TEACHER PERCEPTIONS OF STUDENTGROWTH MEASURES

IN PERFORMANCE EVALUATION

A DISSERTATION IN

EDUCATION

Presented to the Faculty of the University
of Missouri-Kansas City in partial fulfillment of
the requirements for the degree

Doctor of Education

by

RYAN CURTIS MOST

B.A., University of Kansas, 2001
M.A., University of Kansas, 2004
Ed.S., University of Missouri—Kansas City, 2007

Kansas City, Missouri
2019

TEACHER PERCEPTIONS OF STUDENT GROWTH MEASURES

IN PERFORMANCE EVALUATION

Ryan Curtis Most, Candidate for Doctor of Education Degree

University of Missouri—Kansas City, 2019

ABSTRACT

The accountability measures for schools originally outlined in the No Child Left Behind (NCLB) legislation established links between student and teacher performance. Race to the Top (RttT) initiatives and waivers for the original Elementary and Secondary Education Act (ESEA) legislation required these links be used for teacher accountability in the form of including student growth measures, value-added measures, or student achievement data into teacher performance evaluations. While the most recent renewal of ESEA, referred to as the Every Student Succeeds Act (ESSA), has removed this inclusion as a requirement, the language of ESSA still encourages the use for merit pay and compensation measures. While the intent is clear, to hold teachers accountable and hopefully, improve practice, the question occurs, how do teachers perceive these reforms?

This qualitative narrative multiple case study examined those perceptions through the lived experiences of six core content teachers from a large mid-western middle school whose district has adopted an evaluation tool and process that allows for student growth measures to

be included in a summative performance evaluation score. Through interviews, narratives, and shared artifacts, participants contributed to the understanding of the phenomenon.

Each participant was treated as a separate case, and their in-depth interviews and responses to narrative prompts were analyzed using the three-dimensional narrative inquiry space. Cross-case analysis was used to analyze individual cases with the intent to identify similarities and differences across multiple cases. The results were used to answer the research question: How do teachers perceive the use of student growth measures in evaluations used for performance evaluation for intermediate grades teachers at an urban mid-western middle school?

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Education have examined a dissertation titled "Perceptions of Student Growth Measures in Performance Evaluation," presented by Ryan C. Most, candidate for the Doctor of Education degree, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Loyce Caruthers, Ph.D., Committee Chair
Division of Educational Leadership, Policy & Foundations

S. Marie McCarther, Ed.D.
Division of Educational Leadership, Policy & Foundations

Dianne Smith, Professor Emerita
Division of Educational Leadership, Policy & Foundations

Eva Tucker-Nevels, Ed.D.
Executive Directory of Secondary Instruction, Kansas City, Kansas Public Schools

CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

DEDICATION

To my wife and kids, Jennifer, Montana, and Macy whose constant support and

understanding allowed me to achieve my goals.

To my parents, Lannie and Jane, and parents in-law, Steve and Carol,

for the constant encouragement and faith.

To my Grandma Opal and my Grandma Ruth who believed education is a noble profession

and supported my pursuit to be an educator.

CHAPTER 1

INTRODUCTION

How do you know if someone effectively performs their duties? For some

individuals, accountability is quite clear; the doctor's patients are healed. When the mechanic

returns your car, it runs without the previously observed problems. When the roofer is

finished, the rain no longer drips in the house. The sales person makes their sales quota. The

architect's designs are selected and the building is built quite well. There are multiple

variables to account for the best and worst within these fields and; ultimately, it can be

determined if individuals are effective in their professions. Similarly, such variables can be

applied to education. How do administrators or leaders know when teachers effectively

perform their duties? Did students learn stated outcomes? How does the leadership know the

students have learned? How do they know the students learned from teacher A rather than

teachers B and C or from home or peers? Did the students learn enough? Did the students

learn the correct information and skills? How was teacher effectiveness measured?

The current educational setting in the United States includes many reform efforts and

proposed changes. Among these, the debate rages about how to improve teacher quality

(Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Marzano, 2012;

Murphy, Hallinger, & Heck, 2013; Papay, 2012; Shinkfield & Stufflebeam, 2012; Toch &

Rothman, 2008). Too often, from my vantage point as an administrator, have I seen poor

performing teachers remain unprompted to change or reassigned to a role that would limit

their negative impact. When teachers underperform, next steps include improving teacher

practice or removing them from classrooms. Before these next steps, evidence must be

present of the teacher's failure to perform expected duties including increasing student

achievement on standardized, high stakes assessments (Collins & Amrein-Beardsley, 2014; Drake et al., 2015; Milanowski, 2011b; Nixon, Packard, & Dam, 2011; Odden & Kelley, 2008; Paige, 2012; Thomsen, 2014).

When I began as a teacher and later as a building principal, the evaluation process consisted of the teacher and principal creating goals based on broad, vague points such as classroom management, engagement strategies, preparedness, or lesson delivery. None of these areas had descriptors, criteria or clear expectations. My early experiences as a teacher and principal reinforced the cycle and belief that effective teachers had organized classrooms and well-managed students regardless of student growth, achievement, and teacher performance. Not much has changed from these early systems, and evaluation systems continue to reflect obstacles, problems, and inaccuracies that impede their effectiveness. Danielson (2011) identifies outdated criteria, limited shared values of effective teaching, lack of precision in evaluating performance, use of a one-way communication method, and an inability to separate between novice and experienced teachers as limitations in many current systems. Marzano and Toth (2013) recognized similar hurdles and added a lack of consistent, clear standards of practice, inadequate time and staff for effective evaluation, and limited administrator experience. Darling-Hammond (2013) contributed student related components to ineffective evaluation systems including little consideration for student outcomes, and limited procedures that do not include the considerations for needs of teachers. Further, she suggested limited focus on improving classroom practice and evaluation results can be influenced by student assignment to classes and a separation between evaluation and professional development.

While repeating many of the previous stated hurdles, Toch and Rothman (2008) added that multiple variables influence student achievement making high stakes assessment scores difficult to include in evaluating teachers, and many local evaluation systems are riddled with assumptions about teaching that may be inaccurate or no longer applicable. Duffet et al. (2008) concluded that only 26% of teachers found evaluations useful. Another study established only 38% of teachers found administrators monitoring classroom practices useful (Leithwood et al., 2010).

I recall a colleague receiving a glowing end of year evaluation from a building administrator that had never stepped into the teacher's classroom the entire year. The teacher was more offended by the blatantly uniformed positive evaluation than if she would have a received a mediocre but accurate evaluation. I wish I could say this was an isolated incident, but many stories have reached me of equally similar poor practices.

Drake et al. (2015) explored the use of teacher effectiveness data by administrators in a multi case study of six large urban school districts by interviewing almost 200 members of central office leadership and building principals; conclusions were administrators preferred to rely more heavily on classroom observation and teacher development rather than on value-added measures and dismissal. I began this query, with the question: How was teacher effectiveness measured? I propose to study the perplexities of the problem for millions of teachers as they continue to wonder what an effective system would look like?

The Problem

Teacher evaluation and measurement methods inaccurately portray teacher performance and are riddled with a myriad of hurdles previously mentioned. Given the current state of the evaluation systems, these hurdles are inevitable; however, the opportunity

to improve teacher quality with a strong evaluation system leads to student learning and achievement. The emphasis of a strong evaluation system should focus on teacher development and growth rather than rating or scoring (Danielson, 2007; Marzano, 2012). Reinhorn, Johnson, and Simon (2017) conducted a qualitative, comparative analysis of data collected as part of a larger study of teacher recruitment and retention in high performing, high poverty, urban schools. As a result of this analysis, the researchers found as the schools implemented new evaluation policies and practices, there was a strong and consistent emphasis on teacher development rather than summative evaluation. Ironically, administrators are aware of the characteristics of effective teachers but are less willing to implement evaluation measures that will improve teacher practices among all teachers. The findings from the New Teacher Project in the Chicago Public Schools reflected the struggles of administrators with teacher evaluation and the tendency to identify those teachers needing help, especially among schools identified as failing. The New Teacher Project (2007) published the results of their analysis of Chicago Public Schools, which included an extensive review of human resources data, surveys of 464 principals, 1.446 teachers, 434 applicants, and 7 case studies of building administrators. Findings revealed that 93% of teachers received either an excellent or superior rating and less than one percent received an unsatisfactory rating, while 87 schools during this time met their definition of a failing school. Kimball and Milanowski (2009) concluded that many evaluators lack the skill and will to perform effective evaluations, and as such, many evaluations focused on praise with little to no constructive feedback. Recognizing a potential root cause to poor teacher ratings, principals need to improve their practice and ability to recognize effective teaching practices

and, at the same time, be able to coach those same practices for teachers needing improved

skill sets (Carbaugh, Marzano, & Toth, 2015).

Similarly, Weisberg, Sexton, Mulhern, and Keeling (2009) conducted a study across

four states and twelve school districts including 15,000 teachers and 1,300 administrators and

discovered resemblances across all participating sites, which included the following:

- The highest mark on evaluations received by 94%-99% of teachers.

- The most effective teachers were unrecognized.

- Districts failed to connect or identify professional development needs.

- No additional support for novice teachers.

- No actions taken for poor performing teachers

Kimball and Milanowski (2009) used teacher evaluation ratings and value-added student

achievement data to identify eight school leaders that demonstrated two years of consistent

validity scores. The intent was to interview them about teacher evaluation, decision-making

strategies, and school contexts. The mixed-methods study concluded substantial variation

existed in the relationship between evaluators' ratings of teachers and the value-added

measures of the students; and evaluators' decisions were predominantly based on motivation,

skill, and context observed. These factors highlight the imprecise representations teacher

evaluations reflect of teacher practices. Unfortunately, these misrepresentations of teacher

performance become the basis for personnel decisions, which could impact school districts,

school buildings, and student achievement over time (Darling-Hammond, 2013; Marzano,

Frontier, & Livingstone, 2011).

Because of poor evaluation and measurement methods, building administrators make

inaccurate or ineffective personnel decisions (Drake et al., 2015; Kimball & Milanowski,

2009; Odden & Kelley, 2008; Thomsen, 2014). Making high stakes personnel decisions based on ineffective instruments and practices could negatively impact teacher retention and increase teacher turnover. Using the wealth of data from almost 850,000 observations of fourth and fifth grade students attending New York City public schools across an eight-year period, Ronfeldt, Loeb, and Wyckoff (2013) utilized estimation strategies and regression models to determine the effect of teacher turnover on student achievement. Findings were grade levels with higher teacher turnover scored lower in language arts and math, but also low performing schools experienced an even greater decline. Steinburg and Sartain (2015) incorporated a naturally occurring experimental design created by the staggered rollout of the Chicago Public Schools new teacher evaluation system as part of the Excellence in Teaching Project which allowed them to compare the implementation of an observation and feedback heavy system newly implemented with 44 elementary schools as part of cohort one and then followed the next year by another 48 elementary schools identified as cohort two. While math achievement increased for the schools implementing the new evaluation system, it was not statistically significant, and at the same time the data demonstrated that the schools with consistently higher achieving, lower-poverty student populations made greater gains than the schools with higher-poverty student populations. Even teacher reassignment within the same school could negatively impact student achievement and teacher retention. Blazar (2015) found that grade switching is disproportionately higher for early career teachers and teachers in high poverty, low achieving schools, and these teachers have a higher rate of attrition as well as lower student achievement. Another concern identified in this study is the possibility of moving teachers with lower value-added scores to non-tested grades potentially placing the lowest performing teachers at the earliest and most formative grades.

Until required by federal incentive and expectation, school districts failed to update evaluation systems based on proven, reliable research (United States Department of Education [USDE], 2010). Findings from a review of practices in schools indicate human resources personnel filed evaluation results without attention to successful work or poor performance (Colby, Bradshaw, & Joyner, 2002; Milanowski , 2011a; Nixon et al., 2011). Public education places more emphasis on credentials and certifications teachers have earned rather than on performance in the classroom or student performance (Figlio & Kenny, 2007; Thomsen, 2014; Toch & Rothman, 2008). Schools function on the premise that having a certified teacher present to lead the class meets the criteria of effective instruction, and teachers act as interchangeable parts of the educational system (Weisberg et al., 2009). My first year as an elementary principal, I worked with a highly effective school counselor that established the counseling program and support for students and parents of the school and community. She had the support of the entire staff, community, and administration. Due to a paperwork deadline, missed by a week, her license was temporarily terminated. As a result, she was non-renewed, and we hired a school counselor that fulfilled the basic duties of the position. While the previous counselor went on to correct the licensure dilemma, she was hired by another school district and has since won multiple awards and recognitions.

A combination of factors contributed to the current dilemma including inconsistent standards of effective practice, separation of evaluation results and professional development, and inadequate time and staff to perform effective evaluations (Marzano, 2012; Marzano & Toth, 2013; Papay, 2012). Most teachers indicate the evaluation process fulfills a formality and falls short of improving practice (Darling-Hammond, 2013; Duffett, Farkas, Rothertham, & Silva, 2008). "Evaluation is rarely used to help teachers access professional development

to address their unique learning needs" (Darling-Hammond, 2013, p.5). Work originally created in the 1970s, absent reform, pervades many evaluation systems focused on developing student learning at a low-level (Danielson & McGreal, 2000; Marzano & Toth, 2013).

The renewal of *No Child Left Behind* (NCLB) legislation and the addition of Race to the Top (RttT) initiatives outlined in *A Blueprint for Reform* as well as the adoption of the *Every Student Succeeds Act* (ESSA) require teacher evaluation practices receive a drastic overhaul (Connally & Tooley, 2016; USDE, 2010). Specifically, student growth measures or value-added measures will contribute to teacher evaluations and constitute grounds for making personnel decisions (Thomsen, 2014; USDE, 2010). An unproven metric of effective teaching receives emphasis in these articles of legislation (Milanowski, 2011a; Paige, 2012; Schafer et al., 2012). Hallinger, Heck, and Murphy (2014) examined the new generation of teacher evaluations developed and implemented during the previous decade. They concluded that there is little support for teacher evaluation as a school improvement tool. Additionally, the ability to translate research related to effective teaching into monitoring tools of effective practices fails especially when multiple teachers are engaged with students in secondary schools.

When examining teacher incentives and student performance using the National Educational Longitudinal Survey results and their own follow up survey of those participating schools, Figlio and Kenny (2007) found student achievement was higher in schools that offered individual financial incentives to teachers, but they were unable to determine if the incentives generated greater effort from the teachers leading to increased student achievement or if higher performing schools happened to have an incentive program

as well. Randi Weingarten, president of American Federation of Teachers, spoke out often regarding teacher evaluation. Weingarten compared traditional practices to sports teams watching game tape for preparation only after all the games have been concluded (2010). While she recognized student achievement scores might play a part of teacher evaluation, she advocated for limiting the contribution of the metric to one small part of evaluation rather than a significant measure. Current initiatives, legislation, and funding requirements advocate for increased use and influence of student growth measures (USDE, 2010). Teachers' advocates raise their voice for fair, balanced, and transparent systems to foster growth (Darling-Hammond, 2013; Weingarten, 2010). These goals appear divergent at this time. This divergence can be viewed in the discrepant results of two studies. Adnot, Dee, Katz, and Wyckoff (2017), conducting a quasi-experimental research design to study the first years of IMPACT, the District of Columbia Public School's performance and assessment system that incorporates student growth measures. The researchers discovered replacing poor performing teachers with new or different teachers improved student achievement. Contradictorily, the study of New York City's fourth and fifth grade students demonstrated repeated turnover in teachers result in poorer performance in Language Arts and Math (Ronfeldt et al., 2013). The solutions provided and enacted to address the problem of ineffective teacher evaluation do not immediately address the needs of those at the center of the phenomenon, the teachers.

Purpose Statement and Research Questions

The purpose of this narratological case study was to examine teacher perceptions of student growth measures used for performance evaluations for middle school, core content teachers in an urban mid-western city. Teacher perceptions are generally defined as the realities constructed by participants through constant interaction and adjustment with the

phenomenon of the inquiry. Student growth measures are generally defined as those measures collected by a school district and reported to a state agency that will contribute to a summative evaluation score to be included in a teacher's performance evaluation. The unit of analysis was narrowed to the teachers' perceptions of the measures and their inclusion in evaluation.

The use of case study was intended to create thick, rich description in a context rich environment (Creswell, 2013). The narratological tradition was exemplified in the design by the collection of stories of those that experienced the phenomenon central to the topic (Patton, 2015). By collecting narratives revealing the teachers' perceptions of their work in schools and how it could be measured from the teachers' points of view allowed for comparative analysis to what was suggested in legislation and policy.

With the design of a narratological case study in mind the following central question and sub-questions were formulated:

- How do teachers perceive the use of student growth measures in performance evaluations for core content teachers at an urban, mid-western middle school?
    - o What do teachers understand about the context of job performance evaluation?
    - o How do teachers define student growth when measuring their own performance?
    - o How do teachers describe personnel decisions?

With the intention to examine teachers' perceptions of the inclusion of student growth measures in summative evaluation of their performance, I framed the research traditions to be incorporated as well as established the context within which the study resides.

Theoretical Framework

Using my experiences and perspective of teacher evaluation, I deepened my understanding of teachers' perspectives of the evolving landscape of teacher evaluation and the inclusion of student growth measures through the collection of stories of experience from case study participants. "Good qualitative research is relevant, timely, significant, interesting, or evocative" (Tracy, 2010, p. 840). The topic of teacher evaluation is relevant to a large audience even beyond the field of education, as every parent wants the best teachers for their children. Every educator, administrator, and parent share interest in determining, communicating, and implementing teacher evaluation practices.

The theoretical framework structure was used to create cohesion between the questions being asked, the phenomenon studied, the data collected, and the analysis conducted. The theoretical framework served a similar capacity as that of a conceptual framework, establishing linkages among the identified and defined components of the study as a written product (Maxwell, 2013). Case studies, particularly, need to establish theoretical considerations as these directly lead to the selection of case(s) to be included in the study (Yin, 2014). To this end, the framework outlined here established the purpose of the theoretical traditions to be implemented in the design and analysis as well as established purpose and connection among the components of the study.

With the altered requirements in federal funding policies introduced by ESSA, RttT, and NCLB, more states and school districts have adopted new teacher evaluation methods that include student growth measures and establish data trends to evaluate teachers, potentially significantly reshaping the educational structure. These actions served to eliminate many experienced teachers replacing them with new teachers, or possibly filling

the newly vacant positions with those not fully trained or certified to be teachers. This raises evocative questions such as: Is this change geared to eliminate ineffective teachers or motivate teachers to perform at a higher level? This question stimulated my reflections regarding the assumptions I brought to this study.

I brought several assumptions to this study. "A researcher can never enter a setting as a 'fly on the wall.' Instead, the researcher enters a setting and views the research setting through a specific lens (world view)" (Auriacombe & Schurink, 2012, p. 152). First, one of the important purposes of teacher evaluation includes constructive feedback to support growth and improved teacher practice. Second, school leaders want to participate in a model of reflective feedback and exchange but lack the training and experience to do so effectively. Third, student growth measures do not accurately reflect a teacher's contributions to student learning during the same year of instruction; student learning and test scores are affected by previous years' experiences. Finally, the use of student growth measures in summative evaluations would impede teacher development rather than motivate effective instructional improvement. In my roles as teacher, instructional coach, and administrator, I experienced a wide variety of evaluations leading to the previously stated assumptions. I believe teachers and administrators can effectively communicate and cooperate to improve instruction with an evaluation and growth process. To achieve this, it was important to have a deep understanding of teachers' perspectives regarding past and current experiences and their understanding of the legislative changes occurring that alter current evaluation practices in schools. Experiences of ineffective evaluation practices and limited support for development have created ongoing personal investments and views for the necessity to implement improved evaluation practices. Teachers will connect to the material to be discussed as it

directly relates to their roles, how their supervisor views them, personal and financial growth opportunities, and career advancement. In the ever developing realm of education reform, the negative rhetoric in the media and used by legislators should cause concern about coming changes.

Federal legislation, the *Every Student Succeeds Act* (ESSA), eliminated the mandate to incorporate value-added measures in teacher evaluation systems, but continued to provide grant funding for teacher performance pay systems moving forward (ESSA, 2015). So, while the requirement no longer exists, the effect of many states and districts adopting such practices lingers. Federal policy prior to ESSA's adoption required the implementation of evaluation systems that included student growth measures as a significant portion of summative evaluations. These new systems were to meaningfully differentiate between various levels of effectiveness and to provide meaningful feedback to improve practice and inform professional development (USDE, 2010). The *Blueprint for Reform* (USDE, 2010) assumes student growth measures on standardized assessments accurately reflect teacher contributions to student learning during that school year. Second, it also assumed incentive pay aligned to high performance student scores increases teacher performance and effectiveness (USDE, 2010). Next, the document implied fear of losing employment or premium assignments will result in increased teacher performance and effectiveness. Finally, the blueprint presumed removing teachers with low student achievement scores and replacing them with new teachers or creating new teachers through alternative certification programs would result in higher student achievement, a plug and play mentality. Tennessee, Washington DC, Houston, Wisconsin, Chicago, and other areas of the United States currently utilize evaluation practices incorporating student achievement scores, incentive pay, and

guiding personnel decisions. The short-term results varied and the long-term results were inconclusive regarding the inclusion of incentive pay, and student achievement guided personnel decisions did not show a significant difference from those areas not incorporating the same strategies (Glazerman, Chiang, Wellington, Constantive, & Player, 2011). I provide a deeper analysis of the literature surrounding these complex issues through the examination of four key strands.

The first strand of the theoretical framework investigates the history of supervision and evaluation. Throughout educational practice, the roles of supervisor and evaluator have been closely intertwined. Supervision includes the ongoing act of overseeing and directing practice, and evaluation includes the documentation of the development, practices, and actions developed while under supervision. The role of supervisor and act of evaluation has changed over the last two hundred years, and it continues to evolve with the implementation of new legislation. The second strand examines leadership practices occurring in conjunction with evaluation. Leadership, supervision, and evaluation do not occur separately from the day-to-day practices of a school organization. The actions of a school leader influence the culture, interactions of staff, and the understanding of the goals and vision of the school. The third strand is the teachers' perceptions of power dynamics within the evaluation process. Do teachers feel in control in evaluation settings, or do they feel it is something that occurs to them beyond their will? These power dynamics influence the perceptions of those involved—both teachers and administrators. Lastly, the nature of student growth measures is surveyed. What do these measures actually represent, and what roles have they or might they play in personnel decisions?

History of Supervision and Evaluation

Schools began as extensions of the communities for which they served beginning as early as mid-1600s, often directly connected to the religious sect most prominent in the immediate community (Hiatt, 1994; Kober, 2007; Labaree, 2011; Thattai, 2001). The primary curriculum was the bible, and the purpose of these early schoolrooms was to promote and maintain the faith and traditions of the local sects (Beach, 2007; Hiatt, 1994; Kantor & Lowe, 2004; Thattai, 2001). While supported by the founding fathers, reform for public education did not gain traction until the common school movement supported most prominently by Horace Mann as well as his counterpart Henry Barnard. The common school movement called for publicly funded elementary schools for all students to receive an education beyond the demands of religious sects in order to develop an informed and capable citizenry to support the still evolving nation (Beach, 2007; Dewey, 1903; Hiatt, 1994; Kantor & Lowe, 2004; Kober, 2007; Kress, Zechmann, & Schmitten, 2011; Labaree, 2011; Rousmaniere, 2007; Thattai, 2001). . The criteria used to evaluate teachers included how well they prepared the students for life beyond the school as well as the objectives created by those with authority, influence or power in the community (Beach, 2007; Kantor & Lowe, 2004; Rousmaniere, 2007; Tracy, 1995). More towns became industrialized and schools increased in number and size; and by 1918, all states had established some form of compulsory attendance policy or legislation requiring students to attend some form of school whether public or private (Kober, 2007; Thattai, 2001). Organizational structures changed as well, including the roles of those conducting supervision and evaluation (Rousmaniere, 2007; Blumberg, 1985; Cubberley, 1916).

Overtime, the early vestiges of current political and bureaucratic structures took hold including the creation of principals and superintendents. Cubberley (1916) defined the role of superintendent beyond its original clerical nature to include the leader, inspirer, and arbiter between teachers, community, and the school board. The first principals were lead teachers who took on additional managerial duties in addition to still fulfilling teaching duties, but this position evolved to be a full time position to carry out the orders of the school boards and superintendents while assisting teachers in fulfilling their duties (Cuban, 1988). In *The Principalship in Historical Perspective*, Kafka (2009) recognized four changes to the role of principal making the position more recognizable to today's society: the increase in school district size and enrollment required the handing of supervision duties to building heads, building principals demanded and stood up for more authority and autonomy regarding the management decisions within individual schools, principals established strong relationships with community members and established professional associations to provide support and information for their roles, and principals expanded their role in the supervision and guidance of teachers within their buildings. With the more defined roles of school leadership came more directive supervision and evaluation. Often these practices mirrored those present in industries outside education. One example would be scientific management during the early 1900s (Taylor, 1914). This model included the collection of student data to determine effectiveness and areas for improvement. At its core, the artifacts were a collection of observations heavily biased by the observer, often the supervisor or superintendent (Cubberley, 1916).

Moving further into the1900s, the practices evolved into a collection of common practices for fostering productive environments. The practices became expected procedures

or actions to be taken by the supervisor as well as the teacher (Biber, 1958; Coleman, 1945; Danielson & McGreal, 2000; Kafka, 2009). A gap exists in the literature regarding teacher evaluation beyond professional opinions during the day with the exception of the gender and race transitions following the world wars and into the civil rights era; notably, as the position of principal gained establishment and prestige, more white males sought the positions supplanting females and people of color who had previously held those positions (Danielson & McGreal, 2000; Hiatt, 1994; Kafka, 2009; Rousmaniere, 2007; Thattai, 2001). McHaney and Impey (1992) detail and define clinical supervision as a combination of expected steps including pre-observation conference, observation, analysis and strategy, post-observation conference, and a post-observation. The Madeline Hunter model followed, focusing on fidelity to the Madeline Hunter model still using the clinical supervision cycle (Hunter, 1980).

