a scholar, a practitioner, and a researcher. I gained a sense of confidence in my skills and capacity to accomplish more regarding research surrounding classroom teacher observation and training. When observers support teachers in ways that increase their self-efficacy, it can directly lead to an increase in student achievement, which may lead to opportunities for students to better themselves, their families, and their community resulting in positive social change at all levels. This study identified areas of current training practices that are grounded in past practices and bring with it a negative mindset in areas of meeting the needs of observers. The findings identified in my research study

will enhance the training practices and as a result create a positive experience for observers and trainers. Education has always been a relationship occupation now more than ever since the emergence of COVID-19 and the national pandemic. Support and understanding are necessary now more than ever to continue to create a positive work culture where all stakeholders can flourish. Superintendents and principals must establish numerous relationships with community, administration, staff, and students.

Understanding and meeting the training needs of observers will bring about a positive mindset for administrators and teachers. The training practices and methods discovered in this basic qualitative study will lead to observers becoming confident in their observation skill set. Utilizing the strategies to connect with those being observed while building and maintaining positive relationships through their actions and feedback will result in a positive observation experience for both parties. Relationships are key to bringing about a building culture that will lead to a positive social change in the school building and community. Putting in place the training practices identified in my study while cultivating and maintaining positive relationships will create a school building, district, and community culture that will lead to a positive social change in the relationships of all school and community stakeholders.

References

Archer, J., Cantrell, S., Holtzman, S., Joe, J., Tocci, C., & Wood, J. (2016). *Better feedback for better teaching: A practical guide to improving classroom observations*. Jossey Bass.

Bell, C. A., Yi, Q., Jones, N. D., Lewis, J. M., McLoud, M., & Liu, S. (2014). *Observer use of standardized observation protocols in consequential observation systems* [Paper presentation]. Society for Research on Educational Effectiveness Spring 2014 Conference, Evanston, IL, United States. https://files.eric.ed.gov/fulltext/ED562820.pdf

Cosner, S., Kimball, S. M., Barkowski, E., Carl, B., & Jones, C. (2015). Principal roles, work demands, and supports needed to implement new teacher evaluation. *Mid-Western Educational Researcher, 27*(1), 76–95.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). SAGE Publications.

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches* (3rd ed.). SAGE Publications.

Croft, S. J., Roberts, M., & Stenhouse, V. L. (2016). The perfect storm of education reform: High-stakes testing and teacher evaluation. *Social Justice, 42*(1), 70–92.

Culbertson, J. (2012). Putting the value in teacher evaluation. *Phi Delta Kappan, 94*(3), 14–18.

de Leeuw, E. (2008). Self-administered questionnaires and standardized reviews. In P. Alasuutari, L. Bickman, & J. Brannen (Eds.), *The SAGE handbook of social research methods* (pp. 313–328). SAGE Publications.

Derrington, M. L. (2014). Teacher evaluation initial policy implementation: Superintendent and principal perceptions. *Planning and Changing, 45*(1/2), 120–137.

Dodson, R. L. (2015a). Kentucky principal perceptions of the state's new teacher evaluation system: A survey analysis. *Educational Research Quarterly, 39*(2), 53–74.

Dodson, R. L. (2015b). What makes them the best? An analysis of the relationship between state education quality and principal preparation practices. *International Journal of Education Policy & Leadership, 10*(7), 1-21. https://doi.org/10.22230/ijepl.2015v10n7a634

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 Council on Teacher Quality. https://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading and learning.* National Council on Teacher Quality. https://www.nctq.org/dmsView/StateofStates2015

Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly, 49*(5), 838–882. https://doi.org/10.1177/0013161X13485961

Donaldson, M. L. (2016). Teacher evaluation reform: Focus, feedback, and fear. *Educational Leadership, 73*(8), 73–76.

Donaldson, M. L., Cobb, C. D., LeChasseur, K., Gabriel, R., Gonzales, R., Woulfin, S., & Makuch, A. (2014). *An evaluation of the pilot implementation of Connecticut's System for Educator Evaluation and Development.* University of Connecticut Center for Education Policy Analysis, Neag School of Education.

Dowley, R. G., & Kaplin, N. (2014). Evaluating evaluation: Assessing Massachusetts school districts' implementation of educator evaluation. *Journal of Law & Education, 43*(4), 485–502.

Gargani, J., & Strong, M. (2014). Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education, 65*(5), 389–401. https://doi.org/10.1177/0022487114542519

Green, P. C., & Oluwole, J. O. (2015). Response to applying value-added methods to teachers in untested grades and subjects. *Journal of Law & Education, 44*(3), 401–412.