Susan Kent conducted a study regarding the use of clinical supervision to support the development of student teachers working with cooperating teachers in the classroom (2001). The mixed-methods study employed a quantitative component to create descriptive statistics for comparison purposes among participants. The qualitative component took the form of case studies using across case analysis. Sixteen teachers from a variety of grade levels from second through seventh participated in a three credit hour graduate course training them to use clinical supervision to support a student teacher during the upcoming school year. Only one of the teachers had supported student teachers previously, and this course had been offered to previous cooperating teachers with the intention for it to be offered again. The study followed those cooperating teachers use of clinical supervision with their assigned student teachers in the classroom. Data collected included written plans of intended

implementation, observation notes created during cooperating teachers' observations, conferencing notes from pre- and post- observation conferences, and reflective journals. The researcher found that while many verbally stated approval and intention to perform clinical supervision practices including all of the steps involved, the time-intensive demands of clinical supervision led to fewer cycles being performed by almost all participants. The other drawback reported by participants was the desire to be more directive during observation and post-observation conferences. The researcher also noted the need for ongoing and repeated professional development to maintain the skills needed for effective implementation of clinical supervision. As with many of the practices implemented throughout the existence of the American educational system, it may be easy to implement something, but it is difficult to implement any system well enough to result in improved performance (Fullan, 2006; Fullan, 2007; Hallinger, Heck, & Murphy, 2014; Waldron & McLeskey, 2010).

Up to this point, all supervision and evaluation focused on the supervisor. In implementation, the focus included what the evaluator observed and judged as effective or ineffective. Following the RAND report, *A Nation at Risk*, the focus shifted to eliminating ineffective teachers or rewarding successful ones, language still present in today's reform legislation (ESSA, 2015; USDE, 2010). With only a few exceptions, the focus throughout the history of supervision and evaluation has focused on the supervisor's observations, values, beliefs and judgments (Kafka, 2009; Marzano, Frontier, & Livingston, 2011; Rousmaniere, 2007). Since the passage of NCLB, teacher evaluation has continued to evolve and could include multiple components such as student work samples, lesson plans, self-assessments, and student growth measures, all with various weights of importance depending on state policy, school board policy, and negotiated agreements with teachers groups (Brandt,

Mathers, Oliva, Brown-Sims, & Hess, 2007; Hanzi & Arredondo, 2009; Mathers & Oliva, 2008; Thomsen, 2014). Even with these base models, the supervision and evaluation stems from the supervisor's knowledge, understanding and bias.

Leadership Practices in Conjunction with Evaluation

I find it beyond the scope of this proposed study to determine leadership moves that generate student success across the board in all contexts and settings related to teacher evaluation. Within the scope of this examination may be the opportunity to find where on the continuum from transactional practices to transformational practices, teachers perceive the greatest support and success. Both styles attempt to create change in practice either through extrinsic or intrinsic motivational factors. Current reform efforts include transactional practices such as rewards and recognitions (ESSA, 2015; USDE, 2010). Many districts direct school leaders to adopt more transformative methods including relationship building, fostering personal goals and accomplishments related to student achievement, personal growth and development (Fullan, 2006; Fullan, 2007; Moore, 2009; Waldron & McLeskey, 2010). Adopted frameworks focus on increased understanding of effective practices, extensive training, collaborative discussion, multiple observations, and coaching (Danielson, 2007; Gregoire, 2009; Kimball & Milanowski, 2009; Marzano & Toth, 2013; Milanowski, 2011b).

Ross and Gray (2006) conducted a study examining transformational leadership and collective teacher efficacy while accounting for variable context of leadership style and school culture. Their quantitative study consisted of survey responses from 3,074 participants from 218 elementary schools in two large school districts in Ontario. Leadership models were created and analyzed using participant responses for structural equation modeling with

variance-covariance matrices to analyze the maximum likelihood method. Then the models were further analyzed for best fit using criteria for chi-square greater than .05, AGFI (Adjusted Goodness of Fit) greater than .9 and RMSEA (Root Mean Square of Approximation) less than .08. The researchers split the participant schools of the study to create the initial sample and retest sample for internal validity. From their study, Ross and Gray concluded that transformational leadership contributes to collective teacher efficacy, but only after correlating three variables: a) leadership and mission, b) efficacy and professional community, and c) mission and community. Therefore, collective teacher efficacy is partially mediated by the effects of transformational leadership. Collective teacher efficacy strongly predicts commitment to community partnerships. Transformational leadership directly affected teacher commitment. The strongest outcome, commitment to school mission, holds importance given its role as a strong predictor of effectiveness (Ross & Gray, 2006). Leadership actions directly influence the efficacy of the staff. As leadership actions translate to teacher and student success, efficacy will increase and the likelihood of accomplishment will rise simultaneously. Teachers must establish similar transformational leadership in the microcosm of their classrooms.

For growth and improvement of practice, the focus of evaluation must reside with teachers. By fostering development through support, the likelihood of teacher success increases (Moore, 2009). Success leads to a sense of accomplishment and willingness to continue improving, starting a cycle of self-renewal. Instructional leadership actions enhance teachers' instructional capacity and promote student learning, connecting the instructional leadership to student learning-centered outcomes (Ovando & Ramirez, 2007).

Transactional methods have failed to produce consistent results. Over the past ten years, practices such as reward pay, career advancement, and public recognition for student results have inconsistently produced the results expected (Glazerman et al., 2011). Through a cycle of goal setting, strategy implementation, monitoring, and analysis in a transformational culture, students will find success (Connally & Tooley, 2016; Danielson, 2011; Darling-Hammond, 2013; Marzano & Toth, 2013; Milanowski, 2011b; Sartain, Stoelinga, & Brown, 2011; Robinson, Lloyd, & Rowe, 2008).

Teacher Power

Multiple stakeholder groups such as parents, students, teachers, building and district administrators, school board members, and state legislators, are invested in teacher evaluation for different purposes and reasons (Jacoby, 2011; Koski, 2012; Paige, 2013; Winkler, Scull, & Zeehandelaar, 2012). Recognizing the contributors to the debate, the question arises, how much power, control, or influence can teachers exert in this arena?

One option for teachers to obtain power could be to join a union. Stronger teachers' unions negotiate contracts less flexible to administrator managing with fewer management options than weaker teachers' unions (Strunk & Grissom, 2010). Consistently, stronger unions negotiate higher pay and more restrictive contracts (Strunk & Grissom, 2010; Winkler, Scull, & Zeehandelaar, 2012). Historically, unions have acted as a protector or buffer from external forces such as reform efforts and board demands (Ingle, Willis, & Fritz, 2014). Unions acquire their power from the ability to call a strike or withhold teacher service, political interactions such as lobbying, and influencing local and state elections through campaign and endorsement (Strunk & Grissom, 2010; Winkler, Scull, & Zeehandelaar, 2012). Unions have influenced state legislators and statutes to favor their members on many

topics including teacher evaluation (Koski, 2012; Winkler, Scull, & Zeehandelaar, 2012). Factors that strengthen unions include mandatory bargaining, a broad scope of bargaining topics, and the ability to collect dues from some or all teachers to provide union resources (Winkler, Scull, & Zeehandelaar, 2012).

Another opportunity for teacher control could be through collective bargaining of negotiated employment contracts. Collective bargaining often includes teacher evaluation among its topics (Koski, 2012; Strunk & Grissom, 2010). Collective bargaining has also been used to negotiate the use of walk-through and informal observations data as part of evaluations, adoption of evaluation instruments, reduction in force provisions, merit pay provisions, and individual teacher contract renewal provisions (Ingle, Willis, & Fritz, 2014). Even when topics or issues are not allowed by state policy or statute for collective bargaining, they can still be influenced by the negotiation of overlapping topics or by teachers making greater demands in other areas that then influence the original points not available for negotiation (Paige, 2013). Within negotiated agreements, multiple purposes for teacher evaluation have been included, such as to assess teacher performance, the creation of a climate of cooperation, employment decisions, improved quality of instruction, increased student learning, professional growth, and the recognition of exceptional teaching (Ingle, Willis, & Fritz, 2014).

Even with these limited stations of power, teachers still face challenges and opponents demanding reforms including teacher evaluation (Koski, 2012; Paige, 2013; Winkler, Scull, & Zeehandelaar, 2012). As the demand for teacher accountability increases through education reform, many unions find they must evolve their practices from resisting change to guiding the reform efforts that seem to move forward regardless of resistance

(Ingle, Willis, & Fritz, 2014; Jacoby, 2011; Weingarten, 2010). The mandate to implement reform strategies can directly affect teachers' efficacy and belief in potential success (Olsen & Sexton, 2009; Tschannen-Moran & Hoy, 2007).

A quantitative study relying on T-tests, correlation coefficients, and survey instrument validity scales collected responses from 255 teachers from graduate study programs at three state universities or volunteers from two elementary schools, a middle school, and a high school in Ohio and Virginia to examine antecedents of self-efficacy among novice and career teachers (Tschannen-Moran & Hoy, 2007). Specifically, attention was paid to determining whether mastery experiences or verbal persuasion affected novice teachers compared to career teachers. Overall, experienced teachers had higher self-efficacy, and mastery experiences had greater correlations of relationship than verbal persuasion for both groups. While, verbal persuasion was more influential with novice teachers, it held more connection to availability of resources. Ultimately, experiences create teachers' self-efficacy and establish perceptions of future likelihood of success.

Evaluation serves two purposes, judgment of performance and promoting professional growth in improved instruction. As teacher performance improves, student performance rises. "For some principals, it will require a shift away from the idea that they "just know" good practice when they see it, to seeing teacher evaluation as a process of collecting information to deeply diagnose teachers' strengths and weaknesses to improve instruction" (Sartain, Stoelinga, & Brown, 2011, p. 42). Teacher development occurs through focused individual development.

Student Growth Measures

Student growth measures, or value-added measures, attempt to measure the contribution a teacher made during the course of a school year. With a superficial analysis, this makes sense. Measure a student's reading ability at the start of the year, teach that student all year, measure at the end of the school year, the increase must be from the teaching that year, the basic presumption regardless of the topic, grade level, or school. Learning is a much more complex endeavor (Baker et al., 2010; Corcoran, 2010; Fuhrman, 2010; Goe, Bell, & Little, 2010). Often ineffective strategies taught previous years continue to be used skewing current year's results. Instructional supports outside the classroom can cover up failure in the classroom.

Houston Independent School District (HISD) is reported as the largest district with over 200,000 students and 13,000 teachers to implement value-added measures to inform high stakes decisions such as compensation, recognition, and retention. To examine the intended and unintended consequences of such policies by HISD, a case study examined four teachers terminated due to reports generated by the value-added system (Amrein-Beardsley & Collins, 2012). The participants worked at different schools within the district and all had different supervisors. Some participants had different supervisors during the course of data collection and observation cycles conducted as part of the value-added system. In all four cases, the strongest correlation between observation scores conducted by trained supervisors occurred only 50% of the time. The teachers received bonuses based on student score performance, but this relationship was not clearly established or understood as district teachers described the monetary awards as having "won the lottery" since they did not change their practices from year to year. The publisher of the evaluation program

implemented claimed to be able to account for variables such as second language learners and special education students, but the data collected and analyzed for this study did not support those claims. All four participants began the litigation process to reclaim their former positions alleging denial of due process. Only one saw their complaint all the way through to a mediator and conclusion. The determination being that the data provided to justify termination was at best inconclusive and not great enough grounds to justify termination on its own. As more districts and states adopt such high stakes practices based on value-added measures, a greater likelihood of resolving the legitimacy of value-added measures through judicial processes occurs (Baker, Oluwole, & Green, 2013; Marshall, 2012; Paige, 2012; Popham & DeSander, 2014).

Numerous other districts and regions have adopted the use of student growth measures as well, including parts of Florida, Tennessee, parts of New York, Chicago, and Washington DC. Data in these areas, over extended periods of time, show no differences to data trends prior to the adoption of student growth measures (Glazerman et al., 2011). Current federal legislation strongly encourages the incorporation of student growth measures in evaluation (USDE, 2010). Many states have adopted the inclusion of student growth measures in the evaluation process. These same states have also adopted new curriculum (common core or similar standards) and new tests for student measures (PARCC or similar exams). Therefore, students assess their knowledge of unproven curriculum with unproven exams by teachers overwhelmed by changes and the added pressure of possibly being terminated if students perform poorly.

The strands of the framework come together to complete the full image of the phenomenon for study. The history provides context for the actions occurring and the

reactions to these changes. Leadership actions influence the adoption, implementation, and potential success of any changes occurring. Now in the field of education, teachers represent those most directly influenced by changes in the evaluation systems and legislation. Their roles, opportunities to influence change, and perception of capability to influence the larger environment are critical components to generate greater understanding of their perceptions of the phenomenon. While, at face value, many may believe they know what and how student growth measures can be implemented, they are actually a complex metric, easily misunderstood by legislators, administrators and teachers equally. These concepts framed the study and directly guided the decisions apparent in the methodology of the study.

Design and Methods

This narrative case study captured the perspectives of teachers as they faced the adoption of student growth measures as a significant part of summative performance evaluations. Qualitative research created opportunities to examine relationships and connections at a deeper level (Maxwell, 2013, Creswell, 2013; Holley & Colyar, 2012). "One strength of qualitative research is its ability to illuminate the particulars of human experience in the context of a common phenomenon" (Aryes, Kavanaugh, & Knafl, 2003, p. 871). This study was qualitative because it sought to understand the meaning made by participants within the context of including student growth measures in evaluation (Maxwell, 2013; Holley & Colyar, 2012). "Given its strengths, qualitative research is essential for uncovering deeper processes in individuals, teams, and organizations, and understanding how those processes unfold over time" (Bluhm, Harman, Lee, & Mitchell, 2011, p. 1871). Case study was appropriate because the study delved deeply into the real-life, current context issue facing teachers within a bounded system (Creswell, 2013). "No idea or insight about the data

can be used to interpret the data set until it has first been shown to be important in individual experience. Insights from one account sensitize the investigator to similar information as it occurs in other accounts" (Aryes et al., 2003, p. 872). Narrative traditions were used because the information gathered was biographical in nature, and the perspectives of the teachers were shared as sacred, cover, and secret stories detailing their lives in their classrooms (Clandinin, 2013; Clandinin & Connelly, 2000; Creswell, 2013; Holley & Colyar, 2012, Hayes & Wood, 2011).

The study took place at a middle school (grades 6-8) in an urban mid-western school district. The district had adopted an evaluation program, meeting state guidelines and requirements, which included the ability to incorporate student growth measures into summative teacher scores to be reported to the state in an aggregate format. The school was selected based on the principal being in the current position for at least three years, representative of the district in socioeconomic status as well as cultural and linguistic diversity. The district hosted more than 80% of its students with free and reduced lunch and had more than 25% of the students classified as second language learners.

The participants were classroom teachers from core contents such as Math, Science, Social Studies, or English/language arts. The final reporting of findings was a sample of six participants representing different amounts of experience with evaluation in this school district. These individuals would have been selected through a narrowing process. All middle-school teachers were asked to participate in a survey collecting demographic information, interest in participating to a further extent, and requesting a brief narrative of their best or worst experience with evaluation. E-mail surveys made it quite easy to cast a large net for initial participants, but it must be considered if this may have affected their

27

responses or if this method recruited a different type of participant than a face-to-face method (Palys & Atchison, 2013). Those willing to participate further were invited to participate in more in-depth interviews and provide narratives as case study participants. Due to the low number of volunteers, all were included in the study. It was important to use self-reflexivity about the subjects and biases present to create sincerity and transparency (Tracy, 2010).

A variety of data sources were used in order to gain crystallization, and provide different insights related to the unit of analysis; the teachers' perceptions of the measures and their inclusion in evaluation (Babchuk & Badiee, 2011; Ellingson, 2009; Ellingson, 2014; Richardson & St Pierre, 2008). First, the survey responses from the intermediate teachers were available. Transcripts from the individual interviews served as data sources. Each case study participant was prompted to share their stories of evaluations during the interviews. "The goal of qualitative research using a narratological tradition is to understand the human experience through interpreting narrative forms of qualitative research data" (Hayes & Wood, 2011, p. 293). To provide additional context and insight, a document review was included in the analysis. Documents included federal, state, and school district guidelines and policies related to teacher evaluation as well as any internal communication from the school detailing procedures and training related to evaluation. In keeping within the intent of researcher as instrument (Patton, 2015), I included my reflections as a data source as well. The reflections took the form personal memos and journal entries.

Multiple methods of analysis were incorporated. Both within case and a cross case analyses were used for the primary analysis of case study. Within-case analysis identified themes in the data sources for case. "Interpretive techniques designed to be used within individual accounts or cases provide a wealth of contextual richness and person-specific

information without which that case cannot be understood" (Ayres, Kavanaugh, & Knafl, 2003, p. 873). The cross case analysis was used to examine the themes across multiple cases. It was useful to use both since the target participants had different years of experience, thus, possibly different experiences and perspectives. The interviews and collections of narratives were analyzed through three-dimensional narrative inquiry space—interaction, continuity, and situations (Clandinin & Connelly, 2000). Another important aspect was to perform such analyses throughout the process of collecting data to guide reflections and even possibly adjust the methods to discover the information needed to best answer the research questions (Auriacombe & Shurink, 2012, p. 148-149).

In summary, this study focused on six participants with variable years of experience teaching core content subjects at the middle school level. They were selected through processes of criteria, participation, and the ability to reflect and communicate about the topic of the study. Throughout the process, I was reflexive and reflective in nature in order to adapt the study to the learning that occurred while not ranging so far afield as to jeopardize the ability to answer the research questions. Clandinin (2013) established the justification for incorporating narrative beginnings, which would be my autobiographical narrative to capture my relationship with the content of the study. While it was my priority to protect the credibility of the study during its pursuit, there was an equal, if not greater, responsibility to protect the participants of the study. This area will be discussed in Chapter Three: Limitations, Validity, Reliability, and Ethical Considerations.

Significance of the Study

My experiences with teacher evaluation have been filled with the best of intentions from all parties, to provide high quality education to students. Unfortunately, those same

experiences fell short due to a lack of administrator training, limited time to implement quality evaluation practices, hurried program adoption and weak fidelity in implementation. Working in a large school system, at times, felt as if the system kept moving regardless of the ability of individuals within the system to keep up. The inaccurate and unfair use of student achievement data by school leaders to inform decisions affecting teachers is a violation of ethics and law (Baker, Oluwole, & Green, 2013). State representatives and union leaders need to understand how teachers influenced by reform efforts, policy changes, and program adoptions perceive those changes and the impact of the teachers' work and actions (Ingle, Willis, & Fritz, 2015; Koski, 2012; Weingarten, 2010). School board members and district leaders updating district policies and adopting teacher evaluation programs to ensure a highly qualified staff would benefit from this study, and findings would contribute to the board members' and district leaders' understanding of teachers' views of policy changes and program adoptions (Donaldson & Papay, 2012; Lee, 2011).

Administrators, as the individuals typically conducting the evaluations, should be familiar with how their practices, no matter how constrained, can be perceived by the teachers they are charged with supporting (Danielson, 2011; Darling-Hammond, 2013; Marzano & Toth, 2013). The study brings awareness to teachers as they seek understanding, information, and influence related to their professional worlds (Jacoby, 2011; Superfine & Gottlieb, 2014).

The study added to the literature in the field, representing a point in time during nationwide legislative changes (ESSA, 2015; USDE, 2010). At the same time, the study also contributed to the understanding of the topic from the point of view of those impacted by such policies and by offering examination of the perspectives of teachers across ranges of

experience. The study increased understanding of the phenomenon from a targeted viewpoint in multiple ways. First, it allows building administrators a deeper understanding of those being supervised and supported, allows more accurate and necessary steps to create success for all teachers. Secondly, it allows those charged with creating and adopting policy at a district, state, and federal level to grasp how their actions could be interpreted by the educational system. Finally, practice can be improved by providing teacher voice to the discussions where decisions regarding them are made. Without a deeper understanding of the affective consequences of continuing policy adoptions related to student growth measures, it is impossible to fully determine if such policy adoptions are creating the opportunities for the intended successes.

Conclusion

Federal policy changes and reform movement demands to increase the quality of teachers have led to the examination and in many cases adoption of teacher evaluation tools and practices of incorporating student growth measures as a significant representation of a teacher's ability to perform his or her duties (ESSA, 2015; USDE, 2010, Thomsen, 2014). The adoption of such practices has the potential to lead to greater teacher attrition and lower student achievement (Drake et al., 2015; Kimball & Milanowski, 2009; Odden & Kelley, 2008). By examining teachers' perceptions of the incorporation of student growth measures in summative evaluations, this study provided insight of those most affected by such policy changes.

In chapter two the theoretical framework is developed more fully as the literature review clearly bounds the context of the study in the four identified cornerstones of history of supervision and evaluation, leadership practices in evaluation, teacher power in evaluation,

and the nature of student growth measures. Chapter three details the design of the narrative

case study and the methods of data analysis incorporated. Chapter four details the results of

the study and the analysis. Finally, chapter five responds to the research questions and

illuminates the discussion of further research needed to explore the topic.

CHAPTER 2

REVIEW OF THE LITERATURE

Teaching is a complex form with aspects of art and science used to convey meaning and build capability and knowledge in others. Evaluating teaching is no less complex and holds within it the purpose of increasing capability of teachers while ensuring quality instruction. The history of evaluation, discussed in this review, demonstrates the evolution of community supervision into clinical supervision, and now formative observation and professional development. Educational systems, at times being very slow to change, adopt evaluation models, move on to new ones while keeping parts of the previous, and slowly build their current practices and systems. At times, this creates competing theories of practice in place, or misunderstood programs used in rudimentary fashion; failing to serve the purposes as originally designed. Through all of it, different leadership styles, characteristics, and approaches support or invalidate the evaluation practices in place.

A review of the literature was conducted using internet accessible databases including JSTORE, ERIC, Google Scholar and WorldCat. Key search terms included: teacher, teacher evaluation, leadership, leadership style, instructional, transformational, transactional, teacher evaluation practices, student growth measures, SGM, value-added measures, and VAM, value-added models. Additionally, items inaccessible through the database lists were acquired through the university library system when possible. The terms "teacher evaluation" and "leadership" in conjunction yielded the most useful results of accessible published works since 2009. These works consisted of both qualitative and quantitative studies as well as journal articles describing and defining leadership and evaluation theories and practices. Recognizable gaps in the literature centered on the perceptions of those immediately

influenced by the changes, the teachers and administrators. Additionally, the longer-range influences of these changes on those in the profession or choosing the profession were thin in the research (Ballou & Springer, 2015; Goldhaber, 2015; Harris & Herrington, 2015; Johnsons, 2015). With time and attention, these areas can be filled in with this study to contribute to the former.

In order to better understand the larger context of this study, an examination of the literature focused on four aspects of teacher evaluation. First consideration was the history of supervision and evaluation. Over the course of two hundred years, the act of teacher evaluation evolved from the oversight of prominent community members and religious organizations through the creation and implementation of superintendents, head teachers, principals, and instructional leaders. The evolution of school leaders moves to an examination of the leadership practices occurring in conjunction with evaluation. Simultaneous to the creation of new leadership roles in schools, the duties, demands, and practices of these leaders influence the effectiveness of their schools and teachers. Naturally, this leads to an examination of those being led. What works? What doesn't? How does this influence the adoption of new reforms? What power can teachers exert over their changing environment? Lastly, what is the new tool that uses multiple names such as value-added measures (VAM) and student growth measures (SGM)? Teachers must interact with these tools as part of their evaluation process in the continuing evolution of evaluation reforms.

History of Teacher Evaluation

When examining teacher evaluation during the history of the United States, two particular metaphors come to mind. First, much like technology has undergone increasingly rapid changes, improvements and evolutions, so has teacher evaluation. In earlier days, there

were long stretches of common practice just as there were long stretches of common technology, but as the years progress, the time spans shrink, and with each passing timeframe, changes occur and the next is shorter. While it is difficult to conceive, the question arises: Will there be a terminal timeframe in which it cannot be any smaller while demonstrating advancement? This applies to teacher evaluation as well. At what point, do we stop innovating to determine if the practices utilized are working? The second metaphor is the most common in the teaching field, and that is the wheel. Common advice from senior staff members of any school is to wait a length of time, as all initiatives, strategies, and programs will return under a different name or guise. In reviewing the last 400 years of educational practice related to teacher supervision and evaluation, this becomes evident (Fullan, 2015; Goldstein, 2014; Paley, Heilig, Cole, & Sumbera, 2014). Even as we embark on new methods, educators can see the seeds planted 20, 30 or 40 years prior. As the common saying goes, it is important to know what has gone before, or you shall be doomed to repeat the mistakes. At what point do we stop the educational wheel from bringing us back to past practices? How do we know what works and what does not?

An important note for the purpose of the review of the history of teacher evaluation, due to the incredibly close relationship between supervision and evaluation during previous and at times current practices, includes extensive references to supervision as part of evaluation. Many practices and programs defined supervision as the ongoing interaction between the teacher and supervisor to improve practice, and evaluation is the culmination of the supervisory interactions for a summative determination of effectiveness (Danielson, 2010; Marzano & Frontier, 2011; Marzano & Toth, 2013; Mette, Range, Anderson, Hvidston

& Nieuwenhuizen, 2015). As such, while supervision could occur without evaluation, evaluation cannot occur without supervision.

The history of teacher evaluation will be discussed according to five eras. First, I examine pre-twentieth century education and evaluation. This looks at the formation of early American schools, oversight, the earliest reforms in the formation of common schools, and the creation of the position of school principal. With the early twentieth century, what is commonly recognized as school districts, structures, and leadership are in place, and a closer investigation is given to evaluation procedures and structures as they continued to evolve through the 1900s. The mid-twentieth century captures a snapshot of evaluation as the country continued to change and develop socially having participated in World War II and the Korean War following the social upheavals of the early twentieth century such as World War 1, the Great Recession, prohibition and women's suffrage. While any part of this time period could be given a more detailed inspection, the goal of this section of the literature review is to provide a larger view perspective. During this time, the country participated in multiple large-scale conflicts across the globe, multiple social movements and changes such as women's suffrage in the early part of the century and civil rights protests and advancements later in the century. The late twentieth century provides the context from the civil rights movement and the educational reforms that followed the RAND report, *A Nation at Risk* (1983). Finally, the developments of the twenty first century led to current practices and developments of the accountability era of education.

Pre-twentieth Century

From the earliest days of settlement into the 1800s, America's school systems were completely products of the local community, and typically extensions of the most prominent

religious sect (Hiatt, 1994; Kober, 2007; Labaree, 2011; Thattai, 2001). For many

communities, this was often only a single teacher. As described by Tracy (1995) local

community leadership during this colonial time would visit the classroom and afterward

provide direction on expectations for implementation and improvement with the goal for

student success focused on the ability to read the Scriptures and to reflect the mores of the

community.

Entering the 1800s and the industrialization era, communities grew into much larger

towns and cities, and the common school movement began to gain traction. The founding

fathers of the nation supported public education, but Horace Mann of Massachussettes took

the steps to convince the nation of their need. The common school movement included all

students to receive a publicly funded elementary school education beyond the requirements

of local religious sects for the purpose of supporting the developing nation with a capable

and informed citizenry (Beach, 2007; Dewey, 1903; Hiatt, 1994; Kantor & Lowe, 2004;

Kober, 2007; Kress, Zechmann, & Schmitten, 2011; Labaree, 2011; Rousmaniere, 2007;

Thattai, 2001). Excluded from these developments were enslaved and free Blacks (Anderson,

1988) as well as Indians (Adams, 1995).

Teacher evaluation at this time included how well teachers prepared students for life

beyond school and the objectives created by authority figures and influential community

members (Beach, 2007; Kantor & Lowe, 2004; Rousmaniere, 2007; Tracy, 1995). The

structures of schools changed; a hierarchy of state, county, superintendents of education and

schools grew as did the demand and expectation of student participation reflected by the fact

that by 1918, every state had compulsory attendance laws requiring students to attend

elementary school (Kober, 2007; Thattai, 2001). Leadership structures changed with the

creation and evolution of different roles beyond classroom teachers such as the most

educated teacher or sometimes most experienced teacher would take on the role of

"principal" teacher or de facto leader of the school as the superintendent often supervised

multiple schools spread throughout an entire county (Rousmanier, 2007; Blumeberg, 1985;

Cubberley, 1916). This role eventually developed into the school principal and the role of

supervision and evaluation would be delegated to this role (Cuban, 1988).

Scientific examination of schools and teacher performance were in their infancy prior

to the turn of the century with many schools and districts beginning to adopt a recognizable

structure and organization by today's standards, but the research and study of this field began

to greatly increase with the nationwide adoption of scientific management principles shortly

after the turn of the century (Dexter, 1905; Lagemann, 1997; Luckey, 1903; US Congress,

1992). A small qualitative survey (N=5) of Midwest educational leaders in 1904 with the

purpose of defining the ideal secondary teacher determined that the survey participants

emphasized the need for social, charismatic educators with strong interpersonal skills and

effective communication tools while recognizing post-secondary education experience and

interest in improving practice as important but still secondary to social skills (Halleck,

Brown, Brooks, Nightingale & Brown, 1905). A much larger, mixed methods study to gauge

the effectiveness and preparedness of teachers surveyed 1,305 principals of public high

schools, approximately one fifth of all high schools at that time, across twenty-two states and

territories as reported by a 1902 report of the United States Commissioner of Education.

Principal respondents (N=583) represented 4,219 teachers distributed approximately to 37%

males and 63% females. Teachers holding college degrees included 70% male and 53%

female with an additional 7% and 10% respectively having at least some college experience.

Also,120% of the men had post-bachelor's degree experience1and 7% of the women. Only 5% of the men and 6% of the women had participated in training experiences related to pedagogy, whether they had attended college or not due to the new adoption of pedagogy related to teacher preparation. On average, high school teachers taught three subjects. The qualitative component of the study included written responses from the principals, which illustrated that more than half of the principals did not deem college experience as necessary for elementary school teachers, but consider it significant for high school teachers (Dexter, 1905). School districts and systems became more organized and developed, and as they did so, educators began to look for ways to improve their practice. This becomes quickly evident with the adoption of new practices following the turn of the century.

Early Twentieth Century

At the turn of the twentieth century, Frederick Winslow Taylor introduced scientific management. In the introduction of his work *Principles of Scientific Management*, Taylor (1914) shared the perspective that while directed towards engineers and demonstrated with manual laborers, the practices of scientific management apply to all areas of a person's life including church and home. He maintained "maximum prosperity" could be accomplished through meticulous observation and standardization of practice through training. Taylor performed social experiments at Bethlehem Steelworks and published the successful results. Through multiple illustrations, he demonstrated the potential to increase performance and output through initial observation, breaking tasks into discrete components, and training individuals to perform those discrete components more effectively than previously practiced.

Cubberley (1916) took Taylor's principles and applied them to public school education in his book *Public School Administration*. Cubberley outlines how similar

principles could be used to manage schools using data collection to measure student learning and teacher effectiveness. Based on observations and student work samples, teachers can benefit on feedback regarding areas to improve (Au, 2011). Adding to this work, Wetzel (1929) argued for the needs of student assessment data; in this case, aptitude tests, accurate scoring of student results and grading, and the necessity to accurately design curriculum to the meet the needs of an increasingly diverse student population. While using the same principles established by Taylor, Wetzel distance the work of teachers from the factory metaphor and recognized the need to serve diverse populations of students described more by socio-economic backgrounds than racial demographics.

With the context established by Taylor and Wetzel during the turn of the century, the newly developing position of principal began to take definition. The early roles of the principal were clerical in nature to facilitate larger buildings and their communication, but Cubberly (1916) recognized the principalship developing into the inspirational leader negotiating the needs of the community, teachers, and students. With the concepts of scientific management influencing the development of the principalship, Kafka (2009) identified the adoption of four roles more familiar with those serving as principals in today's schools. First, principals took on the building management duties as the schools increased in size and enrollment. Next, building principals demanded autonomy and authority to manage their individual schools. Also, principals established strong community relationships and fostered professional organizations to support their practice. Finally, they also took on the supervision of instructional practices of teachers.

During the implementation of scientific management principles, new theories of education and learning began to develop and be debated by such notable figures as John

Dewey and Edward Thorndike. Dewey (1916, 1938) advocated that students do their best in an environment where they are allowed to experience and interact with the curriculum, and all students should have the opportunity to take part in their own learning. Thorndike (1931) established psychological groundwork for the basics of learning including punishers, negative associations with a stimulus, and reinforcers, positive association with a stimulus, and their influence on learning. Moving further into the twentieth century, common practices for fostering productive environments came to be expected of supervisors and teachers based on combinations of these early principles and theories (Biber, 1958; Coleman, 1945; Danielson & McGreal, 2000; Kafka, 2009). A gap in the literature exists regarding teacher evaluation and supervision with two exceptions: topics specific to gender and race transitions as white males return from the war, or opinion, perspective pieces based more on author expertise rather than empirical research (Danielson & McGreal; Hiatt, 1994; Kafka, 2009; Rousmaniere, 2007; Thattai, 2001). During this early part of the century, roles and responsibilities of principals and superintendents became more defined while establishing the concepts and theories from which the next part of the century would build.

Mid-Twentieth Century

In her 1945 article "The Supervisory Visit," Coleman describes the basis of the modern evaluation process. First, the supervisor must know the teacher and the context in which that person is teaching. According to Coleman, the background knowledge should include the community, interests, aptitudes, and training the teacher received. It is even recommended to visit the teacher's home or attend their church to gain as much information as possible. Next, a relationship should be established to prepare the teacher for a supervisory visit. The actual visit described by Coleman is a shared experience. Depending on how the

students and teachers are engaged, the supervisor should participate as a support, to co-lead, or as a learner. Finally, a discussion about the visit should occur as soon following the visit as possible. During this conversation, the plans, implementation, and results are analyzed and evaluated. This establishes a clear description of a modern cycle, which includes, pre-observation conference, observation, and post-observation conference. The comparison to clinical supervision as defined later in the twentieth century as a process of pre-observation conference, observation, analysis and strategy, post-observation conference, and post-observation analysis becomes clear (Hunter, 1980; McHaney & Impey, 1992; Reavis, 1978).

Lewis and Leps (1946) described the transition of duties from superintendents to building principals. In their description, the principal is the key person planning and executing the school program. Up to this point, supervision is the responsibility of those outside the school, but with the development and formalization of the building principal role, the duties migrated. The authors describe the principles needed to keep the principal grounded in the important work and not let the authority of the role lead the principal away from the intended goals of the school program. Mildred Swearingen (1946) added that many duties were assigned to the supervisor, but chief amongst them were improving the curriculum, improving teaching personnel, and improving the teaching-learning situation.

As noted previously, the early references and materials do not mention evaluation directly, but it is often included or implied in the description of duties and responsibilities of supervisors of these time periods. This is clear as Thompson (1952) describes the roles of a supervisor; she includes many duties, an important one being helping teachers solve problems. Thompson's view was a supervisor does not and should not necessarily wait for an invitation to supervise and address concerns in a teacher's classroom. While an invitation

may suggest the teacher is open to the concepts and ideas available for improvement and problem solving, the lack of an invitation is not an impossible barrier to feedback and addressing concerns.

Whitehead (1952) conducted a study examining teachers' attitudes towards instructional supervision and improvement during this period with a compilation of1questionnaires from 115 teachers in North Carolina. These teachers were spread over the majority of the state, but each was employed at a segregated school established to serve students of color. The size of schools ranged from 4 teachers to 57 teachers, most of them coming from smaller schools. The schools operated on a democratic leadership model in line with Dewey's theories as described in questionnaire responses. The questionnaire consisted of six parts, but only the first two related to aspects of evaluation and feedback to staff to improve practice. Part I of the instrument gauged teacher opinions related to unplanned classroom visit practices. Responses included: One hundred percent of respondents felt a well-planned visit followed by an individual conference was beneficial to improvement, 80% believed the classroom visit was made for the purpose of improving instruction, 20% stated the visit was made to give teachers ratings and inspect the physical classroom space, 72% vocalized the visits were unscheduled, and 57% stated the principal did not stay through the end of the lesson. Whitehead concluded that while improvements regarding classroom visits had occurred, improvements were still needed such as the principal remaining for the entire lesson and following up with an individual conference. Part II of the instrument gauged demonstration teaching and these responses were: Seventy-nine percent of the participants stated the instrument was utilized successfully, 96% identified demonstration lessons as voluntary, 95% stated demonstration lessons were followed by an individual conference, and

70% stated the demonstration lessons were rehearsed in advance. Whitehead asserted that the

rehearsed nature of the demonstration lessons greatly reduced their effectiveness due to

artificiality and stressed perfection. Overall, the study demonstrated the primary purpose of

the administrators should be to improve effective teaching. While the role of principal began

as a fellow teacher and building manager, the role continued to evolve with principals

accepting the responsibility for improving the practice of colleagues through supervision.

The role would continue to evolve as the demands for improved educational practices

increased in the last part of the century.

Late Twentieth Century

The next noticeable wave of teacher evaluation attempted to marry the two previous

approaches; the scientific principles started with Taylor and the human relations components

evident following World War II based on Dewey and Thorndike's work. The result coined

clinical supervision, modeled from medical training practices current at the time (Reavis,

1978). At its core, there is a five-step process: pre-observation conference, observation,

analysis and strategy, post-observation conference, and post-conference analysis (Hunter,

1980; McHaney & Impey, 1992; Reavis, 1978). Reavis (1978) found that in order for the

process to positively impact teacher behavior, the supervisor and teacher must trust each

other and work together, going through the motions of the process will not result in any

changes. In summarizing the research at the time, Reavis consistently found that clinical

supervision was preferred to traditional supervision even though many of the studies

contained potential flaws. The goals of clinical supervision, changing teacher behavior and

improving student learning, mirror previous supervisory and evaluation practices, and

research suggested clinical supervision improved teacher behaviors when conditions of

teamwork and mutual goals between supervisor and teacher were present (Coleman, 1945; Whitehead, 1952; Hunter, 1980; Kent, 2002; McHaney & Impey, 1992; Reavis, 1978).

Lerch (1980) later clarified the work of Reavis (1978) by comparing clinical supervision to the diagnostic and prescriptive approach used by physicians such as examining the patient, analyzing test data, diagnosis, and prescribing next steps. An area of emphasis to Lerch's approach focused on a single teacher behavior to change and attend to rather than examining all possible aspects of a teacher's practice to determine next steps. The mutually agreed upon aspect of the teacher's practice increased likelihood of changing teacher behavior because the teacher is part of the identification process rather than the recipient. In analyzing the arguments for and against clinical supervision, it is determined that if the supervisor and supervisee are working collaboratively together for improvement, then change in behaviors for the better will occur, while if they are antagonistic, then positive change will not occur (Hunter, 1980; Kent, 2002; McHaney & Impey, 1992; Reavis, 1978).

The Association for Supervision and Curriculum Development (ASCD) sponsored a research project during the 1979-1980 school year to examine four major instructional improvement processes: curriculum development, clinical supervision, staff development, and teacher evaluation (Cawelti & Reavis, 1980). The study investigated the ways school districts provide instructional leadership in these areas and perceptions of trends related to them. Included were seven large cities, six medium cities, and three suburban communities, which resulted in feedback from almost 900 teachers, 115 principals, 127 supervisors, and 17 superintendents and assistant superintendents. Instrumentation included interviews and questionnaires. Instructional leadership, labeled supervisory services by the authors, was determined to be the least adequately provided service in all types of communities. For

45

medium cities and suburban communities, less than a third of teachers and principals reported that clinical supervision occurred in a beneficial manner or time frame with many reporting no experience with the process. For urban schools, the results reported less than 25% of teachers and principals felt current clinical supervision practices were adequate. Regarding teacher evaluation, again, less than a third for all communities reported it as an effective practice. Interestingly, supervisors scored evaluation even lower than all other respondent groups. For all categories, superintendents and assistant superintendents scored all categories higher and more successful. In summary, the study suggested increasing instructional leadership (supervisory services) as it was consistently rated as the weakest aspect of all leadership services across all sites. Those who participated in any type of service consistently scored this process as a higher need; redefining supervisory services to emphasize a more supportive, instructional role.

Grimmett (1981) suggested adjustments to the clinical supervision process to strengthen its use based on the creation of1a conceptual link to teacher thought processes. He maintained that the definition of clinical supervision should enhance student learning through teacher instruction; supervisors and teachers work must together in the classroom environment with a shared goal. Grimmett asserted, in order to effectively change teacher behavior, the clinical supervisor must work with the teacher on-the-job, in the moment, to coach and support teachers from falling back into previous patterns, behaviors, and routines. The supervisor, therefore, encourages the teacher to keep with the changes and maintain innovative practice through difficult times and lessons. The supervisor assists the teacher with reflection while offering advice and encouragement for continued refinement of practice. Using the clinical cycle in a formative manner assures Grimmett, teachers are able

to explore new teaching strategies without fear of evaluative judgments. This use of the cycle reinforces the relationship required to effectively incorporate clinical supervision. Throughout the process, the supervisor and teacher continue to develop shared perceptions, common language, and common understanding which also reinforces professional relationships and communication. Grimmett believed the repeated cycling through the formative process protects the developed relationship from the critical analysis that occurs during evaluation. Additionally, if the teachers do not self-monitor, the clinical supervisor could partake in stimulating that skill and instilling reflective practice in teachers. As illustrated in this manner, teachers are deemed successful if they perform the teaching expectations, which reflect only half of the purpose of clinical supervision. To address the remaining 50% of the purpose, enhancing student learning, an adjustment must be included. The adjustment is the reintroduction of the analysis to the original setting to determine if student learning is influenced. Rather than completing a full cycle of clinical supervision, Grimmett described a move backward to the classroom setting following the post observation conference to witness the potential impact of the analysis, holding participants accountable to the entire process rather than being able to continue through the process without incorporating changed behavior. This would again shift the role of the supervisor to include maintenance of the new behaviors of teachers as they demonstrate success.

Another major development from clinical supervision was the widely adopted Madeline Hunter model. The focal point of the model consisted of seven aspects of lesson planning; anticipatory set, objective/purpose, input, modeling, checking for understanding, guided practice, and independent practice. Along with this unit framework, Hunter (1980) identified and described the use of six types of supervision and evaluation. As part of a

clinical supervision cycle, one of five supervisory conferences is suggested, with a variety being used over the course of a school year. Type A conference consisted of identifying effective teaching techniques used during a lesson and explaining why they were effective to reinforce their correct use. Type B conference consisted of identifying effective techniques used and assisting the teacher to brainstorm additional techniques that might serve the same purpose to increase the teacher's range and capacity. Type C conference is used to problem solve with a teacher on how to increase the effectiveness of lessons, and there is an emphasis on collaboration for this conference. Type D conferences may carry negative aspects as this conference is specifically used to address errors in instruction and build other strategies to avoid those errors in the future. Type E conferences are used to recognize excellence and encourage those individuals to take on new roles, leadership, and self-challenges. Finally, the sixth conference type is specifically evaluative and should be conducted as a summary of multiple previous conference types. Additionally, Hunter separated supervisory actions and the evaluation conference, but the two are still closely linked, as the evaluation is a summation of supervisory interactions.

Hunter also laid out a larger framework identifying teaching as a science and an art (Hosford, 1984). It is important to take note of this development as it occurs again as a foundational component of Marzano's later work (Marzano, 2007). Hosford (1984) reports those aspects of teaching, which are a science that anyone can learn and develop as well as those that are more of an art, which become more difficult to learn. Hunter groups the components into content decisions, learner behaviors decisions, and teaching behaviors, explaining that excellence and mastery are attainable for all aspects but only through dedicated practice with knowledgeable observation and feedback provided and responded to

48

for improvement (Hosford, 1984). Hunter described the teacher as a professional deserving of autonomy to make professional decisions related to his or her practice, and as such, teachers should behave as professionals by making intentional decisions that can be supported by data and expertise rather than by whim or fancy. The role of the principal or supervisor then becomes that of the teaching specialist to act as an observer, to motivate, and to validate intentional decisions. One of the most important skills needed for the role of pedagogical specialist is to be able to articulate the knowledge in a manner that enhances teacher understanding and performance. Unfortunately, due to time constraints, training, unwillingness to engage in this work, or other factors, much of the supervisory or evaluation work led more to routines, paperwork practices, and ineffectual check-in conversations rather than actual increased performance or student learning (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985).

During this same era, there was a movement recognizing the teacher's role, authority, and power in the supervisory and evaluation process. While much of the work, research, and theory to this point had focused on the supervisor, system, or top-down authoritative aspects, the teachers' roles were beginning to be viewed as well (Blumeberg & Jonas, 1987). First and foremost, in order for teachers to change their behavior, teachers must be willing to change their behavior. While this may seem clear and obvious, it did not gain much ground and attention until it became a push back component against the more prescriptive Hunter model type programs. Blumeberg and Jonas (1987) discussed teachers' control over supervision, recognizing that in order for improvement in practice to occur, the teacher must be a willing participant in the work, recognize the value in change, align those changes with beliefs about teaching, recognize the value of the supervisor's input, and feel valued

throughout the process. If any one of these pieces was missing, the teacher may go through the required motions, but changes in teacher behavior would not occur in a manner to positively influence student behavior.

Following the National Commission on Excellence in Education (1983) report *A Nation at Risk: The Imperative for Educational Reform*, several questions arose related to the field of teaching and many connected to teacher evaluation either as a tool for eliminating ineffective teachers or as a means of identifying successful teachers for recognition or reward. Stemming from those questions, the RAND Corporation conducted a study which has become a keystone component when investigating teacher evaluation. At its core, the study investigated, "how teacher evaluation can be used to improve personnel decisions and staff development." (Wise et. al. 1985, p. 63)

The initial study consisted of 32 school districts. Following the initial round of interviews and surveys to gather data about the districts and evaluation practices, the study narrowed to four school districts for case study comparisons of the different evaluation processes, purposes, and methods. Of the original 32 districts, six schools were selected from each district and six teachers from each school along with the principals and specialist staff participated in the information gathering process. The initial round included teacher perspective information, also points of view of community representatives from each of the schools (Wise et. al.1985)

Four purposes of teacher evaluation were identified: individual staff development, individual personnel decisions, school improvement, and school status decisions (Wise et. al., 1985). This illustrated a problem when examining the different evaluation processes and models. While the primary goal of teacher evaluation is to improve student performance

through improving teacher performance, the inclusion of other applications, diffused the effectiveness of any model implemented. Further, most model designs are geared towards identifying effective teaching, but lack design elements for changing teacher behaviors when found to be below an acceptably effective level. This study also recognized the necessity for teacher satisfaction and cooperation throughout the evaluation process if any changes in teacher behavior are to occur.

The initial survey including all 32 districts brought to light important information. Consistently, the evaluation process included in collective bargaining agreements focused on procedural applications rather than substantive elements. District personnel and teachers had been dissatisfied with previous evaluation models before adopting current practices. Teachers strongly promoted a more standardized evaluation as the narrative evaluations from before were insufficient and subjective. Also, different states had different levels of influence, as some were very open in that the districts keep records of evaluations while others were more prescriptive in all aspects of teacher evaluation.

As the study continued (Wise et al., 1985), findings suggested that on the broad descriptors the districts maintained many similarities in their evaluation models. They included a pre-observation conference, observation, post-observation conference, and some included a written component of expectations of future practice. Where the districts differed were in the particular details such as number of conference/observation cycles to occur, timing of the cycles, and self-evaluation components. Consistently, funding did not exist for training evaluators, nor was the amount of time needed to evaluate effectively recognized in the distribution of duties.

Problems identified as part of teacher evaluation included: (a) inaccuracies occurred in evaluations due to insufficient knowledge and weak convictions of principals, (b) teacher push-back or indifference, (c) inadequate evaluator training, and (d) principals lacked content knowledge in advanced courses (Wise, et al., 1985). In essence, there were many instances where principals were untrained or incapable of effective evaluation procedures or did not evaluate effectively in order to maintain relationships with staff members. Teachers, similarly, refused to participate or did not care about the results because they recognized the evaluators' ineffectiveness or feared finding they, themselves, were ineffective. Potential positive effects identified from teacher evaluation included more effective communication and clearly established goals and expectations (Wise et al.). The more teachers and principals spoke to each other, the more understanding and communication improved related to expectations and practices.

The study (Wise et al., 1985) resulted in five conclusions with twelve recommendations. These are discussed as follows:

- Conclusion 1: the evaluation system used must align with other district practices and beliefs. Related recommendations include: districts should maintain alignment of programs and beliefs while adopting an evaluation model that fits, and states should not be overly prescriptive in their requirements and expectations.

- Conclusion 2: District commitment and resources are more important than procedures and checklists. Related recommendations include: time must be granted to evaluators to perform duties effectively, school districts should regularly assess the quality of their evaluation systems, training should be given to evaluators.

- Conclusion 3: school district must identify the main purpose of the evaluation and then match the process to that purpose. Related recommendations include: districts should examine current systems and match appropriate processes and the district must determine if one or more evaluation systems are needed for different purposes then be clear about each.

- Conclusion 4: teacher evaluation must be seen as useful which depends on is efficient use of resources. Related Recommendations include: the district must allocate appropriate resources and the district should target resources to achieve benefits from teacher evaluation.

- Conclusion 5: teacher involvement and responsibility improve teacher evaluation. Related recommendations: districts should involve expert teachers in the supervision of peers, districts should involve teacher organizations in the design and oversight of evaluation, and districts should hold teachers accountable to standards of practice to perform on behalf of their students.

Historically, with few exceptions to this point, the literature has focused on supervisors' observations, values, beliefs, and abilities, but at this point in time, the era of accountability, the focus shifts (Kafka, 2009; Marzano, Frontier, & Livingston, 2011; Rousmaniere, 2007). The new focus, still found in our current legislation and policy, focuses on teacher ability, performance, accountability, and student achievement (Every Student Succeeds Act [ESSA], 2015; United States Department of Education [USDE], 2010). Stemming from her work with Educational Testing Service (ETS) to create a framework for state and local organizations to make licensure decisions, Danielson designed and published a comprehensive description of teaching, its expectations, and its supervision, *Enhancing*

Professional Practice: A Framework for Teaching (1994, 2007). The framework is divided into four domains; planning and preparing, classroom environment, instruction, and professional responsibilities. Seventy-six elements are categorized across the domains, and the elements are described in detail related to knowledge, skills, and dispositions. The framework acknowledges the complexity of teaching, maintains a common vocabulary for professional discussion, and allows for clear assessment across performance as unsatisfactory, basic, proficient, and distinguished. In her explanations describing the frameworks use for supervision and evaluation, Danielson is emphatically clear that it is a collaborative process enacted with the teacher, and it can only occur with trust present between the parties involved.

Entering the twentieth century saw a sharp turn towards studying schools and education as a field as demonstrated by the increased volume of empirical research on the topic. The research included a strong emphasis on clinical supervision, influenced by the medical model. Entering the twenty first century witnessed another sharp turn with the adoption of legislation defining educational practices for years to come.

Twenty First Century

Entering the current century, student achievement and the quality of teaching have been brought to the forefront of public consciousness. Legislation enacted by the Bush administration, *No Child Left Behind* (NCLB), required 100% of students to be performing at statistically impossible levels (Burch, Steinberg, & Donovan, 2007; Forte, 2010; Porter, Linn, & Trimble, 2005; Springer, 2008). Lee and Reeves (2012) in an examination of National Assessment of Educational Progress (NAEP) scores, both pre- and post- adoption of NCLB found that when characteristics of different states were accounted for including

variability of student performance across grades and diverse groups of students, no

sustainable evidence supported the high stakes accountability policies present in the

legislation.

Rather than repealing or adjusting this legislation, the Obama administration added

Race to the Top initiatives along with Arne Duncan's *A Blueprint for Reform* to continue to

add pressure for increased student achievement and other educational reforms including

teacher evaluation by offering exemptions from the NCLB requirements if other targeted

reforms were put into place (United States Department of Education, 2010). The USDE

conducted a two-prong study to determine the effectiveness of the school improvement

grants (SIG) awarded as part of this legislation to states and schools. First, a descriptive

analysis compared 290 SIG schools to 190 non-SIG schools examining what strategies

outlined by SIG guidelines were used. Secondly, a regression discontinuity design was used

with data from 460 schools to determine the effectiveness of strategies used. Key findings

included: (a) no causal link between SIG funding and strategies implemented, (b) no

difference between strategies used for English language learners at SIG schools versus non-

SIG schools, and (c) no significant impact on math or reading test scores as a result of SIG

(Dragoset, et al., 2017).