Guest, G., MacQueen, K. M., Namey, E. E. (2012). *Applied Thematic Analysis*. SAGE Publishing.

Halverson, R., Kelley, C., & Kimball, S. (2004). *Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice.* https://web.education.wisc.edu/halverson/wp-content/uploads/sites/33/2012/09/HalversonKelleyKimball-TREA.pdf

Hazi, H. M. (2014). Legal challenges to teacher evaluation: Pitfalls and possibilities in the states. *Clearing House, 87*(3), 134–139. DOI:10.1080/00098655.2014.891898

Herlihy, C., Karger, E., Pollard, C., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record, 116*(1), 1–28.

Hill, H. C., Charalombous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough. *Educational Researcher, 41*(2), 56–64. DOI:10.3102./0013189X12437203

Ho, D. A., & Kane, J. T. (2013). *MET project research paper: The reliability of classroom observations by school personnel.* The Bill & Melinda Gates Foundation. https://k12education.gatesfoundation.org/download/?Num=2520&filename=MET_Reliability-of-Classroom-Observations_Research-Paper.pdf

Holstein, J. A., & Gubrium, J. F. (1995). *Qualitative research methods, Vol. 37: The active interview.* SAGE Publications.

Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher performance.* Center for Public Education. http://www.centerforpubliceducation.org/Main-Menu/Evaluating-

performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-Evaluation-Full-Report-PDF.pdf

Jacob, A. (2012). Examining the relationship between student achievement and observable teacher characteristics: Implications for school leaders. *International Journal of Educational Leadership Preparation, 7*(3). https://files.eric.ed.gov/fulltext/EJ997469.pdf

James, S. (2014). *The leadership skills, knowledge, and training required of high school principals to effectively evaluate classroom teachers.* (UMI No. 3680857)[Doctoral dissertation, University of Southern California, 2014.

Janesick, V. J. (2011). *"Stretching" exercises for qualitative researchers* (3rd ed.). SAGE Publications.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011a). Evaluating teacher effectiveness: Can classroom observations identify practices that raise achievement? *Education Next, 11*(3), 1–7.

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. **DOI** 10.3386/w15803

Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making with a standards-based evaluation system. *Educational Administration Quarterly, 45*(1), 34–70. https://doi.org/10.1177/0013161X08327549

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, *24*(1), 120–124. DOI: 10.1080/13814788.2017.1375092Kraft, M., & Gilmour, A. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher, 46*(5), 234–249. https://doi.org/10.3102/0013189x17718797

Kraft, M., & Gilmour, A. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, *46*(5), 234–249. https://doi.org/10.3102/0013189x17718797

LaMorte, W. W. (2018). *Social cognitive theory*. Boston University School of Public Health.

Lasswell, T. A., Pace, N. J., & Reed, G. A. (2008). Weighing in: Rural Iowa principals' perceptions of state-mandated teaching evaluation standards. *The Rural Educator, 29*(3), 40–44. https://doi.org/10.35608/ruraled.v29i3.465

Loewus, L. (2017, July 13). Principals are loath to give teachers bad ratings: Most teachers still rated as effective. *Education Week*, *36*(37). https://www.edweek.org/ew/articles/2017/07/13/principals-are-loath-to-give-teachers-bad.html

Marzano, R. J. (2003). *What works in schools: Translating research into action*. ASCD.

Mathers, C., Oliva, M., & Laine, W. (2008). *Improving instruction through effective teacher evaluation: Options for states and districts.* National Comprehensive Center for Teacher Quality. https://files.eric.ed.gov/fulltext/ED520778.pdf

Maxwell, A. (2013). *Qualitative research Design: An interactive approach* (3rd ed.). SAGE Publications.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation.* Jossey-Bass Publishers.

Michigan Department of Education, (2019). *Michigan educator evaluations at-a-glance.* https://www.michigan.gov/documents/mde/Educator_Evaluations_At-A-Glance_522133_7.pdf

Michigan Legislature, n.d. (2019) http://legislature.mi.gov

Moustakas, C. (1994). *Phenomenological research methods*. SAGE Publications.

Mowrer, K. (2014). *District response to Michigan's teacher evaluation laws* (Publication No. 364589) [Doctoral dissertation, Michigan State University, 2014. ProQuest Dissertations and Theses.

Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *The Journal of Psychology*, *89*(1), 158–164. DOI: 10.1037/0021-9010.89.1.158

Norman, A. D. (2010). Assessing accomplished teaching: Good strides, great challenges. *Theory Into Practice, 49*(3), 203–212. DOI 101080/00405841.2010487755

Norris, M., Van der Mars, H., Kulinna, P., & Ambrein-Beardsley, A. (2017). Administrators' perceptions of physical education teacher evaluation. *The Physical Educator, 74*(14), 730–756.

Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE Publications.

Papay, J., & Kraft, M. (2016). The myth of the performance plateau. *Educational Leadership, 73*(8), 36–41.

Pies, S. (2017). *Teacher evaluations: Do classroom observations and evaluator training really matter?* (Publication No. 10641881) [Doctoral dissertation, Indiana State University, 2017. ProQuest Dissertations and Theses.

Pisciotta, S., (2014) Relationship between Teacher Self-Efficacy and the Teacher Evaluation Process. Walden University ProQuest Dissertations Publishing, 2014. 361rt5ff5125.

Praetorius, A., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, *22*, 387–400. https://doi.org/10.1016/j.learninstruc.2012.03.002

Putman, H., Ross, E., & Walsh, K. (2018). *Making a difference: Six places where teacher evaluation systems are getting results.* National Council on Teacher Quality. https://files.eric.ed.gov/fulltext/ED590763.pdf

Quinn, A. E. (2014, September). Looking at the bigger picture with Dr. Robert Marzano: Teacher evaluation and development for improved student learning. *The Delta Kappa Gamma Bulletin, 81*(1), 12–18.

Revised School Code Act 451 of 1976, Michigan Common Laws, §§ 380.1249, 2011.

Rowna, B., Schilling, S. G., Spain, A., Bhandari, P., Berger, D., & Graves, J. (2013, December). *Promoting high quality teacher evaluations in Michigan: lessons*

from a pilot of educator effectiveness tools. University of Michigan, Institute for Social Research.

Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.). SAGE Publications.

Schmitt, N., Noe, R. A., & Gottschalk, R. (1986). Using the lens model to magnify raters' consistency, matching, and shared bias. *Academy of Management Journal, 29*(1), 130–139. https://doi.org/10.1111/j.1744-6570.1986.tb00950.x

Semmelroth, C. L., & Johnson, E. (2014). Measuring rater reliability on a special education observation tool. *Assessment for Effective Intervention, 39*(3), 131–145.

Smylie, M. A. (2014). Teacher evaluation and the problem of professional development. *Mid-Western Educational Researcher, 26*(2), 97–111.

Steinberg, M., & Sartain, L. (2015, January). Does better observation make better teachers? *Education Next, 15*(1), 70–76.

Stewart, K. (2016). A phenomenological study of practicing educators' personal and collaborative experiences within a climate of high stakes individual accountability (Publication No. 10583034) dissertation, Western Michigan University 2016 . ProQuest Dissertations and Theses.

Stitt, J. K., Simonds, C. J., & Hunt, S. K. (2003). Evaluation fidelity: An examination of criterion-based assessment and rater training in the speech communication classroom. *Communication Studies, 54*(3), 341–353. DOI:10.1080/10510970309363290

Strunk, K., Weinstein, T., & Makkonen, R. (2014). Evaluating teacher effectiveness: Can classroom observations identify practices that raise achievement? *Education Policy Analysis Archives: Arizona State University*, *22*(100), 1–38.

Taylor, S. E., & Tyler, H. J. (2012). Can teacher evaluation improve teaching? *Education Next*, *12*(4), 79–84.

University of Washington Center for Educational Leadership. (2012). *5D+ teacher evaluation rubric* (version 2).

Vinney, C. (2019, January 20). Social cognitive theory: How we learn from the behavior of others. ThoughtCo. https://www.thoughtco.com/social-cognitive-theory-4174567

Walsh, K., Joseph, N., Lakis, K., & Lubell, S. (2017). *Running in place. How new teacher evaluations fail to live up to promises.* National Council on Teacher Quality. https://www.nctq.org/dmsView/Final_Evaluation_Paper

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* The New Teacher Project. https://tntp.org/assets/documents/TheWidgetEffect_execsummary_2nd_ed.pdf

White, R. S. (2018). Who do state policy makers listen to? The case of teacher evaluation. *Phi Delta Kappan*, *99*(8), 13-18. https://doi.org/10.1177/0031721718775671

Appendix A: Interview Questions

DEMOGRAPHIC

1. Number of years as an administrator?
2. Total number of students and teachers in school?
3. Type of school? Urban? Suburban? Rural?

OBSERVATION

4. Please tell me about yourself as an administrator/observer and your experience with observing teachers.
5. How often are teachers observed and are classroom observations performed separate from the end of year evaluations?
6. Are you able to spend sufficient quality time in the classroom to accurately evaluate the performance of your teachers and what is a typical duration you spend in classrooms for each observation?
7. How has the observation process with the 5D+ influenced your classroom observations?
8. Describe the teacher observation process you use when observing teachers. (pre-observation conference, observation, and post-observation conference)
9. What, from your perspective, are the keys to effective evaluation?
10. What are the strengths and weaknesses that you perceive of using the 5D+ Teacher Evaluation Rubric?
11. What specific element of the 5D+ rubric do you find effective, and why?