Through these endeavors, states and school districts were required, among other

reform efforts, to include student achievement measures in teacher evaluation processes. "We

are calling on states and districts to develop and implement systems of teacher and principal

evaluation and support, and to identify effective and highly effective teachers and principals

on the basis of student growth and other factors" (USDE, 2010, p. 4). Another important

emphasis, among these accountability reforms, called for student achievement to be included

in teacher evaluation; new student assessment systems were incorporated at the local and state levels (USDE). The most recent legislation, ESSA (2015), has removed the requirement of student achievement measures in evaluation, but it still encourages their use as a base for incentive or merit pay. Whether used for merit pay or performance evaluation, almost all states require some form of student performance measures to be included as part of a teacher's summative evaluation (Doherty & Jacobs, 2015; Gandha & Baxter, 2016; Hewitt, 2015; Lomax & Kuenzi, 2012).

Teacher evaluation continues to evolve and develop to include a variety of components including student work samples, lesson plans, teacher self-assessments, student growth measures, peer assisted reviews, and supervisor observations with a variety of importance on different components reflective of state, school boards, and negotiated agreements with teachers groups (Brandt, Mathers, Oliva, Brown-Sims, & Hess, 2007; Hanzi & Arredondo, 2009; Mathers & Oliva, 2008; Thomsen, 2014). Regardless of the evaluation practices adopted by school districts, it will fall to school leaders to implement them effectively for the betterment of teacher practice and student growth and learning.

Leadership Practices in Conjunction with Teacher Evaluation

Danielson (2010) established the purpose of evaluation as a growth tool and identified four essential questions every effective growth model evaluation system must have. How good is good enough? Good enough at what? How do we know? Who should decide? Danielson argued, shared understanding and common language are essential to effective communication and improved practice. By maintaining a consistent definition of good teaching, professional conversations with trained evaluators would lead to improved teacher practice and increased student learning. A challenge facing this implementation from my

own experience and those shared with me by fellow administrators is finding the time to implement reflective conversations and teaching practices with trained evaluators, who must engage in reflective conversation to build shared language, understanding, and clear expectations. Teachers share the work of the evaluation by knowing what is expected and participating in the review of the work that has occurred. Moving from the traditional model of passive teacher participation (happening to) to a framework that happens with the teacher through shared observation, conversation, reflection and establishing next steps builds the pathways for increased student achievement.

The following sections examine three broad leadership styles and their relationship to teacher evaluation. Largely, the forms of transactional, transformational, and instructional leadership were selected based on their reflections in the history of teacher supervision and evaluation. Transactional leadership structures were very present in Taylor's scientific principles and the description of principal duties of the early twentieth century. Transformational leadership moves can be seen in the mid to late twentieth century development of targeted feedback in the clinical cycles. Finally, the use of instructional leadership structures is apparent in the framework adoptions and implementations with teachers to improve practice. As will be shown in the discussion of each, the leadership styles continue to be implemented in a variety of forms and situations, but an evolution of emphasis can also be recognized when comparing these practices to the previously discussed developments.

Transactional Leadership

Clearly, teacher evaluation evolved with education reform including leadership practices associated with the evaluation process and relationships between administrators and

teachers. Highlighted in the history of teacher supervision, the practices associated with the process were often instigated and carried through by the administrator. The principal would speak with the teacher, observe the teacher, and potentially speak with the teacher again before offering a summative evaluation of the teacher's performance. Even as the form and processes changed, the core relationship remained the same for much of the history of teacher evaluation. Transactional leadership resided at the core of these practices. Burns (1978) defined transactional leadership in terms of a contract or bartering with the supervisor offering concrete rewards in return for services provided. In this case, the supervisor would offer a beneficial judgment if teacher checks off all the boxes during their classroom performance. As with many different leadership styles and approaches, benefits and detriments arise. The school would likely demonstrate organization, compliance, adherence to policy and uniform delivery of instruction while being inflexible to meeting the needs of teachers and diverse learners.

A meta-analytic study including 117 independent samples over 113 primary studies demonstrated transactional leadership was more effective for individual level task completion (Wang, Oh, Courtright & Colbert, 2011). Grissom and Loeb (2009) conducted a quantitative 42-item task inventory survey study of hundreds of principals and assistant principals and thousands of teachers and parents along with evaluations of participating principals conducted by their assistant principals in the Florida, Miami-Dade County Public School system. The purpose was to determine which principal skills influenced student outcomes; findings suggested only a principal's organizational management skills consistently predicted student achievement growth. More specific to teacher evaluation, policy frameworks are typically overly prescriptive and rigid lending themselves to the organizational skillset of a

transactional leader (Baker, Oluwole & Green, 2013). Also, principals help interpret district and state policies, thereby influencing if and how policies are enacted (Clifford & Ross, 2011; Halverson & Clifford, 2006; Zmach, 2006). Due to the demands across all areas of a school campus, principals spend more time performing management functions and less time on instructional leadership (Camburn, Spillane & Sebastian, 2010).

Unfortunately though, teacher evaluation to this point in time has not derived the expected positive impact regarding the quality of teaching and student achievement, and, likely, this is in part due to its narrow and rigid nature defined by school districts (Baker, Oluwole & Green, 2013; Davis & Annunziata, 2002). A change in approach is suggested in order to spark a change in individuals. Learning organizations, incorporating authentic leadership, built on relationships create opportunities to share the duties and responsibilities of leadership. Leading together as an organization generates the setting to frame dilemmas and address them together as an organization. Fullan states, "The big problems of today are complex, rife with paradoxes and dilemmas. For these problems there are no once-and-for-all answers. Yet we expect our leaders to provide solutions. We place leaders in untenable positions" (2001, p. 2).). Since the transactional leadership practices hold limited success for effective and collaborative evaluation systems, the adoption of other leadership styles occurs. Particularly, transformational leadership, as the name implies, is incorporated to transform teacher practices and performance.

Transformational Leadership

A shift in leadership practices regarding teacher evaluation may occur, and subsequently transformational leadership becomes more common (Bass & Riggio, 2006; Chin, 2007; Fenn & Mixon, 2011; Leithwood & Jantzi, 2006). Burns (1978) characterized

transformational leadership as engaging with followers with a focus on followers' needs and an awareness of the significance of outcomes and new ways those outcomes may be achieved. "In transformative leadership, leaders and followers are united in pursuit of higher-level goals common to both" (Sergiovanni, 1990). Transformational leadership qualities include idealized influence, inspirational motivation, intellectual stimulation, individualized consideration, and reflects a process of mutual influence regarding collaborative leadership and collective leadership (Bass & Riggio, 2006; Fenn & Mixon, 2011; Hallinger & Heck, 2010). Group settings, contextual settings, and including efficacy contribute to more effective components of transformational leadership (Wang, Oh, Courtright & Colbert, 2011).

Principal roles have changed and the roles determine definitions of success (Clifford, Behrstock-Sherratt & Fetters, 2012). As result of a survey of 92 teachers, using descriptive statistics and t-tests, a study determined transformational leadership must be a ready tool of highly qualified leaders (Quin, Deris, Bischoff & Johnson, 2015). The relationship between principals and teachers impacts more than just how they feel about each other and their work place. Teacher efficacy influences principals' actions (Walker & Slear, 2011), and principals are a determining factor in teachers' decisions to join and remain in a school (Boyd et al., 2010; Ladd, 2009; Clifford & Ross, 2011).

A quantitative study of 400 teachers in Texas found transformational leadership correlates to more supportive administrators, more engaged teachers, and fewer frustrated teachers (McCarley, Peters & Decman, 2016). Research has found positive relationships between transformational leadership and school climate (Allen, Grigsby & Peters, 2015). Principals influence the school climate (Williams, 2009). Sebastian and Allensworth (2012) found that differences across schools in instruction and achievement were a result of learning

climate. Hallinger and Heck (2010) conducted a quantitative longitudinal study of 192 elementary schools in a single state over a four-year period and found significant direct effects that collaborative school leadership can positively impact student gains Part of the skillset of transformational leaders includes providing for the needs of staff. New teachers need more support and modeling while experienced teachers need emotional support and inspiration (Walker & Slear, 2011). Thus, another aspect of transformational leadership includes emotional intelligence, the awareness and capability of managing one's own emotions and those of others (Harms & Crede, 2010; Hebert, 2011). Hebert (2011) conducted a quantitative correlational study of emotional intelligence and leadership practices that included 30 principals across elementary, middle and high schools and approximately 180 teachers that worked with those principals. She concluded a positive relationship exists between emotional intelligence and transformational leadership while no significant relationship existed between emotional intelligence and other leadership styles. A meta-analysis of sixty-two samples representing 7,145 leaders found a moderate relationship between emotional intelligence and transformational leadership, and while lower than expected, some of the variation was attributed to the wide variety of tools used to measure emotional intelligence (Harms & Crede, 2010). Fenn and Mixon (2011) determined that transformational leadership could be taught to and learned by school leaders as part of their survey study of 215 Texas superintendents that attended an annual leadership academy over a ten-year period between 2000 and 2010.

Not only do administrators need to learn the skills of effective leadership, but also they must learn to use the tools and policies of their respective districts and buildings. For every school district there is a different evaluation tool, process or framework. Even those

districts adopting published evaluation frameworks and being trained to use the tools will develop at different rates of speed, fidelity to model, and validity of use. Teachers' perceptions and satisfaction with different models could influence their effectiveness. A case study of 229 kindergarten through fifth grade elementary teachers was conducted in three rural districts in Northwest Georgia to examine the satisfaction of teachers across three plans for evaluation and supervision. Arp (2012) examined five questions using two previously published online surveys:

> Is there a difference in perceptions of teachers' satisfaction across three districts using different plans for evaluation and supervision? Is there a relationship between the amount of formal observation and teachers' perceptions of satisfaction with the amount of supervision? Is there a relationship between three districts and teachers' perceptions of overall satisfaction? Is there an interaction effect between the three districts and teachers' years of teaching experience on teacher participation? How do district plans for evaluation and supervision align with models found in literature? (p. 20-21)

The three plans in place were the CLASS keys, Georgia Teacher Evaluation Program, and a locally created plan implemented by y a charter school. Document analysis was used to determine type of supervisory model that aligned with each of the three district plans. One district incorporated collaborative supervision, another used directive supervision, and the third formulated a non-directive supervision plan. The study found no significant difference in perceptions of teacher satisfaction across the three districts. Teachers' perceptions of satisfaction were dependent on the amount of time the district spent in formal supervision. Teachers were more satisfied when participating in non-directive or collaborative supervision rather than directive supervision and preferred to be involved in the plan for supervision and evaluation. Time was not a contributing factor to teacher satisfaction as long as they were involved in the plan through such options as pre/post conferences. In short, they considered the feedback obtained from formal supervision valuable. This study presented a contradictory

view to often negative perceptions of teacher evaluation, which can be attributed to limited attention to context and leadership.

Recalling the purpose of evaluation is to improve teacher performance in connection to increased student learning, it is important to look at leadership characteristics that positively influence classroom practice. Blasé & Blasé's (1999) early research concluded that leaders need to talk with teachers to promote reflection and professional growth. The study sample consisted of male (N = 251) and female (N = 558) elementary, middle, and high school teachers from multiple rural, suburban, and urban school sites with an average age of 37 and 11 years of teaching experience. The Inventory of Strategies Used by Principals to Influence Classroom Teaching (ISUPICT), an open-ended questionnaire, was designed and pilot tested with 30 full-time teachers. A methodological perspective emphasizing the study of human perceptions and meaning that people construct in their social settings was maintained. Findings suggested effective principal and teacher interactions about instruction, inquiry, reflection, exploration, and experimentation result in teachers building toolboxes of alternative strategies.

As highlighted in this section, transformational leadership strengthens positive relationships between teachers and supervisors, creates opportunities for collaboration and discussions for improvement. Transactional leadership reflects tools for maintaining a well-managed, organized school. To reach a level of instructional improvement, leaders must not only incorporate transactional and transformational leadership, but likely need to adopt instructional leadership practices as well.

Instructional Leadership

Robinson, Lloyd and Rowe (2008) found five dimensions of leadership hold the greatest impact when analyzing effect sizes among comparable studies versus comparison across different types of leadership studies. The dimensions included establishing goals and expectations, resourcing strategically, planning, coordinating, and evaluating teaching and the curriculum, promoting and participating in teacher learning and development, and ensuring an orderly supportive environment. The study also determined the relative impact of instructional and transformational leadership. Instructional leadership was almost four times more powerful, but one of the limitations of the study, the process for selecting and locating studies to include for meta-analysis, could have skewed the results. Of the 27 studies included, the transformational leadership measurements predominantly focused on organizational change while the instructional leadership elements focused on student achievement gains. Another important finding identified that the greatest growth would be a combination of instructional and transformational leadership.

Modern evaluation procedures at times utilize a framework for teaching as a foundation for making judgments and providing feedback. One of the more commonly adopted was Danielson's framework for teaching as a standards-based evaluation framework, originally published in 1996 and then updated in 2007 (Danielson, 2011). Danielson's framework includes four domains describing teacher responsibilities that include a total of twenty-two components, which are further broken down into seventy-six discrete elements, strategies, or techniques of instruction and professional responsibility. Supervisors use rubrics for the varied elements to rate teacher performance and provide feedback for improvement. These ratings can also be compiled to provide an overall rating for teachers

that can be combined with student achievement data for a summative rating of teacher performance (Danielson, 2011).

Kimball and Milanoski (2009) studied the validity of teacher evaluation and leadership decision-making within a criterion referenced system, using models as a standards-based evaluation framework. In particular, they looked at how much the validity of the performance rating relationship varied across evaluators. The study compared the variations from evaluator to evaluator with student achievement measures to identify evaluator accuracy. Student achievement measures for reading and math were generated through a mathematical formula to find the average student achievement for each classroom teacher. From the evaluators' scores of teachers with the framework and student achievement scores, correlations were drawn between more and less valid evaluators. A semi-structured interview protocol was used to explore evaluator will, skill, and context. Kimball and Milanoski concluded that there was a large amount of variance between evaluators, confirmed by additional statistical analysis beyond what was originally designed. Also the three domains identified, will, skill, and context, were not enough to fully explain the decision-making process. Evaluators in both the high and low validity groups identified a more formative evaluation goal. They wished to better develop the teachers they were working with rather than make summative judgments. The study recommended extensive training and establishment of inter-rater reliability before using results to make high-stakes decisions about personnel. Future research should consider the extent to which teachers are trained in the use of the evaluation framework. Another avenue to explore would be the extent to which the evaluation framework increases understanding and performance if teachers are trained with evaluators or separately.

The principal plays a key role in a school's success, but many districts' policies and procedures are very prescriptive regarding principal actions and responsibilities (Baker, Oluwole & Green, 2013). Ovando and Ramirez (2007) utilized a multiple case study approach to investigate six elementary, middle, and high school principals' and assistant principals' perspectives on what instructional leadership actions they could take to improve student performance within the context of their district's evaluation system. Instructional leadership was defined by a comprehensive understanding of knowledge, skills associated with teaching and learning and the ability to coach and improve educators' practices. The key research questions investigated were: What are school principals' perceptions regarding their actions within the performance appraisal of teachers to improve instruction in successful schools? What are the similarities and differences in principals' perceptions by school level? Triangulation was accomplished through multiple data sources consisting of interviews, observations, and journaling. Data were analyzed inductively to identify significant patterns to create a framework to communicate the essence of the data. The study found instructional leadership actions at all levels included setting clear expectations, monitoring instruction using walk through observations, and providing professional development opportunities according to teachers' needs (Ovando & Ramirez).1Administrators relied on instructional leadership actions pertinent to each grade level's needs and applied the performance appraisal system as a basis to enhance instruction and improve student achievement. Essentially, the study emphasized instructional leadership actions in relation to teacher evaluation targets to enhance teachers' instructional capacities regarding student learning.

Instructional leadership holds the power to improve schools (Seashore Louis, Dretzke & Wahlstrom, 2010). A longitudinal qualitative inquiry designed to examine the connections

between leadership behaviors and student achievement used full day, in-person observations of 100 principals across three years. Findings indicated if principals spend their limited time on broad instructional leadership actions then that investment will not pay off in student achievement, but targeted instructional investments seemed to predict year to year gains in student achievement. Walkthroughs, as examples of broad instructional supervision, had a negative prediction of student achievement gains. Time spent on teacher development, evaluation and coaching paid dividends. (Grissom, Loeb & Master, 2013). Job embedded professional development provides the mechanism to shift from punitive evaluations to improved teaching and learning (Coggshall, Rasmussen, Colton, Milton & Jacques, 2012; DuFour & Mattos, 2013). The opportunity to improve teacher practice exists without requiring dismissal for poor performance. Relationships between supervisors and teachers influence interpretations of messages about instructional leadership and how they implement the evaluation process (Rigby, 2015). "Using what is known about how teachers learn and creating evaluation systems with integrated opportunities for aligned job-embedded professional learning that is more learner-centered, knowledge-centered, community-centered, and assessment-centered will more likely capture the energy-generating potential of teacher evaluation reform " (Coggshall, Rasmussen, Colton, Milton & Jacques, 2012, p.21). Within a school, the principal's leadership influences achievement indirectly through quality professional development and fidelity of programs (Sebastian & Allensworth, 2012). To improve school leaderships' development, a tool for measuring instructional leadership effectiveness should be developed (Porter et al., 2010).

There also continues to be a shortage of people willing to assume the role of principal (Clifford, Behrstock-Sherratt & Fetters, 2012; Pijanowski, Hewitt, & Brady, 2009). A study

of 11 key industries' management practices, found the education field was the least likely to intentionally work and plan to hire and retain quality talent (Ringo, Schweyer, DeMarco, Jones, & Lesser, 2008). High need, low performing schools in particular are in short supply for strong leadership (Clifford, Behrstock-Sherratt & Fetters, 2012; Horng, Kalogrides, & Loeb, 2009; Papa, Lankford, & Wyckoff, 2002; Rice, 2010). In addition to highlighting the shortage of principals in high need schools, Rice (2010) presented to the National Center for Analysis of Longitudinal Data in Education Research (CALDER) that (a) principal experience is a predictor of effectiveness, (b) principal efficacy is a predictor of what tasks are given time as well as (c) what approaches are taken toward those tasks. Essentially, principals spend their time where they are confident. Additionally, principals require support to address their knowledge and management concerns in order to implement education reforms such as new, high stakes evaluation systems (Derrington & Campbell, 2015).

In an attempt to gauge the usefulness of a daily log for measuring principal leadership practice, Camburn, Spillane and Sebastian (2010) examined 50 urban principals as part of a mixed-methods study that included daily logs, observations, interviews, and an experience sampling instrument. As a result of their work, they confirmed the bulk of a principal's time is spent on managerial tasks, and the log was a useful tool when consistent, direct, and strategic. Contact and discussion of the contents of the log occurred as a follow-up process. Like with teachers, in order for principals to improve, they need targeted feedback to increase instructional leadership practices. Principal evaluation instruments include limited coverage of behaviors regarding curriculum and instruction, and often assessments of leadership lack documentation or do not occur (Goldring, Cravens, Murphy, Porter, Elliott & Carson, 2009).

Similarly, principals are inconsistent in their practice of teacher evaluation. Principals

evaluate inconsistently based on understanding, relationships and experience (Rigby, 2015). Principals will use the highest ratings when they may not be warranted and inflate ratings to preserve relationships (Sartain, Stoelinga & Brown, 2011). Allen, Grigsby, and Peters (2015) conducted surveys of six principals and almost 60 teachers in a small school district in Texas to examine the correlations between principals' transformational leadership and student achievement. The primary findings included that principals did not directly influence student achievement in reading and math, but did influence school climate. An interesting by-product of the study found that very little correlation occurred between principals' perceptions of their own transformational leadership actions and teachers' perceptions of their leadership. Principals need to improve conversations with teachers about improving teaching (Sartain, Stoelinga & Brown, 2011).

Principals hold direct and indirect influence over the school (Clifford, Behrstock-Sherratt & Fetters, 2012). Teachers make decisions about where to work based on school principals, and more effective principals are able to recruit and retain more effective teachers (Rice, 2010). Rubrics of instructional implementation can assist in creating shared language and conversations that lead to improved instruction, but ultimately the tool is only as good as its users (Sartain, Stoelinga & Brown, 2011). Principals receive a variety of messages regarding teacher evaluation (Rigby, 2015). When principals were evaluated on their ability to assess teacher performance, they were more likely to invest more time on that duty (Sun & Youngs, 2009). Many states mandate student achievement data be incorporated into teacher performance evaluation (Piro, Wiemers & Shutt, 2011). Variations occur between evaluator ratings and value-added measures of teachers, and evaluators' decisions were complex combinations of motivation, skill, and context (Kimball & Milanowski, 2009). Principals'

experience and insight offers valuable insight to teacher performance beyond summative assessment results (Rice, 2010).

High performing school principals employ all leadership practices more than principals in low performing schools, and a shared vision along with challenging the accepted processes are the most effective practices to increase student achievement (Quin, Deris, Bischoff & Johnson, 2015). Both Transformational and Transactional Leadership are needed, and training for both is significant as well (Pepper, 2010). Meaningful reform efforts and positive school outcomes result from a school leader able to act with emotional intelligence, manage relationships effectively, understand the larger picture, and be flexible with change (Saxe, 2011).

In summary, most evaluations occur in a transactional construct with narrowly defined expectations, timelines, routines, and required forms. Transformational practices increase professional development and are preferred by teachers interacting in evaluation procedures. Research continues to investigate the use of instructional leadership practices along with other forms as teacher evaluation continues to evolve. Using frameworks to establish common language for reflective conversation in a more formative evaluation format could very likely lead to discoveries and developments in evaluation practices that have been lacking up to this point, namely consistent improvement of teacher practices across school districts. Largely lacking in this discussion has been the leadership practices directly connected with incorporation of student growth measures into summative evaluations. This is a noticeable area for potential future research. Some studies reference instructional leadership as those practices leading to increases in student achievement on standardized assessments, but that was not the categorical definition for this review. Even with that particular

perspective, it still would not address specific actions implemented in systems where student growth measures are adopted into the performance evaluation of instructional staff. A component of this proposed study includes teachers' perceptions of instructional leadership actions in regard to student growth measures as a component of the performance evaluation. When considering teachers' perceptions, an examination of teacher influence and efficacy or lack thereof arises. What do teachers think of the educational reforms occurring around them?

Teacher Power

Teachers' perceptions can be powerful. Do teachers perceive events and acts unfolding as happening to them? With them? For them? Because of them? On the day-to-day experiences, teachers likely experience a variety of controls or lack thereof regarding their classrooms, schools, districts, and the field as a whole. In examining education reforms, specifically teacher evaluation reform, what roles do teachers play? What roles can they play? Do they feel they have the power or voice to participate? As evaluation reform pushes through, are teachers steering the vehicle, being rundown by the vehicle, or do they feel they are outside of the event altogether? To that end, the topics included in this section are unions, politics, teachers' perceptions, and improving performance. On one side, when teachers' voices are quiet or feel overwhelmed by the changes around them, what circumstances generate such an effect? Equally important, what do teachers' believe will improve their performance? Throughout this section, consideration is given to typical teacher power constructs such as unions, collective bargaining and political arenas. Either through their strength in numbers or through political lobbying and campaign, teachers have been able to exert influence and raise their voice on topics that matter to them.

Unions

Typically when considering teacher power, unions come to mind, strength in numbers. Commonly, unions serve five purposes: (a) raising wages, (b) growing membership, (c) increasing share of representation of labor pool, (d) prevent pay based on performance, and (e) minimize competition from non-union entities (Coulson, 2010). Increasing their share of the labor pool and minimizing competition of other entities serves the union as an organization more than directly supporting teachers. Wages, some could argue, is the primary goal of unions, which includes benefits and working environments. While unions vary in strength from district to district and state to state across the country, some trends are common. School districts with stronger unions' pay scales favor experienced teachers, while districts with weaker unions favor new hires in their salary schedules (Guthrey, 2018). Union strength is associated with education expenditure increases and increased salaries for experienced staff while holding a slight negative relationship with student achievement (Cowen & Strunk, 2015). School districts with strong teacher unions used acquired state and federal aid towards teacher compensation while districts with weaker unions used acquired aid to reduce tax burdens and hire new teachers with the former demonstrating student achievement gains over the latter (Brunner, Hyman & Ju, 2018). Coulson (2010) argues unions resist differentiated pay or merit pay, as it is a threat to collective bargaining. As often happens when discussing money in education, the question comes up, does increased spending or wages increase student achievement?

A popular pro-union argument includes stronger unions affect student achievement by reducing class sizes, but it is difficult to connect unionization and student achievement results directly due to the varied definitions and measures of student achievement and varied union

power (Guthrey, 2018). Guthrey (2018) also noted, districts that unionize have higher dropout rates, but there have been no causal links between these simultaneous occurrences. School districts with strong unions dismiss more low-quality teachers while retaining more high-quality teachers resulting in higher average student achievement, due to higher wages holding more influence for recruitment and retention (Han, 2015).

Anti-union political movements frame the argument of disadvantaged kids against greedy teachers to push political agendas (Smith, 2015). Levinson and Theisen-Homer (2015) conducted an examination of teacher reductions in force in Los Angeles based strictly on teachers' value added measure (VAM) performance. They concluded the fairest method would have been to include student evaluations, administrator evaluations, value-added measures, and staff seniority. The authors of the study emphasized the need to recognize students and teachers as allies not combatants in the debate or negotiation of how to handle such matters. When unions and school districts become entrenched combatants both sides lose, but when working with each other, both sides have the possibility to gain.

In order to hold a position of power in collective bargaining, membership and participation represent strong indicators of a union's strength. Union membership is highest in those schools and districts hardest to staff and with the poorest working conditions (Fowles & Cowen, 2015). Pogodzinskie and Jones (2014) discovered unions might need to evolve to meet the needs of the new generation of teachers based on surveys of 184 novice teachers across 11 school districts in Michigan and Indiana along with a similar number of surveys of experienced teachers from the same districts. Novice teachers held different views and perspectives of education reforms such as welcoming teacher accountability and considering unions less of a necessity for their career compared to the more experienced

educators (Pogodzinskie & Jones, 2014). Wage compression, a result of union collective bargaining, is less appealing (Coulson, 2010). While often the relationship between school districts and unions plays out in collective bargaining, it also occurs at a local, state, and federal political levels.

Politics

Unions lobby and campaign for candidates spending huge money (Coulson, 2010). Measuring teacher union political activity, based on contributions to candidates for state office, found higher political activity predicts fewer education reform policies such as merit pay and vouchers from occurring (Hartney & Flavin, 2011). As a result of legislation passed in Wisconsin to limit collective bargaining and union contributions to political entities, teacher turnover increased, teacher salaries decreased, and student achievement declined (Baron, 2018). A common accusation exists that the presence of unions and collective bargaining at the state level leads to increased education spending, but an examination of pre- and post- state policies implementing collective bargaining shows these states did not increase education spending (Paglayan, 2018). States that enact mandatory bargaining laws create contexts forcing the creation of unions; which in turn, become political entities that lobby government, support political campaigns both verbally and financially, and act as an activist instigator encouraging teacher members to participate in political rallies and protests (Flavin & Hartney, 2015). Unions declare teacher voices have not been lost but are hibernating, and in order to be heard during the current climate, teachers must become more politically active and collaborate with national networks (Baird & Heinen, 2015).

Education has become politicized, and the representation of unions and collective bargaining is used by pro-reform organizations and candidates to target Democratic support

and allies (Wade, 2015). Collective bargaining agreements maintain status quo and resist reforms and changes over large periods of time (Cowen & Fowles, 2013). Union political contributions and collective bargaining are associated with higher public employee wages and more employment for state and local government including teachers (Crowley & Beaulier, 2018). Teacher pay, benefits and employment have shown little influence from collective bargaining even as union presence and participation increases among teachers. Meanwhile, firefighters have seen large increases in pay and police officers have received shorter work weeks for slightly more pay (Frandsen, 2016). Higher needs schools, disadvantaged by poverty, are forced to higher less experienced and capable teachers due to seniority based transfer rights, often negotiated provisions in larger, urban districts that allow more experienced teachers to transfer to more affluent schools (Anzia & Moe, 2014). While having found collective bargaining to cause teacher quality gaps in larger, high-needs school districts due to seniority-based transfer rights, Koski and Horng (2014) identify the lack of generalizability in their work and those of similar studies needed to encourage the elimination of such provisions. Consistently, with-in district transfer rates are higher for senior staff with disadvantaged students than novice teachers working with similar students when strong seniority transfer provisions are part of the collective bargaining agreement. Related, senior teaching staff members tend to remain at advantaged schools when strong seniority provisions are present in the collective bargaining agreement (Goldhaber, Lavery & Theobald, 2016). Strong, formal institutional-union partnerships and collaboration predicts positive student achievement over time (Rubinstein & McCarthy, 2016). The question then arises; do teachers believe such partnership can exist?

Teachers' Perceptions

Cooperation, while possible, is difficult when goals and perspectives differ among the negotiating parties. Current reforms, based on legislation actions, (NCLB, RttT and ESSA) encourage teacher evaluation as a tool to improve student achievement. Teachers and their representative unions may view teacher evaluation as a weapon to weaken their autonomy, authority and ability to deliver services to students. Studies of teachers' perceptions found teachers feared implementation of summative evaluations based on student metrics because they feared new systems would contribute to control and sanctions of their practices (Finnegan, 2016; Flores, 2012; Morgado & Sousa 2010; Tuytens & Devos, 2009). Increases in student achievement, as part of education reform, most likely occurs when principals and teachers cooperatively work to improve instruction through the use of pre- and post-conferences of classroom observations (Abdo, 2017; Mette, Range, Anderson, Hyidston & Nieuwenhuizen, 2015). Maine's teacher and principal evaluation program experienced success such as deeper understanding of professional models, use of evidence, reflection, and feedback, and more consistent evaluation practices (Fairman & Mette, 2017). Fairman and Mette also noted Maine's teacher and principal evaluation program had multiple challenges such as time demands, technical questions regarding student assessments, need for to improved calibration, and evaluation of other administrative roles.

Even when districts and schools use similar evaluation models and ratings, implementation varies across schools (Abdo, 2017; Riordan, Lacireno-Paquet, Shakman, Bocala & Chang, 2015). Teachers disagreed while administrators agreed with the belief that new evaluation systems accurately capture teacher performance using student-learning outcomes as a major component (Finnegan, 2016). The ability of administrators and systems

to maintain fidelity to evaluation models influences teachers' trust of the evaluation system. Both administrators and teachers feel an effective evaluation tool consumes too much time and resources, and lacks needed training (Riordan, Lacireno-Paquet, Shakman, Bocala & Chang, 2015). Evaluation reforms that create common frameworks and language improved principals' feedback and conversations with teachers, but were so demanding for time, they were often implemented poorly, accompanied by poor training at the beginning of the process (Kraft & Gilmour, 2016). Teachers' understandings do not necessarily align with principal's intended messages (Abdo, 2017). Some districts look to train teachers as evaluators or hire third party organizations to conduct evaluations (Fairman & Mette, 2017). Both teachers and administrators agree the professional standards are a useful and accurate tool, and both groups are less supportive of the incorporation of student achievement data (Finnegan, 2016).

The purposes of supervision and evaluation are different but overlap greatly. Principals must employ instructional leadership skills and provide professional development during supervision in order for principals and teachers to gain effective and accurate evaluative summaries of performance (Abdo, 2017; Mette, Range, Anderson, Hvidston, Nieuwenhuizen & Doty, 2017). In cooperation with the teachers' union, Denver adopted a pay for performance system as a pilot and then later as a district wide tool. As a result student achievement increased compared to before, but the district was unable to separate causes of gains from pay for performance or other district initiatives occurring at the same time. Some bonuses appear targeted and effective while others do not (Goldhaber & Walch, 2012). Teacher evaluation reform policies including student growth measures do more harm than

good while effective strategies to improve teacher performance such as mentorship practices are ignored (Abdo, 2017; Smith & Spatariu, 2015).

This brings the focus back to school settings where teachers handle the day-to-day business. Teachers may look to unions, politics, and collective bargaining to influence larger needs or exert more influence on the system as a whole, but at the school level multiple variables still provide context for teacher power. Administrators greatly influence teachers' decisions to stay or leave through their relationships and by creating and managing the day-to-day working conditions (Boyd, Grossman, Ing, Lankford, Loeb &Wyckoff, 2011). Teacher confidence and competence are also reflected in student performance. Elementary school teachers demonstrate modest statistical significance by teaching specialized subjects rather than all subjects when examining students' reading and math scores (Goldhaber, Cowan & Walch, 2013). Teachers' background and education contribute to feelings of confidence as those teachers who attended more prestigious universities were more willing to support teacher accountability and charter schools (Bell, 2018). Age and generation contribute too as new teachers see new evaluation tools as an opportunity for accountability and alignment (Bush, 2017). Donaldson (2012) conducted a study in a medium-sized, urban school district by interviewing 92 teachers and school leaders during the second year of implementation of a new teacher evaluation system. Donaldson found (a) teachers were positive about setting own goals and working towards them, (b) teachers agreed evaluation reform was necessary, (c) teachers did not want to have a system imposed, but would rather have input on the creation of a tool, (d) teachers disagreed about the fairness and objectivity of the new evaluation program, (e) teachers with high marks were more positive in general, and teachers with low marks were more negative in general, and (f) the majority of teachers

said the evaluation program did not influence their lesson delivery, but for some it did change their planning practices (Donaldson, 2012). Teachers that are more experienced, Democratic, or unionized typically rally against reforms while teachers that are less experienced, Republican or non-unionized are more supportive of reforms (Bell, 2018).

In contrast Satty (2016) conducted a smaller, similar interview study to Donaldson's (2012) in Broward County, Florida and found teachers believed their new evaluation structure, recently introduced for performance pay, (a) was enacted unfairly by evaluators, (b) was inaccurate, (c) caused stress and anxiousness, (d) did not accurately measure student gains, (e) and triggered overall misunderstanding of the new system. A study of 12 teachers at a Midwest high school across a variety of experience levels discovered the implementation of a new teacher evaluation model led to teachers' perceptions and beliefs of (a) lack of trust in the state's understanding of the school's needs, (b) teachers lack of control or input in development or implementation of the new model, and (c) feelings of mistrust and fear of ulterior motives (Bush, 2017). A separation of experience was noted in this study as the seasoned and mid-career teachers saw the new tool as a negative occurrence; the new teachers viewed the tool as an opportunity of accountability and alignment of practice. Overall, teachers hold a negative opinion of new evaluation systems (Case, 2016). Teachers feel they have lost control as a result of federal and state mandates regarding practices such as evaluation reform, student achievement testing, and subject emphasis on math and reading to the exclusion of other content (Vardas, 2014). Policies and human resources procedures devalue people and create distance between the acts being evaluated and the results that occur (Hardin-Bartley, 2014). So, as a collective group, teachers generally demonstrate dislike and skepticism towards the adoption of new evaluation practices, but within the larger

context new teachers and confident teachers can find the positive outlook while experienced and less confident teachers harbor strong concerns.

Adding performance pay to the conversation appears to only contrast and highlight those differences among teachers. An analysis of 3,264 teacher surveys about tenure, accountability, and school choice highlights those divisions (Bell, 2018). Bell found teachers, more confident in ability or having trained at a more prestigious university, were more willing to trade tenure for increased pay. Additionally, high school teachers were more likely to support merit pay than elementary teachers. The performance pay system established in Washington DC called "IMPACT" demonstrated an increased likelihood of voluntary resignation when receiving poor performance evaluations, and those that remained demonstrated increased performance. Also, those receiving incentives demonstrated further improved performance (Dee & Wyckoff, 2015). Interestingly, a group of economists examined pay for performance from a different angle by specifically comparing two groups of teachers. One group received their monetary incentive for student performance at the beginning of the school year had to pay it back if students did not achieve results, loss framed incentive. The other group received a gain framed incentive, in that the teachers received a bonus at the end of the year if their students met achievement marks. Loss framed incentives showed statistical significance, while gain framed incentives did not show statistical significance (Fryer, Levitt, List & Sadoff, 2012). Framing such transactional interactions provides insight into a very narrow aspect of improving teacher performance, but teachers demand much more from leaders in order to reach their fullest potential.

Improving Performance

While wages often arise as a topic when discussing the wants, needs, or in some cases demands of teachers, specifically when discussing improving teacher practice, feedback looms are needed as vital tools for teacher improvement. Teacher evaluation that includes collaborative, constructive feedback, routine discussions of student data and performance, as well as structured peer observations can improve teacher quality and efficacy; but the implementation of these structures requires skillful leaders, abundant resources, and policies and procedures that allow for site-based control (Reinhorn, 2015). Administrators need to provide consistent and constructive feedback, limit subjectivity, and improve the school culture and climate related to evaluation procedures to ensure teacher effectiveness and success (Shugart, 2017). Teachers receiving performance- based pay based on principal observations and one-on-one conversations increased their constructivist instructional practices (Liang & Akiba, 2015). A mixed-methods study of a New Jersey's teacher evaluation system implemented across fourteen high needs charter schools determined data were used to inform collaborative feedback between principals and teachers and to guide professional development, but barriers in the form of time, knowledge, and training existed for both administrators and teachers (Mathews, 2017). But not all teachers agree, as with new and mid-career teachers feedback is needed to improve practice, while some seasoned teachers do not think it is necessary (Bush, 2017). Similarly, Braslow (2016), in his study including interviews of 10 teachers from Massachusetts and 20 teachers from Georgia, whom all had years of experience spanning previous evaluation models as well as more recently adopted evaluation tools, observed experienced teachers: (a) did not feel feedback improved their instruction; (b) when feedback was provided, it may have been prompt and accurate but

lacked next steps needed to improve; (c) prompted emotional responses rather than reflective consideration; and (d) felt the intention of administrators was to make what was wanted obvious, but they did not regularly achieve that practice.

As referenced in the previous section of leadership practices, it can be speculated regarding how much of the teachers' ability to improve or practice power is limited by effectiveness of school leaders. From school to school and district to district, principal leadership varies greatly. Even when a group of administrators are trained together using the same evaluation model, their delivery and explanation of that model to teachers varies greatly (Byford, 2018). The observational components of teacher evaluation systems are generally inconsistently addressed concerning the variable reliability and validity of scores from administrator to administrator (Herlihy et al., 2014). Variability is not limited to administrators though, as teachers vary within their own practice as well. Using standardized test data for 132 teachers for five consecutive years, researchers determined that neither teacher performance, nor effectiveness, were highly consistent over the course of study, and it was determined teacher performance was likely inflated at low performing schools (Morgan, Hodge, Trepinksi & Anderson, 2014). The changes in students affect the variability of student achievement success. Teachers receiving the highest performing students from the previous year were more than twice as likely to receive the highest ratings possible, while the teachers receiving the lowest performing students had an increased likelihood to receive the lowest ratings possible (Steinberg & Garrett, 2016). To further complicate the matter, Taylor and Tyler (2012) found teachers increased productivity during the year of evaluation, but productivity as measured by student performance was even greater the following year as a

result of having had the opportunity to receive feedback and implement new reflective information and strategies.

Recognizing feedback, coaching, collaboration and opportunities to practice can improve teacher practice, principals and teachers, together, could improve student achievement with a focus on professional development. Braslow (2016) in a previously discussed study also found teachers reported no changes to professional development as a result of principal observations of their instruction. Moreover, teachers and administrators disagree on what professional development topics are needed (Byford, 2018). In regard to the actual evaluation tools, teachers need professional development about evaluation use, assessment use tied to evaluations, and assessment strategies to improve teaching and learning in order to ensure effective evaluation systems (Shugart, 2017). Training is required in order to determine if evaluation systems have the potential to meet intended goals (Wright, 2015). When considering what and how much professional development should occur, it can become a measure of time and investment of time.

Consistently when examining challenges to successful evaluation implementation, time for learning, implementing, and reflecting is limited. Doty (2018) maintained that during the task of piloting new teacher evaluation systems, perceptions were positive regarding a focus on professional development, but the scarcity of time negatively impacted the ability of administrators to provide constructive feedback, and the student growth measure mandate was unhelpful in teacher development. Peer observation as a form of professional development was perceived as effective and created opportunities for teachers to critically reflect and collaborate with peers, but the process was challenging due to the time needed to participate in the process (Klingelhutz, 2017). Administrators lacked time to

conference during the evaluation process, skill to conference effectively with teachers about performance, and insufficient content knowledge to improve teaching of subject knowledge (Wright, 2015). Teachers and administrators need time and training in order to implement an effective evaluation system that improves student achievement (Raymond, 2017).

A lack of time is not the only challenge present. Another finding from Braslow's (2016) study was the new evaluation tool led to divisions of staff either from competition or from differences in ratings and hurt feelings. Byford (2018) also reported misalignment of teachers' and administrators' beliefs of what makes an effective teacher leads to barriers difficult to overcome. Principals and teachers held different understandings regarding how evaluation data are used with administrators focused on the removal of ineffective teachers and teachers wanted to improve ineffective teaching practices (Byford, 2018). When examining New York State's legislation requiring professional performance review of teachers, the administrators consistently believed teacher practice and performance had improved by increasing student engagement while teachers believed instruction had not improved as teachers now taught to the test (Kramer, 2016). Additionally, the possibility of bias or prejudice can come into play regarding who is evaluated. Smith (2015) found disproportionate numbers of African-Americans, women, higher paid teachers and those over the age of 55 teachers were often referred to peer assisted review (PAR). Those participating in PAR suffered from high stress, depression, insomnia, hostility, high blood pressure and other similar health conditions and negative perceptions. . Vague policy language open to interpretations of others was blamed for biased and ineffective implementation of PAR procedures.

Revisiting the questions do teachers perceive events and acts unfolding as happening to them? With them? For them? Because of them? From my experience, I find it depends on the teacher as an individual, their background, education, experience, relationships with others, politics, and so many other variables. Teachers' perceptions of their own power are greatly influenced by the contexts and circumstances around them. Confident teachers, new to the career teachers, younger generation teachers, Republican teachers, and non-union teachers likely view reforms, accountability, value-added measures as positive steps towards improving the field of education and their own practices (Bell, 2018; Bush, 2017; Pogozinskie & Jones, 2014). Alternately, experienced, older teachers, Democrats, and unionized teachers may be more likely to push back against evaluation reforms and changes (Bell, 2018; Bush 2017; Case, 2016; Donaldson, 2012; Hartney-Flavin, 2011). Regardless which group wins or loses in this area, to be effective and improve instruction, teachers and administrators alike need to participate in professional conversations, professional development, active dialogue with each other, provide and be open to feedback, and ultimately invest the time needed to improve (Fairman & Mette, 2017; Hardin-Bartley, 2014; Kraft & Gilmour, 2016; Riordan, et al., 2015; Shugart, 2017). Unfortunately, there is no student assessment or measure of performance that can make any of those happen, but often, student achievement can be used as a catalyst to start those conversations for improvement. The question comes back to how are such student measures used?

Student Growth Measures (SGM)

At the root of this study is an examination of teachers' perceptions of the inclusion of student growth measures in performance evaluation. To that end, who, what, and why use student growth measures, and why might they be avoided? This section takes a closer look at

what is meant by student growth measures or other like terms. Also, why have they become an emphasized tool in today's reforms? Why are some concerned about the use of these measures? And finally, who has already adopted them or is about to do so? First the what, then the why and finally the who.

What are Student Growth Measures (SGM)?

A variety of terms reference and describe the metrics and methods for objective teacher evaluation such as value-added measures, student growth measures, and student learning objectives. A Value-added Measure (VAM) was succinctly described by Goldhaber (2015) as "an objective measure that does not rely on human interpretation of teacher practices, and by design, it is a system in which teachers are evaluated relative to one another rather than relative to an absolute standard" (p. 88). Student learning objectives (SLO) are a method of connecting student performance results to teachers in non-tested grades or subjects for the purpose of evaluations (Marion, DePascale, Domaleski, Gong, & Diaz-Bilello, 2012). Just using student assessment data does not automatically create the use of value-added measures, as some assessments do not measure student growth or even expected growth, but the assessments measure performance in relation to other factors such as grade level standards (Lomax & Kuenzi, 2012). Student growth measures (SGM), for the purposes of this study, catches the core meaning of VAM, SLO, and other references as using student assessment data meant to measure student growth in achievement and learning over time and applied to measuring teacher performance in an objective manner. Measuring teacher performance brings the focus to why SGM would be used.

Rationale for use of SGMs

No Child Left Behind (NCLB) and Race to the Top (RttT) legislation both required

more rigorous teacher evaluation processes, eventually requiring the adoption of the use of

SGM became a requirement (Duncan, 2010). ESSA (2015) while removing the requirement

after many states had already implemented the practice, still encourages the use of such

metrics for merit pay for educators. Prior to educator reform efforts through legislative

initiatives, the collective research demonstrated regularly that teacher evaluation systems

failed on many counts, such as: (a) poor evaluation instruments, (b) limited district guidance,

(c) lack of evaluator time, (d) absence of quality feedback, (e) few consequences for poor

performance, (e) deficiencies in evaluator skill, (f) dearth of evaluator will, and (g) obscene

inflation in teacher ratings so that more than 90% received the highest ratings possible

(Donaldson, 2009; Donaldson, 2010; Marzano, 2012). Qualifications used to determine a

teacher's certification and salary, with the exception of years of experience, have exhibited

inconsistent or no influence on student achievement (Rockoff & Speroni, 2010).

The importance of the value of teachers is consistently noted in the literature; higher

teacher quality results in higher student achievement, particularly in contrast to other school

interventions (Aaronson, Barrow, & Sander, 2007; Goldhaber & Hansen, 2010; Rivkin,

Hanushek, & Kain, 2005). Chetty, Friedman, and Rockoff (2012) found students of high

value-added teachers are more likely to attend college, attend more prestigious universities,

earn higher salaries, live in better neighborhoods, save more for retirement, and less likely to

become teen parents. The quantitative study began with more than six million student-year

data sets that were created by linking individual student achievement scores in reading and

math within cohorts of students across multiple school years. Of these, 2.5 million children

from a variety of school year cohorts were linked to parental tax returns, school and school

district teacher performance data, and connected to college attendance at age 20. These

students reported earnings on taxes at ages 25 and 28. As the age checks furthered, the

sample size decreased with 376,000 samples at age 28 providing a more than large enough

sample for estimates of teachers influence on future learning. Essentially, based on this study,

high value-added teachers positively impact and change lives for the better beginning as

teens all the way through retirement, students at all levels have access to better lives when

taught by high value-added teachers (Chetty, Friedman, & Rockoff, 2012)

Teachers greatly influence students' future earning power, and test scores are an

accurate indicator for determining successful teachers (Chetty, Friedman, & Rockoff, 2012).

In order to isolate the teacher effect when analyzing student test data, other factors such as

socio-economic status must be accounted for by using covariates within the statistical

models, which isolate the teacher effect and make it more accurate but also much more

complex and difficult to calculate. VAM are statistical estimates of teacher performance for

comparing teachers to each other within a school, district, state, or other setting and as such

they are compared to an average model, meaning half of all participants will always be rated

below the average indicator (Lomax & Kuenzi, 2012).

A quantitative comparison of four different methods of computing value-added scores

using data from a cohort of 5th grade Arkansas math teachers determined that teacher

rankings across the four models as well as teacher ratings were statistically similar. Also,

each year examined, across all four models less than two to one percent of teachers were

identified as ineffective; hence all models identified the largest percentage of teachers as

effective (Blackford, 2016). The study compared value-added measures with mean test

scores and determined they were strongly correlated suggesting the use of mean test scores rather than the more complicated value-added measures (Keels, 2014). Shneyderman and Froman (2015) compared three methods of incorporating student growth measures into teacher evaluation ratings including a state value-added model, a district –level model using single level regression, and a common approach of student growth percentiles. They maintained findings at the basic level of the models reveal similarities. When the different techniques for aggregation are computed, the results are very similar. If SGM were easy to understand, implement, and readily improve teacher practice, then it makes sense that they should be adopted, but the process is more complex.

Limitations of VAM as part of teacher evaluation systems include: (a) the majority of standardized assessments cover a range of variability, (b) most state assessments vary in precision, (c) nearly all state assessments measure grade level standards without including methods of measuring growth, (d) some content or instructional positions are not measured by standardized assessments, and (e) the validity of inferences generated by VAM are difficult to guarantee even when meeting the most technical requirements (American Educational Research Association Council, 2015).

Concerns with SGM adoption

Not surprisingly, teachers recognize many of these limitations and bring many of their own concerns to the forefront when confronted with evaluation reforms. Negative teacher perceptions that have resulted from the use of student growth measures include (a) teachers attempt to influence the process either by roster manipulation or teaching only to anticipated test items, (b) teachers will leave education for other careers, (c) teachers will intentionally avoid certain schools or groups of students, (d) there is an increase in stress and

stress related illness, and (e) collaboration will be sacrificed for competition (Hewitt, 2015; Hunt 2016). Teachers that participated in schools that included student growth measures as part of the evaluation, compensation, or both demonstrated statistically significant negatively predicted autonomy and job satisfaction (Huss & Eastep, 2011; Wright, Shields, Black, Banerjee, & Waxman, 2018). Value Added Measures are often critiqued for: (a) limitations to guide teacher improvement, (b) unavailable scores for non-tested subjects and grades, (c) fear of uneven comparisons of teachers, (d) lack of trust due to misunderstanding related to1VAM calculations (Gandha & Baxter, 2016).

Lee (2011) asserts most teachers perceive the use of VAM as part of performance evaluation practices and a negative attack on teachers regardless of how they perform, especially when VAM ratings are made public to the parents and community. The exception is a small subset of teachers who received the highest ratings, thus feeling celebrated and recognized by the publication of ratings (Lee). Storie and Denner (2015) found, when analyzing Idaho's recent merit bonus pay system, a statistically significant negative relationship between the percentage of students eligible for free and reduced lunch and the amount of bonus pay awarded to teachers. At the same time, when examining differences between rural, suburban, and urban schools, rural schools were most likely to earn the merit pay. Challenges found when using student test scores to measure teacher performance include (a) ignoring estimation errors, (b) incorporating measurement error into t-statistics, (c) failing to revise value-added estimates when needed or justified, (d) manipulating of testing rosters by teachers, and (e) monitoring own students during exams (Ballou & Springer, 2015).

Even as teachers and those pushing back against evaluation reform efforts vocalize

their concerns, others, in support of the reforms, or in the many states that have adopted their use, have found useful opportunities for objective evaluation. The Measures of Effective Teaching Project funded by the Bill and Melinda Gate's Foundation concluded: (a) the observation instruments studied were consistent predictors of student achievement results, (b) multiple observations were required to accurately determine teacher performance, (c) combining observation scores with student survey feedback and student achievement results increased the predictive power and reliability of the instruments, (d) the combined measures were able to determine which teachers would produce larger student achievement gains compared to years of experience or graduate degrees, and (e) the students of teachers with higher combined scores reported more enjoyment of education, greater confidence and the ability to demonstrate deeper conceptual understanding of material presented (Kane & Staiger, 2012). Teachers perceived a strong relationship between the evaluation system and student learning when given ample training with the evaluation system, and verbal and written feedback from supervisors improves teaching, learning, and professional development (Junor-Carty, 2017).

Objective performance data consistently correlated with principals' beliefs of teacher performance, and the more accurate the data, the more principals trust it unless they hold strong subjective beliefs previously of poor teacher performance (Rockoff, Staiger, Kane & Taylor, 2012). Comparing the value-added scores of 24 middle school math teachers from a population of 222 New Mexico teachers along with survey and observation data obtained from participants resulted in correlations between the value-added scores with teacher mathematical knowledge, quality of instruction, and with the population of students being taught (Hill, Kapitula, & Umland, 2010). Springer, Swain, and Rodriguez (2015) discovered

the Tennessee teacher retention incentive program that rewarded successful teachers in the highest priority schools with a $5000 bonus based on student performance scores was substantially and significantly successful for teachers in grades and subjects that include standardized testing, but the program was unsuccessful in generating retention for those in non-tested grades and subjects. Sorting students potentially eliminates the validity of teacher effects identified by some value-added models that are incredibly biased, but by using complex value-added models over multiple years, correlations, while not at zero, were far from one, thus demonstrating a negated sorting bias (Koedel & Betts, 2009). Using interviews with 32 teachers and survey data over two years from 12,000 Chicago Public Schools teachers, following the adoption of their new teacher evaluation system that incorporated student growth measures, findings suggested teachers overall were positive about the new system, in particular the observation process, but they had concerns about the incorporation of the student growth measures (Jiang, Sporte, & Luppescu, 2015).