12. How can observers best accomplish the goals of observation/evaluation as it relates to the 5D+ rubric?

13. What do you think are the challenges of the classroom teacher observation process for the observer?

TRAINING

14. How long has it been since your last training for evaluation/observation?
15. How much training have you had using 5D+ or teacher evaluation/observation in general and that type of calibration process or rater certification have you gone through if any?
16. What type of training did you receive in the area of teacher evaluation that you found most effective?
17. What type of training would you like to attend that would assist you in better evaluating all of your classroom teachers?
18. Describe the training you received from your school district in the evaluation process. (Follow-up question: Was the PD effective? If, (not / yes) why was it / or why wasn't it effective?
19. Who provided you the 5D+ training?
20. How much training Is the district requiring of you as an observer and describe your observation/evaluation training experience?

Final Statement

If I have any additional or follow-up questions, may I contact you? (yes / no). I want to again thank you (interviewee's name) for participating in this research. I know how important your time is and your responses have been extremely helpful.

Appendix B: Cover Letter

Hello Participant Name:

I hope you and your family are staying healthy and safe in these unchartered waters. As a building principal myself I totally understand what you are going through. It's really hard to lead a building of teachers when we are all distance learning and the daily contact is non-existent. It is also very difficult to reach out to my colleagues in my challenge to finish our school year regarding graduation and other end of the year activities let alone try to complete my research study which is my reason for contacting you. I need your help in completing a research project I have taken on to complete my doctorate degree.

I am a Ph D. candidate at Walden University conducting research on the training of teacher observers who are using the 5 D's Plus Teacher Evaluation Rubric. For my research study, I am looking to interview 10 to 12 administrators, gathering information on their perceptions about their training for conducting observations. The University of Washington's Center for Educational Leadership is interested in the final results of this study and may use the findings for future professional development. I have attached 2 documents and I am hoping you will look over the documents and possibly accept a phone call later this week where we can go over any questions you may have and set up a time for an on-line meeting. The attached documents are the introduction letter for you to review along with the consent form. I appreciate your time and consideration and also appreciate what you do on a daily basis.

Thank you,

Mark A. Williams, Principal

Appendix C: Interview Recording Protocol

Name of Researcher: Mark A. Williams | Date of Interview:

Name of Interviewee: | Job Title:

Years in Current Position: | Interview Location:

Interview Start Time: | Interview End Time:

Interviewee Anonymous ID Number:

Introduction:
Good (morning, afternoon, evening) (Interviewee's name) my name is Mark Williams, and I am a doctoral student at Walden University. I want to thank you for allowing me the opportunity to interview you. The interview that I will be conducting is a standardized open-ended interview with the goal and purpose of learning about your observer training, observer experiences, specifically your experiences with the 5D+ Teacher Evaluation Rubric. This is a fact-finding interview with no right or wrong answers so please answer as truthfully as possible.

As part of the process, I am interviewing Junior High and Senior High school principals in the state of Michigan. During our conversation, I am hoping to learn more about your observation practices and personal experiences with regard to the training provided for the 5D+ Teacher Evaluation Rubric.

I will be writing a report based on my findings in the interviews that I will be conducting, and I want to ensure you that the information gathered from our conversation will remain confidential and anonymous. I would like to tape record the interview so that I will have an accurate record of our conversation. Would that be okay? (Yes / No). I will also be taking some field notes as a secondary option for capturing the information discussed during our interview.

The interview should take approximately 60 minutes and can be stopped at any time per your request. Do you have any questions for me before we begin?

Appendix D: Original Codes and Times Referenced

Original Code	**Number of References**
Experience	13
Challenges	42
Time to schedule observations	4
Scripting	5
Indicators	12
Time to do observations	10
5 Dimensions Plus Rubric	158
Scripting	15
Weaknesses	21
Time to complete observations	10
Rubric	42
Growth tool	12
Strengths	22
Strengths	27
Growth model	27
Future training needs	65
Video training with colleagues	4
Practical training then debriefs	7
Network with another admin	13
Refresher course	12
Training	137
Should be continuous	4
Overwhelming	9
extensive	9
Video training	8
Inter Rater Reliability	7
Rater reliability training	14
Refresher training	16
Who Provided Training	19
Feedback	75
Relationships	4
Coaching	28
Observations	163
In-class process	11
Length of time	13
Wonderings	14
Post-conversation	10
Scripting	15