Not all research related to SGM has been supportive of its inclusion in teacher evaluation.1A quantitative analysis of 5th grade teachers in a school district just outside of Chicago to determine a relationship between a measures of academic progress exam scores and ratings given to teachers by administrators conducting observations using the Danielson framework found no correlation between the two (Alexander, 2016). Following the implementation of a new evaluation system that included a base component of the Danielson Framework and multiple student achievement measures, it was recognized that scores in math and reading increased on some assessments, but not all. At the same time, no direct correlation between the new evaluation system and change in scores was determined (Mathus, 2017). Steinberg and Kraft (2016) used data from the Measures of Effective

Teaching Study to perform simulation analyses and determined that principals consistently rated teachers higher than external raters, and the thresholds determined for different teacher ratings greatly determine the distribution of teacher performances. The second result is important because the thresholds were presented as being able to be determined differently by different districts; thus, making it difficult for states to compare teacher performance across districts.

The intention and logic of policies implemented to improve teacher performance through evaluation and value-added models does not correspond to the findings of empirical research. Other methods would be more effective for improving teacher practice (Hallinger, Heck, & Murphy, 2014). Even under the most controlled circumstances, estimators of teacher value added to student achievement have an above average likelihood for misclassifying the 'worst' and 'best' teachers (Guarino, Reckase & Wooldridge, 2015). Administrators conducting the observation portion of the teacher evaluation process often gave inflated scores to teachers to offset expected low VAM (Education Analytics, 2014). An analysis of survey and interview data of six school districts implementing new evaluation systems found that principals rely more on their classroom observations than student test data for human capital decisions (Goldring, et al., 2015). Harris, Ingle, and Rutledge (2014) discovered through a quantitative analysis of 32 schools that value-added measures and informal principal evaluations were positively correlated, weakly, and through a qualitative analysis that some principals give high-value added teachers lower ratings due to perceptions of low effort and less contribution to the overall school. Value-added models are insufficient tools for teacher evaluation, but they could prove useful as additional data wealth is generated by years of experience (Koedel & Betts, 2009). While there are many potential upsides to the

incorporation of SGM, there are equally as many or more concerns about such adoptions. For some school districts, theory was enough to begin the use of such practices before fully realizing the potential positives and negatives.

Adoption of Student Growth Measures (SGM)

A reasonably well-known example of incorporating SGM into teacher evaluation was Washington, D.C, Public School system, under the leadership of Michelle Rhee. D.C.'s IMPACT evaluation model weighs different measures to calculate a teacher's final score with classroom observations scores accounting for 35%, student test scores for 50%, commitment to the school community at 10%, and overall school performance at 5%. For teachers without standardized testing for their grades or subjects, the observation component is moved to 75% and 10% for teacher assessed student data (Headden, 2011). Using a quasi-experimental model to examine the first years of the IMPACT evaluation model adopted by Washington, D.C., a statistical significant finding (0.14 SD in reading and 0.21 SD in math) was identified regarding the removal of poor performing teachers to be replaced by more successful teachers (Adnot, Dee, Katz, & Wyckoff, 2017). Most teachers are labeled effective or minimally effective, the middle categories, so the teachers receive an acceptable rating or are given a year to improve. Only a very small percentage of teachers receive the lowest rating of ineffective which makes them eligible for termination, but administrators may choose to give them a year to improve also (Headden, 2011).

Los Angeles also had a very public display of their adoption of SGM, when local news agencies reported and published all LA teachers' performance ratings as established by the VAM system. An analysis of the published performance ratings of Los Angeles teachers based on the school district's value-added model, demonstrated a quantifiable difference in

teacher quality, but found only a weak correlation between the ratings and teachers' years of experience, educational degrees earned, teaching credentials, race and gender. A re-analysis of the same data concurred with the differences in teacher quality and correlations with the exception of finding a strong correlation of teacher performance and years of experience and education (Briggs & Domingue, 2011). Bacher-Hicks, Kane, and Staiger (2014) used a quasi-experimental model to examine the value-added estimates present in the Los Angeles schools' evaluation model. Of their three findings, they determined (a) the value-added model is an unbiased predictor of teacher performance, (b) the predictors are accurate regardless of teachers changing schools within the district, and (c) there are systematic differences of effectiveness regarding student race, ethnicity, and prior achievement.

In addition to these very public displays of adoption outcomes regarding SGM, Both the state of Tennessee and Dallas Public Schools in Texas have used accountability systems for more than 20 years to apparent success, but a lack of transparency on how they are computing the ratings and scores and confusion among analysts to determine the models being used exist (Lomax & Kuenzi, 2012). The Southern Regional Education Board, which supports school districts in sixteen states, reported that all sixteen states had required some component of student growth measures to be included as part of teacher summative ratings ranging from being a small contributing factor to a significant factor of half or more of the overall ratings (Gandha & Baxter, 2016). As of 2015, 40 states and Washington, D.C. incorporate some form of student growth measures as part of teacher evaluation (Hewitt, 2015). Doherty and Jacobs (2015) report in their *State of the States: Evaluating Teaching, Leading, and Learning* published by the National Council on Teacher Quality that 43 states require objective measures of student growth in teacher evaluation, in 17 of those states,

SGM counts for half or more of the summative evaluation score, and in 23 of the states, student performance informs tenure and termination decisions. Their report also acknowledges that one of the pushes for reform was that past evaluation systems recognized almost all teachers as being at the highest ratings, but those states that have already fully adopted the new evaluation practices still report almost all teachers at the highest levels of performance even when student achievement remains at the previous levels of performance.

All teacher evaluation systems have some flaws, and value-added models have many. The most effective tool would be to use a hybrid evaluation system with multiple components (Hanushek & Rivkin, 2010). Interview and survey data of more than 200 school leaders led to the determination that teacher evaluation data is rarely used for dismissal without first providing development and support for improvement, so, while the data may lead to termination, the primary focus becomes development and improvement (Drake, et al., 2015). When teachers believe there is a possibility of job loss, their students perform slightly better the next year (Rockoff, Staiger, Kane & Taylor, 2012). Jacob (2011) found principals consider multiple factors when contemplating termination and are more likely to do so when faced with certain teacher characteristics: (a) high teacher absence rate, 12% are more likely to be terminated, (b) teacher that received satisfactory performance ratings or lower (as opposed to excellent or superior) were 22% more likely to be terminated, (c) inversely, the more prestigious the teacher's college or university, the less likely he or she would be terminated, (d) male teachers were 3.8% more likely to be terminated, (e) teachers over the age of 50 were 10% more likely, and (f) black teachers were 2.1% more likely to be terminated.[1]This study consisted of more than 24,000 teachers across almost 600 schools, all part of the Chicago Public Schools system from 2004-2007 during a policy change that

allowed for easier dismissal practices along with the adoption of a new evaluation process (Jacob, 2011). When student growth measures are incorporated as a component of summative evaluation, those teachers with established trust with the administrator conducting the evaluation, the concerns were addressed and put to rest (Wilson, 2016).1However, teachers without strong trust in the administrator remained filled with uncertainty and behaved in a model of self-preservation in fear of the incorporation of student performance scores (Wilson). Typically, the classroom observation scores place teachers at a higher level than the value-added scores generated by student assessments (Headden, 2011). While recognizing teacher effectiveness varies greatly, only years of experience have demonstrated a consistent relationship to student performance (Rockoff & Speroni, 2010). The Ohio Teacher Evaluation System (OTES) required 50% of a teacher's performance rating be based on same-year student achievement data with the remainder of the rating established by supervisor observations along with pre- and post- observation conferences (Elam 2017). Additionally, supervisors, typically principals, must complete a rigorous training process in order to perform the observations (Elam). A mixed methods study (Elam, 2017) and a qualitative study (Harper, 2016) both included interviews with principals and discovered similar findings. Principals remained wary and doubtful of the ability of OTES to improve teacher practice because the focus of conversations tended to be the teachers' ratings and how they were calculated rather than improving practices in class (Elam, 2017; Harper, 2016). Even as systems attempt to guard against the subjective nature of supervision and evaluation, these studies illustrate at the core of the interaction, people are interacting with people, and they will always view their world through their own perceptions.

Principals and teachers alike remain wary of changes forced upon their schools and

practice, but with state adoptions of evaluation reforms and implementation of new demands, principals and teachers must work with their districts and states to advocate and implement effective practices that best support student achievement. Now that ESSA has reduced the requirements of SGM measures, it remains to be seen if states will begin to shift back to previous practices or remain resolute with their recently selected courses of evaluation implementation. Ultimately, SGM can be a component of successful evaluation if all stakeholders come together to create a system with multiple measures that accounts for all the demands placed on teachers' time, including their input in how and when such a system would be implemented.

Summary

While current practices and calls for reform may appear alien from the origination of supervision and evaluation, the core purpose remains the same. Student learning will be enhanced through improvement of teachers' abilities. Early evaluation efforts involved community members responsible for visiting the teachers during their work to provide direction and guidance. This responsibility flowed to superintendents and was later delegated to supervisors, as a result an educational hierarchy was created. With continued development of responsibilities and roles, building principals took on evaluation tasks with the potential for incorporating student achievement as a necessary component for determining effectiveness. Looking back, more attention was given to the leading measures of teacher observation rather than the emphasized concept in current reforms of incorporating the lag measure of student standardized test scores. Consistently, when supervision and evaluation were described as successful, there existed a mutual relationship and purpose between supervisor and teacher regarding the development of practices and strategies to benefit

students. Equally consistent, ineffective supervision and evaluation occurred when other factors were introduced and obstructed the original purpose. In order for success to occur, time must be devoted to effective evaluation practices, training must occur for all parties involved, and professional relationships based on meaningful teaching must be fostered to increase access to changing teacher behaviors for the betterment of students.

The remaining chapters detail the methodology of the study and analysis of data, reporting on the findings, and implications for future practice and research. Chapter three includes the design of the narrative case study in detail including sampling, data collection, and methods of analysis. Chapter four provides description of raw, collected data along with refined results gleaned from the data. Finally, chapter five includes responses to the research questions and discusses the implications for current and future practice in the educational setting while also suggesting future research to be conducted.

CHAPTER 3

METHODOLOGY

Teacher evaluations, which influence high stakes personnel decisions made by building administrators and district leaders, poorly depict teacher performance and are presented with many obstacles (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Marzano, 2012; Murphy, Hallinger, & Heck, 2013; Papay, 2012; Shinkfield & Stufflebeam, 2012; Toch & Rothman, 2008). The majority of teachers receive the highest ratings while very few teachers receive an unsatisfactory rating (Kimball & Milanowski, 2009; The New Teacher Project, 2007; Weisberg, Sexton, Mulhern, & Keeling 2009). High stakes personnel decisions based on ineffective instruments and practices could negatively impact teacher retention and increase teacher turnover (Drake et al., 2015; Kimball & Milanowski, 2009; Odden & Kelley, 2008). Teacher reassignment, attrition, and turnover negatively influence student performance (Blazar, 2015; Ronfeldt, Loeb, & Wyckoff, 2013; Steinburg & Sartain, 2015). The poor state of teacher evaluation and the predicament of its current context include separation of evaluation results and professional development and inadequate time and staff to perform effective evaluations (Marzano, 2012; Marzano & Toth, 2013; Papay, 2012).

The purpose of this narratological multi-case study was to examine teacher perceptions of student growth measures used for performance evaluation for core content teachers at an urban middle school in a mid-western city. The unit of analysis was narrowed to the teachers' perceptions of the measures and their inclusion in evaluation. As a result, this study responded to the identified research question and sub questions:

- How do teachers perceive the use of student growth measures in performance evaluation for core content teachers at an urban mid-western middle school?

  - What do teachers understand about the context of job performance evaluation?

  - How do teachers define student growth when measuring their own performance?

  - How do teachers describe performance evaluation?

The study adds to the literature in the field by representing a point in time during nationwide legislative changes directly related to this topic. This study also contributes to the understanding of the topic from the point of view of those impacted by such policies. Another manner in which this study contributes to the research is by offering a deep examination of the perspective of teachers across ranges of experience. The literature on this topic is predominantly quantitative representing student achievement or the lack of it related to the consideration of the variable of student growth measures in the evaluation process (Glazerman & Seifullah, 2012; Hill, Kapitula, & Umland, 2011; Kimball, White, & Milanowski, 2003; Rockoff & Speroni, 2010; Taylor & Tyler, 2012). The literature is enriched by capturing teacher perceptions of the inclusion of student growth measures in high stakes performance evaluation and the thick description of how student growth measures impact teachers' realities related to providing service to students, parents, schools and districts (Campbell, 2013; Ladd, 2011; Nelson, 2012; Stronge, Ward & Grant, 2011). Without a deeper understanding of the consequences of continuing policy adoptions related to student growth measures, it is difficult to determine if such policy adoptions truly create the opportunities intended or if they actually undermine the original intentions in ways not readily observed at this time.

This chapter includes the rationale for qualitative inquiry including a description of the theoretical traditions that guide the methodology, role of the researcher, the design of the study which entails a description of the setting, including participants and sampling techniques, and methods for data collection and analysis. I conclude with an examination of the limitations, validity, reliability and ethical considerations.

Rationale for Qualitative Inquiry

This study examines the lived experience of teachers as they make-meaning of reform efforts altering their performance evaluation process, which fits a naturalistic paradigm. Qualitative research allows for an in-depth examination and understanding of the human experience within natural settings (Ayres, Kavanaugh, & Knafl, 2003; Grbich, 2013; Hayes & Wood, 2011; Holley & Colyar, 2012; Stake, 2010). "Qualitative research is a situated activity that locates the observer in the world" (Denzin & Lincoln, 2011). Qualitative research designs require assessing, reassessing and potentially redesigning the interactions of the design, making adjustments if needed in contrast to quantitative studies' more prescriptive design requiring set, specific steps to be followed through with fidelity before analyzing for adjustments in future studies (Maxwell, 2013). "By quantitative we mean that its' thinking relies heavily on linear attributes, measurements, and statistical analysis" (Stake, 2010, p. 11). In other words, quantitative research focuses on variables and their statistical relationship to each other (Maxwell, 2013). Stake (1995) described the differences between quantitative and qualitative approaches simply: quantitative seeks explanation and control while qualitative seeks to understand the complex interrelationships that exist.

Pursuing a qualitative study allowed me to study the teachers' world in terms of people, situations, events and the processes tying them together (Maxwell, 2013).

"Qualitative inquiry is personal. The researcher is the instrument of inquiry" (Patton, 2015, p. 3). A qualitative researcher's subjectivity provides value to the study (Grbich, 2013). Maxwell (2013) asserted separating the researcher from the research is harmful because it creates an illusion of objectivity and leads the researcher to ignore influences of or understanding, decisions, and interpretations related to phenomena of interests, and it also cuts the researcher off from major source of insights, questions and practical guidance in conducting their research.

Further, Stake (2010) recognized qualitative studies as interpretive, experience based, and situational, requiring diverse and complex thinking. Thick description, experiential understanding, and multiple realities derive from qualitative research (Denzin & Lincoln, 2011; Stake, 1995). Researchers look for patterns both expected and unanticipated (Stake, 1995). This stance requires qualitative researchers to be flexible about what might be unearthed as the study progresses (Maxwell, 2013). Patton (2015) notes the researcher is likely to discover the why of patterns in the data:

> Qualitative research often inquiries into the stories of individuals to capture and understand their perspectives, as just discussed. But often the answer to why people do what they do is found not just within the individual but, rather, within the systems of which they are a part: social, family, organizational, community, religious, political, and economic systems. (Patton, 2015, p.8)

In short, qualitative research includes the natural setting, the researcher as the instrument, multiple methodologies and forms of data, complex reasoning with inductive and deductive logic, participant meaning, emergent design, reflexivity, and holistic account (Creswell & Poth, 2017).

I used what Patton (2015) described as *openness of inquiry* by asking open-ended questions, maintaining an open-mind, looking for patterns in hidden details. Working in the

context of the study required understanding of the culture, politics, history, resources and institutions of the school and district (Patton, 2015). As the goal of the study was to understand the complex setting of student growth measures in performance evaluations and teachers' responses to this phenomena, qualitative study design was selected. Qualitative research design has a multitude of frameworks and traditions from which to construct a study (Creswell & Poth, 2017; Patton, 2015). The theoretical traditions used to explore the teachers' perspectives include multiple case studies as the major technique through the lens of narrative inquiry.

Case Study

Qualitative case study has a popular history in the social sciences such as psychology, medicine, law, and political science. Initially, fields including anthropology, history, psychology, and sociology conducted case studies to examine lived experiences (Simons, 2009; Stewart, 2014). Early forms of case study focused on quantitative collections of data such as the work done by Charles Darwin, but with the development of grounded-theory by Glaser and Strauss merged more qualitative methods of analysis with the original quantitative methods of data collection in the forms of case studies (Harrison, Birks, Franklin, & Mills; 2017). Current qualitative case study research can be traced back to the 1920s and 1930s through the fields of anthropology and sociology (Creswell & Poth, 2017). Anthropologists at the University of Chicago School of Sociology conducted lengthy case studies using field-based observations to study different social and cultural experiences on their campus (Creswell, Hanson, Plano Clark, & Morales, 2007; Stewart, 2014). During the late twentieth century and into the early twenty first century, three educational research methodologists developed and defined contemporary, qualitative, case study research methods that vary in

purpose, strategy and technique: Robert Yin, Robert Stake, and Sharan Merriam (Yazan, 2015).

In order for a case to be present for study, it must be a "bounded system," and as such people and programs qualify as good cases. Events and processes fit this description less well and would not fit within a case study without imposing clearer boundaries (Stake, 1995). Yin (2014) provides a more detailed, analytic approach to defining case study. He described a case study as an investigation in depth and in real-world context that includes many more variables than data points and relies on multiple sources of evidence that converges in a triangulation fashion. Yin describes his interpretation of case study as being more fully inclusive and applicable to both qualitative and quantitative studies while Stake's maintained a qualitative skew. Both, though, focus on an in-depth understanding in a real-world context recognizing the need to see the focal point from multiple perspectives. For this study, the case was bound and defined by narrowing the focus to an urban, middle school teacher of a core content subject in the Midwest who has experience with his or her district's teacher evaluation process during a time of reform and change as outlined by the *Every Student Succeeds Act* (ESSA, 2015).

The use of multiple case study was intended to create thick, rich description in a context rich environment from multiple perspectives and ranges. The focus narrowed in on the specific, observable beliefs, actions, and events as sources of inquiry, and the actual contexts within which these are situated (Creswell & Poth, 2017). While such a small sample size does not lend itself to generalizability to other populations, it created generalizations or petite generalizations about larger theories (Stake, 1995; Yin, 2014). Additionally, the use of multiple cases allowed for comparisons across cases so as to illuminate potential larger

patterns or themes during analysis (Miles, Huberman, & Saldana, 2014; Patton, 2015). A target of six participants were invited to participate in the multiple case study. Each individual participant was considered a separate case, and the incorporation of six cases made this a multiple case study.

Case studies are classified in different manners beyond the number of participants involved. Stake (1995) identified three types of case studies: collective, intrinsic, and instrumental. Yin (2014) also identified three types of case studies: descriptive, explanatory, and exploratory. While each serves different purposes, each of them greatly overlap with the others (Stake, 1995; Yin, 2014). Both collective and multiple case studies refer to the number of participants, and for this study, I used the multiple method as previously identified. Exploratory often acts as a precursor to a larger study, and explanatory serves to explain the phenomenon in excruciating detail as bounded by the case (Yin, 2014). Descriptive case study also desires to describe in explicit detail the phenomenon occurring (Stake, 1995). While these could be utilized, I do not believe they fully capture the intent of the study to examine how the teachers perceived the changes occurring. Instrumental case study begins by selecting the concern first and then examining it by focusing on the case, and recognizing my process of identifying the problem first, followed by determining how to study it suggested this would be a strong connection for the study (Maxwell, 2013). At the same time, intrinsic case study could have met the needs of the study if my focus was on the case, and I studied the case through the lens of the changing evaluation process (Stake, 1995). When stepping back to best evaluate the tool that was most effective for this study, instrumental case design best fit the needs to examine the phenomenon and cases involved because of the focus on the phenomenon in the lives of the cases rather than examining the cases through

the lens of the phenomenon. Thus, the case study is best described as multiple with the use of an instrumental case design. I also incorporated narrative inquiry, which influenced the design of the study as well.

Narrative Inquiry

John Dewey expressed the notion of experience as a form of inquiry. He also held that individuals encountered both personal and social interaction, and that experiences occur along a continuum always building on past experiences and being the support for future experiences (Dewey, 1938). Clandinin and Connelly (1990) identified narrative inquiry in education research beginning with historical accounts of teachers' lived experiences and stories of daily life as well as collecting the experiences of non-traditional students such as second language learners and adult students. "For us Dewey transforms a commonplace term, experience, in our educators' language into an inquiry term, and gives us a term that permits better understandings of educational life" (Clandinin & Connelly, 2000, p. 2). Equally important to experience in narrative inquiry is the relational aspects: the relationship between the participant and the narrative shared, the relationship between the researcher and the narrative understood, and the relationship between the participant and researcher that affects both previous components. Narrative inquiry embodies the relations of those invested and the manner in which they invest (Clandinin, 2013).

The narrative inquiry tradition was exemplified in the design of this study by the collection of stories of those experiencing the changes central to the phenomenon (Patton, 2015). By collecting the stories detailing the teachers' perceptions of their work in schools and how it has been, could be or should be measured from their point of view allowed for thick description and rich understanding of their experiences. "One theory in educational

research holds that humans are storytelling organisms who, individually and socially, lead storied lives. Thus, the study of narrative is the study of the ways humans experience the world" (Clandinin & Connelly, 1990).

Clandinin and Connelly (1999) coined the phrase "personal practical knowledge" to describe teachers as knowledgeable and knowing persons. "Personal practical knowledge is in the teacher's past experience, in the teacher's present mind and body, and in the future plans and actions" (Connelly & Clandinin, 1988, p. 25). Teachers' knowledge from this standpoint is formed and expressed through narrative context and stories (Connelly & Clandinin, 1999). Teachers express their lived experiences through secret stories and cover stories with the first being their lived experiences sometimes shared with trusted fellow teachers and sometimes kept for themselves, and the latter being those stories and narratives shared with others outside the classroom experience that may not know or understand the teachers' experiences (Connelly & Clandinin, 1999). My roles and experiences as an educator created opportunities for me to access the secret stories the participants held regarding their perceptions of teacher evaluations. As this support for a qualitative design was described, qualitative studies encouraged the researcher to discover the lived, human experience while recognizing the researchers' subjective interpretation and meaning-making through the experience. The instrumental multiple case study approach created opportunities to examine the phenomenon across multiple cases to discover themes and patterns in the perceptions of teachers. Narrative inquiry created the avenue to capture teacher experiences and personal practical knowledge, enabling me to learn the secret stories held by the participant teachers regarding their perceptions of student growth measures in performance evaluations.

Role of the Researcher

As mentioned, I am an educator, administrator, and tasked with the duty of teacher evaluation, so it was important to use reflexivity regarding the subjects and reveal my biases to create sincerity and transparency (Tracy, 2010). "The term reflexivity is meant to direct us to a particular kind of reflection grounded in the in-depth, experiential, and interpersonal nature of qualitative inquiry" (Patton, 2015). At its core, the researcher used reflexivity as a tool of deep reflection on how his or her actions influence those observed and the setting the observation occurs in while balancing those influences with what was studied (Carlson, 2010). As I work with the teachers, collecting their voices and stories, my interaction with them broke up their daily routines, interrupted what they would normally be doing if I were not present. The questions I asked may have caused them to see themselves or me differently, or to see their work differently. My existence in their world created ripples in their lives that were not there before my arrival. As a researcher, I made every attempt to keep the ripples as small as possible, realizing their existence cannot be eliminated, and still conduct the study. To be reflexive meant to build a relationship with those being studied while being aware how that relationship might have affected the people and the phenomenon (Creswell & Poth, 2017). While the research focused on the unit of analysis, every interaction was some form of intrusion on the participants, and they were treated with respect (Maxwell, 2013).

As will be detailed more in a discussion of field texts later in this chapter, I maintained a journal to capture my thoughts, feelings, beliefs, and assumptions that occurred during my reflections. Journaling throughout the process helped me to be more cognizant of my biases. It was also important to identify a critical friend that I shared my thoughts and reflections with throughout the data collection and analysis process. I engaged in constant

dialogue with this individual and had this person interview me about my own experiences with evaluation. These experiences were included in the final reporting of data to increase transparency with readers and establish credibility of my understanding and role in the research.

Design of Study

The study took place at a middle school, referred to as Hawkins Middle School (pseudonym) in an urban mid-western school district, referred to as Hilltop School District (pseudonym). The district adopted an evaluation program meeting state guidelines and requirements, which included the ability to incorporate student growth measures into summative teacher ratings to be reported to the state in an aggregate format. The principal has been in charge of the building for more than three years and has established expectations and routines regarding evaluation and performance. The school was representative of the district across multiple demographic categories as exhibited in table 1. Both the school and the district serve diverse communities. While the school serves a larger ratio of Latinx students, the representation of European American students was similar. Also, the ratio of English language learners (ELL) and special education students to the total populations was equivalent. In addition to race and educational needs, the school reflected the district in the area of economically disadvantaged students by both exceeding the threshold of Title 1 qualifications by a large degree.

The stories and interview responses shared by teachers about their evaluation experiences were used to capture their perceptions of including student growth measures in performance evaluation. The teachers' perceptions of these measures are the units of

analyses. Patton (2015) states the unit of analysis directs the data collection on what is happening to the individuals in a setting and how they are affected.

Table 1

Demographics of Hilltop School District and Hawkins Middle School

Site	Enrollment	Latinx	African-American	White	Other	Economically Disadvantaged	ELL	Special Education
Hilltop SD	19,500	50%	29%	12%	10%	85%	41%	14%
HawkinsMS	604	66%	18%	12%	4%	93%	42%	14%

State Department of Education, 2017

Participants and Sampling

The participants were middle school teachers in core content subjects such as math, science, social studies or English/language arts. The goal for final reporting of findings was six participants, each representing an individual case and different years of experience with evaluation in the Hilltop School District. These individuals were to be selected through a purposeful sampling process called maximum variation sampling. Maxwell (2013) describes purposeful sampling as deliberate selection of participants to provide information directly relevant to the research questions that could not be gained elsewhere. Creswell and Poth (2017) recognize maximum variation sampling as ideal for qualitative studies because it can increase the possibility to discover different perspectives about the phenomenon. The goal of incorporating maximum variation sampling was to capture the heterogeneity of the group by identifying variations within the group of possible participants and systematically selecting those that provided the greatest range within the variations (Maxwell, 2013).

When selecting cases for this study, maximum variation sampling was going to be used to increase the number of content subjects represented in the group, diverse perspectives (race, ethnicity, and gender), and a variety of years taught in the building from the available

pool of participants. The subject being taught by the participant was primary as the evaluator's previous experiences as a teacher may influence how or how well the evaluations were performed, thus, influencing the teacher's relationship to that event. Secondly, race, ethnicity, and gender were included to increase the number of personal and diverse perspectives possible. The years taught in the building was selected with the consideration that more experienced teachers would possibly have encountered other forms of evaluation or supervisors that conducted evaluation while newer teachers may have only experienced evaluation with the current supervisor. Ultimately, purposeful sampling was not used to identify participants because only six participants were willing to contribute to the study. A variety of factors could have caused other staff members to decline to volunteer such as the start of spring sports and teachers could be coaches or their own children could be playing, personal responsibilities limiting time available for participation, state assessments occurred during this time and many teachers identify that as a stressful time, or new initiatives and expectations handed down from new district leaders could have cause them to pass on participating.

All core content teachers at Hawkins Middle School were asked to participate in a survey collecting demographic information, interest in participating to a further extent, and requesting a brief narrative of their best or worst experience with evaluation. I made this request in person following a brief description of myself, the purpose of the study, and the commitments it entailed. Surveys made it quite easy to cast a large net for initial participants, but the method was considered that it might have drawn a very particular type of participant versus face-to-face methods might have recruited as well as how it may affect their responses (Palys & Atchison, 2012). The building administrator provided the teachers' names and

contact information. After speaking to the teachers as a whole group to introduce myself, describe my study, and emphasize confidentiality for participants throughout the study, I provided the survey (appendix A) to the core subject teachers using Qualtrics, a secure server and data capturing software with encrypted transmissions of data. The only question requesting their name or identifying information was the optional last question for those interested to self-identify a willingness to participate in the full study. Those individuals that expressed an interest in participating in the study were provided with a consent form for participation as well as a pseudonym to be used for further data collection. Only I had a list of pseudonyms associated with participants, and the list was kept on a password protected computer.

The participants were invited to engage in more in-depth interviews (appendix B) and provide narratives (appendix C) as case study participants. While qualitative research does not require nor expect a representative sample due in large part to the sample size, selecting the greatest diversity available among those variables would have allowed for the greatest amount of exploration among the different cases during analysis (Maxwell, 2013). As such the participants did provide a range in years of experience (4 to 33 years), years in district (1 to 33 years), gender (5 male and 1 female), and all four core content subjects (and some had experience teaching elective courses also). All participants identified as European-Americans. Of all the core content teachers available to participate, approximately 40% were male, 60% were female, 88% were European-American and 12% were of other races. This was representative of secondary schools in this district, and the school board was taking actions to more actively recruit a more diverse collection of teachers regarding race and ethnicity to be more aligned to reflect the population of the students being served.

Data Sources and Collection

A variety of data sources including surveys, in-depth interviews, narrative prompts, and documents were used to gain crystallization. Each provided different insights and information to different degrees (Babchuk & Badiee, 2011). Ellingson (2009) describes crystallization as a means of achieving depth of knowing through the process of capturing large quantities of data in a wide variety of forms to provide a "wide-angle view" of the object of the study. Each data source provided a different lens through which to understand what was being studied and the context surrounding the phenomena. By examining a phenomenon through multiple lenses and combinations of lenses, it was possible for me to have a more complete understanding of the unit of analysis within its context. The pool of data included the survey responses from the teachers followed by transcripts from the individual interviews, transcripts of narratives of evaluation experiences, and the collected documents. I collected documents including: relevant excerpts from the Elementary and Secondary Education Act (ESEA), *NCLB* and *Blueprint for Reform* adoptions and updates. Additionally, the state department of education (SDE) (pseudonym) provided such documents as: state vision for education, list of student performance indicators for teachers, descriptions of state approved evaluation systems, inter-rater agreements and expectations, evaluation timelines and deadlines, evaluation system requirements, state performance rating matrix, state summative scoring matrix, educator evaluation fact sheet, definitions of student performance and identifying indicators, approved application for state ESEA waiver, and state educator evaluation handbook. The human resources department of the school district included documents such as an outline of evaluation indicators, evaluation timelines and deadlines for teachers of various years of experience, calendar of scheduled trainings for

114

administrators, presentation handouts from administrator trainings related to teacher evaluation. The school principal provided the walkthrough tool used in his building as well as an outline effective teaching indicators looked for in classrooms during walkthroughs and observations.

Each case study participant was asked to share narratives about past or recent experiences with evaluation as well as predictive stories about the educational setting of evaluation with the implementation of student growth measures during the interviews. The narratological tradition allowed me to understand the human experience by interpreting narrative forms of data (Clandinin & Connelly, 2000; Hayes & Wood, 2011). As the researcher, I also collected field notes connected to the teachers' experiences, which helped me make meaning and connections across multiple forms of data. These writings took the form of personal memos and journal entries.

Surveys. Qualitative research surveys have three characteristics: (a) descriptive aspects or characteristics of the group, (b) the information is collected by asking questions and (c) the information is collected from a sample rather than all members of a group (Fraenkel, Wallen, & Hyun, 2012). For this study, all characteristics were met. The survey questions collected descriptive information about the group, the questions asked captured data to be analyzed, and I only surveyed the core subject teachers of one school rather than all teachers. In addition to providing descriptive information, the resulting data was used for participants to self-identify interest in participating and it would have allowed me to employ maximum variance sampling if more individuals had volunteered.

While surveys typically fall under the guise of quantitative research or mixed-methods, the open-endedness of some questions created opportunities for qualitative data

analysis (Fielding, Fielding, & Hughes, 2013; Fraenkel, Wallen, & Hyun, 2012). A qualitative survey does not aim to calculate frequencies or distributions of variables as a quantitative survey would, but to identify and recognize variance present among respondents (Jansen, 2010). The survey (appendix A) created an opportunity for initial qualitative responses. This was the first opportunity for participants to share narratives related to evaluation experiences through responses to open-ended questions, but none chose to do so. In summary, the surveys served multiple purposes including self-identification of participants, would have allowed for maximum variance sampling of willing participants, and provided an opportunity, through open-ended questions, for participants to share experiences with the phenomenon.

Interviews. Interviews serve multiple purposes in qualitative research. Qualitative interviews allow for obtaining unique information held by the interviewee and allowing the researcher to find information about something not observable by the researcher (Gill, Stewart, Treasure, & Chadwick, 2008; Marshall & Rossman, 2014; Meriam & Tisdell, 2015; Patton, 2015; Stake, 2010). "Qualitative interviewing begins with the assumption that the perspective of others is meaningful, knowable, and able to be made explicit. We interview to find out what is in and on someone else's mind, to gather their stories" (Patton, 2015, p. 426). Interviews as part of qualitative research should resemble guided conversations more than formal structured interviews (Turner, 2010; Yin, 2014). Clarifying and follow-up probes may be interjected to confirm or add details to participant responses (Harrell &Bradley, 2009; Patton, 2015). Interviews as part of a qualitative study are most appropriate when detailed insights are needed and are useful when the topic may be too sensitive to discuss in group settings (Gill, Stewart, Treasure, & Chadwick; 2008). Through interviews, participants

provided detailed information from their unique viewpoints and experiences (Turner, 2010). At the same time weaknesses in interviewing must be recognized such as individuals being interviewed may be uncomfortable with sharing, may misunderstand the questions, or the questions themselves did not lead to useful or descriptive responses (Marshall & Rossman, 2014).

The interviews allowed for the gathering of teacher understanding and perception of the use of student growth measures in teacher performance. As a part of this work, a general interview guide approach was used. The general interview guide approach allowed for greater flexibility than a standardized open-ended interview by allowing the researcher to be certain all areas had been covered, which was not insured in the informal conversation interview (Patton, 2015; Turner, 2010). The interview guide (appendix B) provided the structure to which topics will be introduced and discussed with the participants while allowing the participants to focus and reflect on those areas of greatest meaning to them within those topics (Harrell & Bradley, 2009).

The more open-ended the questions were designed then the more freedom and opportunity participants had to share deep, thick, descriptive responses that may not be gained through a narrower, structured question format (Harrell & Bradley, 2009; Merriam & Tisdell, 2015; Patton, 2015; Turner, 2010) Additionally, intention and effort was invested to avoid dichotomous questions and focus on singular questions. Dichotomous questions imply a yes or no answer limiting participant response while the dual or multiple-natured format of non-singular questions can be misleading or confusing (Harrel & Bradley, 2009; Patton, 2015). Both limit the thick description hoped for in qualitative interviewing.

The questions asked during interviews were designed to discover information to answer the research questions. The overarching research questions and sub-questions were used to design the questions used during the interviews. From the group of participants that self-identified and volunteered, interview dates and times were scheduled. The interviews occurred at locations of the volunteers' choosing. The length of initial interviews ranged from sixty to ninety minutes. Follow up interaction occurred with each participant once their initial interviews had been transcribed, so the participants could member check transcriptions for accuracy. The follow up contact created opportunities for clarifying and questions regarding their responses and narratives.

Narrative Prompts. In an effort to understand teachers' perceptions of changes in teacher evaluation, I needed to witness and understand them relationally within their setting. Clandinin (2013) describes narrative inquiry as a relational methodology. She goes on to describe the storyteller as defining their relationships with self, setting, world, others and so on in their stories. Narratives capture experiences and shed light on how people see themselves (Creswell & Poth, 2017). Connelly and Clandinin (1999) highlight the concept of sacred stories, secret stories, and cover stories provided by teachers to different audiences to reflect their work in safe ways and confirm their identities to themselves and others. By sharing myself and some of my experiences in my introduction prior to the survey and in my repeated interactions with participants, I acquired the teachers' trust and respect in order to move beyond the cover stories, and receive the secret stories saved for trusted peers that more genuinely reflect how the teachers see themselves and their work.

The interview participants were asked to share stories illustrating their experiences and expectations of evaluation. Participants were allowed to choose to share their narratives

through writing as field documents or orally as part of the interview. A deadline of five days following the interview was allowed for completion of any narratives not provided by the conclusion of the interview. Oral narratives were transcribed and provided to the participant along with the interview transcriptions to validate for accuracy. Participants were prompted for additional detail during the oral narratives or asked clarifying questions. No participants chose to write narratives, but one participant did volunteer copies of past evaluations for use as field documents. Multiple narratives depicting different experiences with teacher evaluation were encouraged for all participants. During the follow up contact, additional details, information, and experiences related to the initial narratives were requested from each participant. I asked if they would like to add to their stories or make changes, but none opted to do so.

Documents. Documents play a primary role in qualitative research; they are reflections of the people and setting from which they come, and detailed analysis can generate understanding of the people and their relationships with each other and the setting (Saldana, 2013). The category of documents involved a wide variety of materials and sources (Grbich, 2013) including the researcher's journals, participant's artifacts, public documents such as official memos, minutes and records (Creswell and Poth (2017). Stake (2010) adds to this list with newspaper articles, internal reports, and correspondence among many others. Given such a range of documents available for any field of study, researchers need to be selective in identifying those that are relevant to their study (Clandinin & Connelly, 2000). "In short, documents refer to any kind of information that exists in some type of written or printed form" (Fraenkel, Wallen, & Hyun, 2012, p. 537).

Bowen (2009) identified five uses for documents in qualitative research: (a) provide data on the context or setting of the participants, (b) suggest questions that should be asked, (c) specify supplementary research data, (d) offer a means of tracking changes and development, and (e) make available a way of verifying findings and corroborating evidence. Patton (2015) asserts "…documents prove valuable not only because of what can be learned directly from them but also as a stimulus for paths of inquiry that can only be pursued through direct observation and interviewing" (p. 377). The juxtaposition of information gleaned from participants and documents can lead to additional clarification as well as new inquiries and clues to further topics of interest depending whether the sources converge or diverge in alignment (Yin, 2014). As relationships with participants develop, the vision of the researcher can become more focused and narrow, eliminating the inclusion of outside influences. The inclusion of document reviews contributes to added perspectives (Clandinin & Connelly, 2000). When working with individuals similar to the researcher, such as an educator studying fellow educators, the potential for limited vision is particularly true. Connections and similarities between the participants and researchers could be limiting, but the lens provided by documents provide additional methods for crystallization of perspective (Bogdan & Biklen, 2007). While documents remain integral to almost every study, the researcher must recognize potential bias present in any document or if the document may have been edited before received (Yin, 2014).

Public documents were collected in the forms of federal and state policies and reports detailing requirements and expectations of evaluations and related reforms. Examples include the portions of *No Child Left Behind* (NCLB), *Race to the Top* (RttT), and *Every Student Succeeds Act* (ESSA) reports that describe the expectations of teacher accountability. These

documents were acquired from the United States Department of Education (USDE) website, State Department of Education (SDE) reports, fact sheets, and related documents that describe state expectations of school districts regarding policy, statute, and implementation of teacher evaluation including reporting methods regarding the state of teacher accountability. Hilltop School District's board policies, human resource directives and memos regarding teacher evaluation, and training materials provided to building administrators to prepare them for teacher evaluation were acquired from the school district website and from the district's human resources department. Other document sources included Hawkins Middle School's internal memos and communication regarding teacher evaluation, materials distributed to the teachers detailing the teacher evaluation process and expectations, and any other materials related to teacher evaluations tools, processes, and expectations provided by the building administrator or participating teachers.

Fieldwork. The data sources of the study, surveys, interviews, narratives, and published documents involved extensive fieldwork over the course of the study. "Fieldwork is the central activity of qualitative inquiry" (Patton, 2015, p. 55). Researchers must maintain descriptive, detailed field notes of all experiences and encounters during the study (Connelly & Clandinin, 2000; Patton, 2015; Stake, 2010). In addition to crafting field notes, artifacts can be collected representing the many facets of participants, study, and relationships involved. Many different artifacts such as photographs, artwork, timelines, records or recordings can represent field texts, most any memory box item kept by a participant (Patton, 2015). The scouring of such artifacts for relevant information and data can increase understanding and provide clearer observation of the setting.

Field notes take any form, and they are the most common form of data (Yin, 2014). "Field notes taken during an interview usually contain a fraction of the actual content" (Miles, Huberman, & Saldana, 2014). Often these serve to spark a memory, story or conversation between the researcher and participant resulting in the illumination of understanding. "But formal write-up usually will add back some of the missing content because the raw field notes, when reviewed, stimulate the field worker to remember things that happened at that time not in the notes" (Miles, Huberman, & Saldana, 2014). While published documents are objective products, field texts are "experiential, intersubjective" texts (Clandinin, 2013). Historically, researchers were warned to remain objective and distant from their research, but that distance narrows opportunities to understand the nature of what is being studied when the phenomenon includes emotion and interpretation (Patton, 2015).

Field notes were collected and maintained in a journal throughout the research process. These notes were generated in the form of reflections and journal entries capturing the process and impressions of the participants and researcher, narratives of lived experience during the research process, personal memos and notes to self. The artifacts shared by the participant were also be included. While no copies of evaluations were requested from participants, one participant chose to share copies of previous evaluations. Items such as these were categorized and treated as artifacts. These artifacts were incorporated into the individual cases' data to be analyzed as part of their narratives.

Data analysis

Analysis of qualitative data inherently consists of sense-making, taking the wealth of raw data and pairing it down into digestible, communicable findings (Patton, 2015). As previously noted, the data sets for this study included surveys, in-depth interviews, narrative

prompts, and documents. The surveys completed by the core classroom teachers primarily served to allow participants to self-identify interest in participating and would have provided the opportunity to identify maximum variance among those participants. The surveys completed by those not participating in the full study were not included in any data analysis. The surveys completed by selected study participants did include open-ended questions whose responses were analyzed using the first and second cycle coding process, described in subsequent sections.

Documents directly related to the subject of the study were examined incorporating the first and second cycle process as well. Each document was analyzed individually because while many documents were applicable to all cases, some were only applicable to individual cases. As an example, the data points, codes, and categories derived from analyzing the school district's evaluation policies and procedures were included for each case, but the data and subsequent codes and categories derived from an e-mail or audience specific memo was only included in the data sets of those cases directly influenced.

The narrative data collected from each participant including in-depth interviews, narrative prompts, and related field texts and artifacts was examined in the three dimensional inquiry space of sociality, temporality, and place. Relying heavily on reflexive practice to guard against my own biases, I re-storied the varied shared, lived experiences combined with understanding gleaned from analysis of the three dimensional inquiry to present a written narrative to each participant capturing, to the best of my ability, his or her reality. Through a member-checking process, each participant was given the opportunity to direct, clarify, add, or revise this collective narrative of his or her experiences creating a cycle of revision and

resubmission for member-checking until the participant were satisfied with the narrative's accuracy.

The data generated by each participant defined each individual case. In this regard, the codes and categories generated by documents and survey responses was analyzed as a collective set using the first and second cycle inductive coding process with in-depth interviews and data generated from narrative prompts analyzed using the three-dimensional narrative inquiry space. As previously noted, conceptualizing the variety of data sets through different lenses from which to view and learn about the case, captured the concept of crystallization. Further enhanced by the incorporation of narratives, Ellingson (2014) depicts crystallization as the incorporation of more artistic approaches with the more traditional creates additional access to reflexive acts. With the cases fully aggregated across multiple data sets into collections of codes and categories, themes and concepts were generated from the similarities, differences, frequencies, and gaps present in the data (Saldaña, 2013).

Surveys and Documents. Data sources including documents and the qualitative surveys were analyzed using enumerative and thematic coding (Grbich, 2013; Miles, Huberman, & Saldana, 2014). The act of a first coding cycle is defined by Saldaña (2013) as the complex process of qualitative data analysis incorporating multiple layers of inductive coding and involving multiple data sets. As noted, coding was incorporated as the primary method of analysis for those sources other than narratives as this took the raw and unprocessed data, reducing the material interpreted to a collection of symbolic, descriptive, and inferential information (Creswell & Poth, 2017; Miles, Huberman, & Saldaña, 2014). Coding entailed the researcher identifying categories for the components of data and determining intensity and quantity of various aspects present within the data. "A datum is

initially and, when needed, secondarily coded to discern and label its content and meaning according to the needs of the inquiry" (Saldaña, 2013). Miles, Huberman, & Saldaña (2014) label codes as a data condensation task as the codes act as prompts to elicit deeper meaning and to connect different data points together.

Rather than a priori codes, having a preselected set of codes to label data, inductive coding was incorporated to allow for the codes and categories to develop from what is encountered in the data (Maxwell, 2013). Additional coding processes incorporated were descriptive coding, using a word to describe the basic topics of passages, and in vivo coding, using the words and phrases from participants as the codes (Miles, Huberman, & Saldaña, 2014). While a variety of coding processes could be incorporated, I believed these assisted in the reflexive practices I employed to guard against my own language overwriting that of the participants. Each relevant data source went through the described first coding cycle of analysis. All recordings were transcribed; artifacts were captured as written field texts, allowing the coding and categorizing process to be applied to every component of data.

The second cycle coding enabled the codes and data to be cycled comparatively to ensure against the loss of concept and meaning through the coding process (Saldaña, 2013). Miles, Huberman, & Saldaña (2014) describe this second cycle as generating pattern codes, looking for similar codes or clusters of codes that lend to similar interpretations. Terms such as categorizing, theming, family codes, pattern codes have been used to describe the process of reducing the coded categories from the first cycle to fewer, but larger conceptual categories in the second cycle (Maxwell, 2013; Miles, Huberman, & Saldaña, 2014; Saldaña, 2013). For my purposes, I refer to the second cycle codes as categories, as the concept of themes will be defined and used in a different manner as part of the data analysis process. A

matrix display using Microsoft Excel was incorporated, capturing the data points, initial codes and categories in a visual manner allowing me to further analyze connections and to also create opportunity to view potential gaps in the connections and relationships of the data points (Creswell & Poth, 2017; Maxwell, 2013; Miles, Huberman, & Saldaña, 2014).

Narrative Analysis. Narratives, as part of their nature, create greater meaning when regarding as part of the larger context (Clandinin & Connelly, 2000). Reissman (2008) declared high quality data is lengthy and include multiple complex themes not easily broken into smaller parts. "Critical and illuminative contextual and structural elements of a story can be lost when it is coded narrowly and reduced to bite size units" (Patton, 2015, p. 131). The components of Clandinin & Connelly's (2000) three-dimensional narrative inquiry space waere used to analyze the participants' narratives as illustrated by in-depth interviews and narrative prompts with field texts used to make meaning of these sources. The three-dimensional narrative inquiry space includes: (a) interaction—personal and social interactions, (b) continuity—past, present, and future, and (c) situation—physical spaces and settings (Clandinin, 2013; Clandinin & Connelly, 2000; Creswell & Poth, 2017). While any one component of a participant's data may reveal different aspects of the three-dimensional spaces, the examination of the collective narrative of each participant revealed a more complete picture. Each component of the narrative shared revealed a personal or social interaction to illuminate the interaction space of the three-dimensional inquiry, but examining the entire set of shared experiences, I was able to glean whether the participant views the phenomenon as being a unique and solo experience or a more collective experience based on the other individuals included across the varied stories. Equally, the natures of continuity and space illustrated by the shared experiences highlighted the perceived phenomena from a

126

greater scope than provided by any one component of the narrative separate from the others. The purpose of collectively examining the narratives of each participant was to gain greater understanding of the lived experiences of each participant.

To further increase understanding, and to avoid contextual and thematic loss, a process of re-storying with member-checking was included. As I analyzed the narratives provided by participants using the three-dimensional inquiry space, I created a cohesive story reflective of the individual participant's stories. Using reflexive practice, guarding against my own bias and interpretive lenses, I generated the story of each participant's lived experiences. These in turn were shared respectively with the participants providing the source material along with transcripts of interviews and other materials shared by the participants. "The inquiry space, and the ambiguity implied, remind us to be aware of where we and our participants are placed at any particular moment—temporally, spatially, and in terms of the personal and the social" (Connelly & Clandinin, 2000). As the participants reviewed their stories and shared experiences, each had the opportunity to share additional stories and information to influence the captured narrative generated by me. I, in turn, revised the greater narrative reflective of the participant's newly shared experiences. This began a cycle of re-storying until the participant agreed the collected narrative captured his or her experience with the phenomenon being examined. It was through this cyclical process of re-storying that a researcher can gain greater insight and understanding of the experiences of the participants to best answer the questions posed by the researcher (Estafan, Caine, & Clandinin; 2016).

Within Case and Cross Case Analysis. Creswell and Poth (2017) describe interpretation as moving beyond coding and categories into larger units. These larger units,

themes and concepts, allow access to addressing the research questions posed by the study (Creswell & Poth, 2017; Saldaña, 2013). For each case, *gestalt,* the holistic understanding of the case in relation to the phenomenon, was achieved. "Interpretive techniques designed to be used within individual accounts or cases provide a wealth of contextual richness and person-specific information without which that case cannot be understood" (Ayres, Kavanaugh, & Knafl, 2003, p. 873).

Each participant represented an individual case. Analyzing the collective data of each participant represented within-case analysis. Within-case analysis occurred by pulling apart all of the components and examining their characteristics and connections that created the whole (Creswell & Poth, 2017; Stake, 1995). The gleaning of understanding generated from each case was juxtaposed with the research questions to determine new understandings from the study.

Upon completion of examining the individual cases, the collection of all the cases was analyzed as a whole set. This is referred to as a cross case analysis. A cross case analysis was used to interpret themes and concepts across the multiple cases. Overarching themes were gleaned across the cases to further increase understanding of the phenomenon and best provide insight into answering the research questions from multiple perspectives. While I had no intent to generate generalizability from such a small sample size, the layering of findings from the multiple individual cases provided greater insight due to commonalities and differences present from the lived experiences of the participants.

In my roles as a classroom teacher and later as an administrator, teacher evaluation continues to be a demand in my list of responsibilities. Through my experiences, I have grown to recognize the complex nature of teacher evaluation and student growth measures,

and I believe it is possible to successfully incorporate student growth measures into teacher accountability. I also believe, at this time, that an effective, accurate method of student measurement has not been utilized to determine student growth; nor do I believe a system has been created to fairly balance the demands on teachers' duties between teaching the content measured on summative exams along with the holistic child development expected of teachers. As discussed previously, reflexive practice on my part was crucial to protecting the validity of the study, and this practice and other measures to ensure validity will be discussed in more detail in the following section, "Limitations including Validity, Reliability, and Ethical Considerations." My pursuit in advancing this study was to capture the perceptions and understandings of teachers as the context of teacher evaluation evolves through education reform. This desire has guided my design decisions including data sources, collection, and analysis methods.

Limitations Including Validity, Reliability and Ethical Considerations

The purpose of qualitative research is to increase understanding. Qualitative data captures and communicates peoples' experiences at a particular time and place with depth and detail (Patton, 2015). When researchers wish to know a more holistic understanding beyond 'how well,' qualitative research provides the avenues (Fraenkel et al., 2012).

> Because a research design is supposed to represent a logical set of statements, you also can judge the quality of any given design according to certain logical tests. Concepts that have offered for these tests include trustworthiness, credibility, confirmability, and data dependability. (Yin, 2014, p. 45)

These tests are used to measure different forms of validity such as external, internal, and construct, as well as reliability. As with all parts of qualitative research, these are open to interpretation and influence by researcher bias (Yin, 2014). To this end, it becomes increasingly difficult to define validity in its standard form when discussing the data

collected and analysis performed. "In recent years, validity, has been defined as referring to the appropriateness, correctness, meaningfulness, and usefulness of the specific inferences researchers make based on the data they collect" (Fraenkel et al., 2012, p. 148). Qualitative research requires data to be trustworthy, relevant, full of thick description and authentic (Maxwell, 2013). The researcher is the instrument collecting this data as well as analyzing it. To believe a researcher could set aside all bias and personal perspective when performing such a task is unrealistic (Maxwell, 2013). With this in mind, the researcher must be very transparent in sharing any bias or perspective that could influence the collection or interpretation of the data. "Note that qualitative researchers often use the term credibility to encompass not only instrument validity and reliability but internal validity as well" (Fraenkel et al., 2012, p. 458). Through sharing all potential influencers, the researcher establishes credibility, not by eliminating personal perspective, but by providing the audience with a clear understanding of how the researcher's presence and understanding may have influenced the data and interpretation. Additionally, the concept of crystallization is incorporated into the study to reinforce validity within each case. Crystallization is defined by the use of multiple forms of data, across multiple forms, genres, or presentation styles to increase the number of lenses or viewpoints with which the phenomenon and participants may be analyzed (Ellingson, 2009; Ellingson, 2014; Richardson & St Pierre, 2008). The use of crystallization allows for a holistic understanding or *gestalt* of each case prior to analyzing the cases comparatively.

Limitations

I recognize five threats present within this study, which act as limitations, and I will address how I guard against them when discussing validity and reliability. First, the data

selected for this study was in itself a limitation. I am unable to observe the formation of teachers' perceptions. Even if I was able to observe them in the process of being evaluated, I would still be unable to fully understand how they relate and understand the process. I must rely on conversations with the participants and analyzing related materials. This results in layers of interpretation lending themselves to possible misunderstandings, missing relevant information or connection, or misdiagnosing connections and relationships. Included within this limitation is the fact that all participants are European American and all but one are male. As such, there could be uniqueness or narrowing to their perspective that could have been realized provided a more diverse collection of participants volunteered to participate. A second limitation stems directly from the first, and that is the interpretation of the data through my personal lens and bias. Having had multiple personal experiences with the phenomenon of evaluation in different roles, I have created my own perceptions, which must be guarded against to prevent them from skewing the analysis and any potential development of understanding of the phenomenon. As with the previous limitation, recognizing my own racial and cultural background represented by the participants had to be guarded against as well. A third limitation was reflexivity during the interviews, possibly causing reactivity. As discussed in the interview section, the interactions present during an interview may cause the information provided by participants to be skewed. This could have occurred through body language, question phrasing, tone of voice, facial expression or any number of other cues provided between people interacting in conversation. Whether the participants were attempting to provide pleasing information or are attempting to be abrupt to speed the process along could have limited the accuracy of data collected. Fourth, would be narrative smoothing. It is important to realize the nuances present in narratives including author

reliability, point of view, audience, what is highlighted in the story, as well as recognizing those components or pieces not present in narratives. Finally, there was the fact that I held the role of a principal and am charged with teacher evaluation. While I am not the evaluator of the participants of the study, I did fulfill those duties in my school and as such hold my own beliefs and understandings related to the experiences.

Validity

Many of these limitations held overlapping dilemmas and as such were addressed by similar precautions. Multiple strategies were employed to buffer against these potential threats including the acquisition of thick description, respondent validation (member checking), discrepant evidence, comparison, wakefulness, and peer debriefing. These strategies established levels of trustworthiness between the reader and researcher through transparency. The relevance and contribution to the field is apparent and supported by the credible nature established through the openly shared use of the strategies mentioned previously.

The collection of thick descriptive data established a deep pool to resource the development of understanding of teacher perceptions. This detailed and diligent description of context and phenomena established credibility and relevance to other settings (Carlson, 2010). The inclusion of the data in analysis as well as reporting guarded against bias present in the selection of data as it is apparent to the reader the nature of the data from which understandings were created (Creswell & Poth, 2017; Maxwell, 2013; Fraenkel et al., 2012). In this sense, it also contributed to providing insight to the audience about the establishment of interpretations and where they originated. Wakefulness, "it is a question of being alert to the story not told as you are to those that are told" (Clandinin & Connelly, 2000, p. 182). In

regard to potential narrative smoothing, wakefulness is the inclusion of thick description to illustrate understanding by what is present and by what is not.

Reliability

Reliability refers to the concepts of accuracy and trustworthiness in regards to the data (Creswell & Poth, 2017). "Since there can be no validity without reliability (and thus no credibility without dependability), a demonstration of the former is sufficient to establish the latter" (Lincoln & Guba, 1985, p. 316). Member checking, or respondent validation, was used to guard against bias and reactivity as well. By allowing participants to review collected materials and written components based on interactions and collected data, it provided opportunities to confirm understanding, interpretation, and effective communication (Creswell & Poth, 2017; Maxwell, 2013; Fraenkel et al., 2012). As language is a limited medium at times, it was important to verify understanding not only to strengthen data collected but to also negotiate the relationships with participants. This also created opportunities to examine for reflexivity and reactivity between the participants and the researcher. Participants were able to contribute in guarding against any bias present from my perspective of administrator and carrying my experiences or perceptions too heavily into write-ups.

The analysis of negative cases and incorporating discrepant evidence pushed a deeper reflection of potential bias of the data selected to use or not use, its interpretation, and reactivity present in the interactions between the researcher and participants (Creswell & Poth, 2017; Maxwell, 2013; Fraenkel et al., 2012). Comparison across multiple cases as well as with-in case analysis increased the likelihood of representative participants (Yin, 2014; Creswell & Poth, 2017; Maxwell, 2013). While generalizability to a much larger group will

not be possible or likely, limiting the study to only one case would have increased the likelihood and increase the question of whether the single case might not be a more extreme outlier than closer to a more representative figure. I was also aware of the dynamics of power connected to my role as an administrator in the district. Peer debriefing, or an external audit, was incorporated to reign in bias and the dynamics of power that might have influenced the collection and interpretation of data due to my role and experiences as an administrator in charge of evaluating teachers (Creswell & Poth, 2017; Fraenkel et al., 2012). These were the potential limitations and how I guarded against them.

Ethical Considerations

My role as an administrator is one of many ethical considerations that was evaluated when appraising how best to protect the participants of the study while contributing to the field. The Belmont Report (1979) established three areas of attention regarding the application of research processes: informed consent, assessment of risk and benefits, and selection of subjects. The study remained well within the direction of each of these three areas of consideration. The selection of subjects, as defined by the report, refers to providing favorable selection for participation to particular groups or to focus potentially negative consequences of research towards a particular group. This study did not provide notoriety nor negatively identifiable information to others. Participants did not receive any financial reward or incentive for participation. The participants may have benefited personally from participation according to the reciprocity model as they had the opportunity to express their ideas, concerns, thoughts, feelings, could have felt esteemed due to being asked to participate, or hoped their participation led to changes made in their contextual system on some scale (Patton, 2015).

To prepare for the study and to knowledgeably protect subjects participating in the study, I participated in Collaborative Institutional Training Initiative (CITI) coursework targeting social, behavioral, educational research. This course work covered ethical principles, human subjects research, federal regulations, assessing risk, informed consent, privacy, and confidentiality among many other topics. Successfully completing the program included passing its associated exam at the conclusion of the coursework. The CITI was completed in preparation of the proposal for submission to the Institutional Review Board. Prior to beginning the selection of subjects or collection of data, the study was submitted for review and approval to the university's Institutional Review Board (IRB). The IRB evaluated the study's proposal to determine if the previously mentioned criteria of the Belmont Report were acceptably met. Approval of the board and adherence to the approved proposal qualified the study as acceptably meeting those aspects of protection of participants.

Another component of consideration was informed consent. In seeking approval through a review board, the study was outlined in detail including processes and steps for participant identification and interaction. This created difficulty in negotiating relationships with participants and building the study collaboratively once the participants had been recruited. Also, by implying that participants were capable of being informed of every aspect of the research and discovery process, implied there was no cause to conduct the study as the results and experiences were already known (Clandinin & Connelly, 2000). With these challenges present, all participants were fully informed of all planned procedures, actions, and uses for the collected information. Participants were volunteers and had the ability to withdraw at any time of their own choosing.

All necessary steps were taken to reasonably secure all materials that could impinge on the anonymity of participants. The interests of security and privacy were maintained at all times. The survey results were downloaded and stored on a password-protected computer. At the conclusion of the interviews, the audio recordings were transcribed; any additional notes and memos were completed following the transcriptions. The original and all copies of the recordings were deleted following participant verification of the transcriptions accuracy. All transcriptions and notes were made electronically and saved on the password protected computer. The file names of the electronic documents did not include personally identifying information.

Finally there was the assessment of risks and benefits. The benefits consisted of the contributions to the field and the increased understanding of their perceptions on the topic. Risks were guarded against throughout the study. Adhering to the concept of "do no harm," anonymity was maintained for all participants. All names were altered in written materials. All recorded materials were transcribed then deleted. Participants had the opportunity to establish stories and reasons for researcher's presence to peers, colleagues, and others present (Clandinin & Connelly, 2000). The three considerations for application were based upon the principles of respect for persons, beneficence, and justice. Throughout the study, the same methods used to guard against threats of the study were used to maintain adherence to these same principles.

In summary, this study focused on six participants with variance in subjects taught and years served at Hawkins Middle School. Participants were self-identified through the survey, and they were able to reflect and communicate about the topic of the study in thick, rich description. The research methods included: surveys, interviews, narratives, and

documents. Throughout the process, the researcher was reflexive and reflective in nature in order to adapt the study to the developments that occured while not ranging so far afield as to jeopardize the ability to answer the research questions. The multiple data sets were analyzed using analytical approaches that fit the various traditions. For example in-depth interviews and narratives were interpreted using Clandinin & Connelly's (2000) three-dimensional narrative inquiry space. Surveys and documents incorporated descriptive and thematic coding. Hence, within-case analysis was followed by cross-case analysis to develop deeper understanding of each individual case and allow for comparisons across cases. While it was a researcher's priority to protect the validity and credibility of the study during its pursuit, there was an equal, if not greater, responsibility to protect the participants of the study as discussed in response to proposed ethical considerations.

The remaining chapters detail the description of raw, collected data along with refined results gleaned from the data. Chapter four portrays the finding of the research including the thematic and narrative analysis within-case and cross case analyses. Chapter five presents the research questions and discusses the implications for current practice and possible future research in the educational setting.

CHAPTER 4

FINDINGS

As school districts implement education reform efforts, the focus at times is on poor teacher performance and the inability of school district's to address this root cause. The blame for such contributors should not be placed at the feet of the teachers while job performance appraisal systems are riddled with inaccuracies and subjectivity. Efforts have been taken to improve such systems leading to criteria based evaluation systems incorporating multiple measures of teacher and student performance for summative appraisal, which then leads to personnel decisions. As these particular reforms increased and became commonplace as result of the recent leadership and policies generated by *No Child Left Behind, A Blueprint for Reform* , and now *Every Student Succeeds Act* (ESSA) (U.S. Department of Education, 2015) teachers have become the subject of reform actions without having the opportunity to contribute to their development. The purpose of this narrative multi-case study was to explore the perspectives of teachers experiencing the changes at the building level of teacher evaluation reform efforts.

The theoretical traditions used to explore the teachers' perspectives include narratology and multi-case study. The narratological tradition is exemplified in the design by the collection of the stories of those experiencing the changes central to the topic (Patton, 2015). By collecting the written stories detailing the teachers' perceptions of their work in schools and how it could be or should be measured from their point of view will allow for comparative analysis to that of what is suggested in legislation. The use of multi-case study is intended to create thick, rich description in a context rich environment from multiple

perspectives and ranges. The actions, beliefs, and events situated within the contexts in which they occur provide focus and depth of understanding (Creswell, 2013).

The research questions were selected in order to more fully investigate teachers' perspectives and understanding. The primary question being: How do teachers perceive the use of student growth measures in performance evaluations for core content teachers at an urban mid-western middle school? Three sub-questions provided guidance for analysis: (1) What do teachers understand about the context of job performance evaluation? (2) How do teachers define student growth when measuring their own performance? (3) How do teachers describe performance evaluation?

The study took place over the course of eleven months beginning in mid-January and incorporated multiple data sets, including surveys, documents, interviews and narratives. Table 2 outlines the research process, including IRB approval, sampling for participants (survey) etc. After a brief presentation of the purpose and methods of the study to the middle school staff a survey was sent to all staff members. The last question of the survey allowed potential participants to identify their interest in moving to the interview phase of the study and subsequent data collection activities. Six participants self-selected from a survey of the teaching staff of an urban mid-western middle school to contribute through interviews and provide narratives of their experiences with evaluation. The intention was to use purposeful criteria based sampling, but with the limited number of willing participants, all were included to participate. If more participants had volunteered, the purposeful criteria used to select participant would have been intentional towards identifying participants with the greatest possible range of experiences and perspectives including age, years of experience, race, gender, and content subject. Contributing factors that may have influenced non-participation

include personal responsibilities limiting available time, spring sports beginning and many teachers coach or have kids of their own participating, participation occurred during the state assessment window which for many teachers is a high stress time, and new district leadership had begun new initiatives requiring teachers' time to learn and implement. Each participant teaches at the same middle school in an urban mid-western school district currently undergoing a reform process related to the teacher evaluation system. The participants have a range of experience from four to 33 years and include both male and female participants.

Using the communication method identified by the participants, I contacted them to set up an interview at a location of their choosing. The interviews were semi-structured and conducted on an individual basis lasting approximately 75 to 150 minutes. The semi-structured nature allowed for a scripted guide while also allowing for the conversational tone to carry the interaction, as it didn't require adherence to a rigid script. Narratives were also collected during this time in the form of oral stories via prompts. Documents included were collected from the state department of education website related to teacher evaluation adoption, expectation and implementation. Additional documents included were gathered from the school district's human resources department, building principal, and participants and included emails, guidelines, implementation practices, training handouts provided to teachers, and an observation protocol used by the building administrator.

Table 2

Outline of Research Process

Process	IRB Application	Sampling of Participants (survey)	Interviews & Narrative Collection	Narrative Analysis & Re-Story	Thematic Analysis	Report of Findings
Timeline	Jan.-Feb.	March	April	May	June-Aug.	Sept.-Nov.

As a new acquaintance with each participant, I wished to respect any personal boundaries they may hold as well as protect their confidentiality as well. With this in mind as I scheduled interviews, I made a point of encouraging them to select a location they would feel comfortable sharing and speaking openly on the topic of the study. With each I recognized they may not wish for me to meet in their school in order to protect their confidentiality. I also expressed a desire to respect their personal lives by holding the meeting at neutral locations rather than their home, acknowledging the inconvenience it may impose to ask them to come to me either at home or work. To this end, different locations were agreed upon with each participant that they were comfortable speaking. Two participants selected different coffee shops. Another participant selected a reserved meeting space at a public library near the participant's home. A third interview was in a participant's home, and two participants selected a local restaurant due to the hours of their availability following a full day of teaching and coaching after school. I assigned pseudonyms for each participant. These were generated by compiling a list of names and then confirming to the best of my ability using staff lists and the school website that no staff member current or in recent years had each name used. Participants were able to approve the pseudonym used for him or her during the member checking process. Pseudonyms were also used for

administrators, school leaders, and other people mentioned in teachers' stories in order to respect their confidentiality.

Each interview began with small talk related to their day or week, events occurring at their school and an intentional attempt to build connection with what they shared with my own school experiences and workweek. I would highlight commonalities in students' and colleagues' recent actions to build rapport. If they had shared extra-curricular activities they sponsored, I would inquire about those as well, again, sharing my own similar experiences as a former teacher and current administrator. This initial small talk would be followed with a review of the purpose of the study and the consent information. The initial questions of the interview guide are descriptive in nature and were intentional to ease participants into sharing their work life and experiences (see Appendix). At the conclusion of the interview, I would inquire what method of communication they would prefer phone, text, e-mail, standard mail, or personal delivery for follow up questions and to provide a copy of the interview and narrative transcripts. Future interactions for follow up and confirmations were made using the participants' preferred method of communication.

As with many first time researchers, even as we guard against bias and influencing our own thinking, I subconsciously expected my own experiences and ideas to be reinforced by the experiences and stories of others. My purpose in pursuing this topic resonated with recognition for potential in a system to support and develop teacher practices, but my own experiences repeatedly illustrated such systems as ineffective. As the study progressed, I continued to recognize the potential for teacher evaluation to support teacher development and even the potential for student growth measure inclusion in such a system. My growing realization over the course of the study highlighted the potential only being as great as the

skills of those involved. As I interacted with participants, their experiences highlighted all the variables associated with interactions between principals and teachers including communication, expectations, relationships, and time investment among others. In order for student performance to be a meaningful measure of teacher performance, the evaluator should account for many other variables that impact time, attention, and available support.

I took intentional actions to guard validity and reliability throughout the study. Participants reviewed transcripts of their interviews and were given opportunities to clarify or add detail to their original responses. The participants own words were used throughout the inductive coding process for descriptive and interpretive codes, known as in vivo coding (Boeije, 2010). The three-dimensional inquiry space was used to analyze the narratives. The three-dimensional inquiry method includes sociality, temporality, and place (Clandinin & Connelly, 2000). Interaction occurs from the researcher analyzing the experiences present in the story and the lens of the narrator. Continuity includes analyzing the narrative across the chronology of the story as well as including what is known before and after the events of the narrative. The situation represents the locations or settings in which the narrative occurs with some examples from the participants of this study including empty classrooms, classrooms with students, school hallways, and administrators' offices. Wakefulness was practiced throughout to guard against being aware of what was not shared or emphasized as well. Additionally, a re-storying process was incorporated to highlight essential experiences, and these were also resubmitted to participants for confirmation of capturing the essence of their experiences.

The repeated return to each participant following the survey, interview, transcript of the interview, narrative transcript, and re-storying generated an increased reliability as it

143

provided for correction by each participant at multiple steps along the way. This also guarded

against my own bias or reflexivity skewing the collection or analysis of the data, as the

participants would have corrected or adjusted any interpretations or presentations of

inaccuracy. The incorporation of multiple cases allows for the examination of negative cases

and while not representative of a larger whole, increases likelihood of representing the

smaller population sample. Finally, I worked with a peer to discuss the potential bias and

power dynamics I may bring to my analysis of the data throughout the collection and analysis

stages keeping awareness of that potential hurdle at the forefront of my mind and strategies

to protect against it.

Data analysis

The data sets for this study include documents, surveys, in-depth interviews, and

narrative prompts. Most documents collected apply to all participants and were maintained as

a data set analyzed separately. The surveys primarily consisted as a method for participants

to self identify willingness to participate. In addition to that role, they include descriptive and

demographic questions providing details that would have come up during the interview but

could be expedited through the survey respecting participants time during the interview

process. Two open-ended questions were included in the survey, and participants' responses

to these questions were included as part of the thematic analysis associated with the

individual participants.

The in-depth interviews included many questions like the survey requesting details

describing participant's roles, responsibilities, priorities and understandings. These

responses, along with participants' survey responses, and if individual's provided additional

documents of their own, such as Eric providing past evaluations, in addition to my journal

notes describing observations and reflections following interactions with participants, comprised the data sets for individual cases to be analyzed through multiple cycles of inductive coding and generating of themes. The in-depth interviews also included many open-ended, reflective questions, and multiple narrative prompts resulting in the generation of stories provided by the participants.

Clandinin and Connelly (1996) defined three types of stories shared by teachers: sacred stories, cover stories, and secret stories. Sacred stories include those recountings provided in reports, public descriptions, and the ideal versions of what occurs in the classrooms. Cover stories are those shared with people familiar with the setting such as fellow teachers or parents. Secret stories include the most factual, accurate events typically reserved for those present in the classroom, typically the teacher and students. The collection of sacred and cover stories shared by each participant create a data set for that individual as well which was analyzed using the three-dimensional inquiry space.

A variety of documents were collected to provide context and expectations of systems and individuals related to teacher evaluation. Among these documents were relevant excerpts from the *Elementary and Secondary Education Act* (ESEA), *NCLB* and *Blueprint for Reform* adoptions and updates. Additionally, the state department of education provided such documents as: state vision for education, list of student performance indicators for teachers, descriptions of state approved evaluation systems, inter-rater agreements and expectations, evaluation timelines and deadlines, evaluation system requirements, state performance rating matrix, state summative scoring matrix, educator evaluation fact sheet, definitions of student performance and identifying indicators, approved application for state ESEA waiver, and state educator evaluation handbook. The human resources department of the school district

included documents such as an outline of evaluation indicators, evaluation timelines and deadlines for teachers of various years of experience, calendar of scheduled trainings for administrators, presentation handouts from administrator trainings related to teacher evaluation. The school principal provided the walkthrough tool used in his building as well as an outline effective teaching indicators looked for in classrooms during walkthroughs and observations.

The analysis of the data collected from the documents involved multiple cycles of inductive coding, the use of descriptive and interpretive coding pulled from the language present in the data. This process involves breaking the data down into digestible fragments that may be categorized for comparison. These components can then be regrouped and collected into broader themes and issues (Maxwell, 2013). The initial descriptive codes used during first cycle analysis were generated from raw data within the language of the documents. The second cycle analysis collected the results of the descriptive coding into particular interpretive codes for cluster analysis. Finally, these interpretive codes were collected into themes with each set of data informing the next and returning to the previous for deeper analysis and stronger connections (Grbich, 2013).

Each participant's survey, interview transcripts minus the narratives, participant supplied documents, and field documents made up individual case data sets. Similar to the inductive coding process used for document analysis, these data were analyzed using first and second cycle inductive coding resulting in interpretive codes and themes. The initial descriptive codes used came from the participants' language in their responses. The generated themes provided insight and understanding of the participants in relation to the research questions (Maxwell, 2013).

The narrative data collected from each participant's in-depth interviews and narrative prompts were examined in the three-dimensional inquiry space of sociality, temporality, and place. Relying heavily on reflexive practice to guard against my own biases, I re-storied the varied shared, lived experiences combined with understanding gleaned from analysis of the three-dimensional inquiry to present a written narrative for each participant to capture cover stories. Through a member-checking process, each participant was provided the opportunity to direct, clarify, add, or revise this collective narrative of his or her experiences creating a cycle of revision and resubmission for member-checking until the participant was satisfied with the narrative's accuracy. The use of multiple forms of data and data sets increases the number of viewpoints from which the phenomenon and participants may be analyzed resulting in crystallization (Ellingson, 2009; Ellingson, 2014; Richardson & St Pierre, 2008).

Document Analysis

Educator evaluation has assumed many forms and approaches over the years (Blumberg, 1985; Hunter, 1980; Lewis & Leps, 1946; Reavis, 1978; Taylor, 1914; Tracy, 1995; United States, 2010; Wetzel, 1929). As outlined earlier, I incorporated in the study such federal documents as Elementary and Secondary Education Act (ESEA), *NCLB* and *Blueprint for Reform* adoptions and updates; state department of education documents, and those at the local district level that involved human resources guidelines for evaluation and principal walkthrough tools that stressed effective teaching indicators. New guidelines expected by the US Department of Education as illustrated by the State Department of Education's waiver request (2012) define evaluation to be "a systemic determination of merit and significance, of someone using criteria against a set of standards" (p. 38) and "…refers to the actual assessment of the teacher's or principal's effectiveness, not necessarily to the

147

number of times a teacher or principal is observed" (p. 3). Documents such as this one and others gathered from federal, state, and district sources provide the context within which the teachers' experiences are embedded. Reform efforts begun at the federal level influence state level decisions and demands upon school districts which adopt policies and practices to be in alignment resulting in building actions, cultures, and climates resulting from their implementation.

Four themes from document analysis illustrate the context of the teachers' experiences: **Evaluation Process**, **Performance Rating***,* **School Culture,** and **Teacher Preparation and Performance.** Each theme is defined by interpretive codes derived from clusters of descriptive codes, which were drawn and defined by participants' language in surveys, interviews and narratives.

Evaluation Process. The first theme derived from the documents was Evaluation Process including what occurs or unfolds as a person is evaluated. The first interpretive code, *inspection,* provided body to the theme of Evaluation Process. Almost every document analyzed included some reference to components falling within the observation act or other phenomena related to observing and monitoring teacher actions. From the administrator training presentation handouts, one of the slides outlined the number of observations required for different staff members:

> Formal Observation Requirements:
> All certified staff being evaluated must have the following number of formal
>   observations:
> • First and second year teachers have four formal observations (2 per semester)
> • 3rd and 4th year teachers have three formal observations per year
> • Tenured teachers have three formal observations per year
> • Off cycle evaluations have three formal observations per year
> • Teachers on warning status are evaluated according to the Teacher Negotiated
>   Agreement. You will work with your HR Director and Advisor on the process

Another slide from the same presentation handout identified the components of a required, formal observation:

> Types of Observations
> Formal Observations are required to include the following:
> 1. Planning Conference (pre-observation conference)
>   a. Evaluator uses the planning conference questions to guide the pre-conference and elicit evidence, review lesson plans
> 2. Classroom Observation
>   a. The observer should remain for the entire lesson OR determine in the planning conference what period of time will allow the observer to see instruction and student evidence
> 3. Reflection (post observation conference)
>   a. Teacher and Evaluator will examine student evidence to arrive at scores.
>   b. Use reflection questions in evaluation tool to guide the discussion and elicit evidence

The evaluation model detailed what should be looked for during evaluations, timelines presented when observations should occur during the school year, training presentations highlighted how observations and discussions of such should occur, and building level created observation protocols emphasized look-for items in individual classrooms.

Proof developed as the second interpretive code upon formation of acts related to performance, assessment, and other evaluation tasks. Documents focused on teacher performance, the act of teaching, evidence of teaching, or evidence of performance. References to assessment, types of assessment, student performance documented by assessments contributed to the interpretive code of *evidence* as well. These interpretive codes capture the act of evaluating in the theme of Evaluation Process, and the next theme illustrates what occurs after the act of evaluating concludes.

Performance Rating. The theme of Performance Rating holds the outcome of the evaluation process including the interpretive codes of *measurement* and s*core*. The state education department (SDE) published on their website a list of more than 70 student

performance indicators to choose from to use as a large portion of a teacher's evaluation score. These measurements include state assessment results in addition to a variety of other methods to measure student learning including grades, local assessments, commercially available assessment tools, and student work. The SDE Educator Evaluation Fact Sheet shows how student performance results should be included in an educator's final summative rating. As part of their guidance for defining student performance and identifying indicators, the notes recognized that districts were required to include student measures as part of teacher performance ratings, but many districts had interpreted the statue as optional. As a result the following note was included as part of the guidance:

> Ironically, state statute already expected that student performance be a part of educator evaluation. It didn't happen to the extent intended. SDE legal counsel has recently (January 2016) interpreted the statute addressing student performance as part of the evaluation not as an option, but as an expectation.

SDE also published a performance matrix explaining tabulation of teacher performance scores that required the use of at least three student performance measures. This matrix also indicated that teacher scores should be categorized into highly effective, effective, developing and ineffective and the method for computing such scores. Prior to earning summative scores educators engage in practices to prepare and improve their practice as illustrated by the next theme.

Teacher Preparation and Performance. The theme of Teacher Preparation and Performance involves the teachers, how they interact with the work and each other. The interpretive codes are *pedagogy* and *education*. The interpretive code *pedagogy* focuses on continually improving professional practice through job embedded professional development, coaching and feedback. A slide from an administrators' training presentation handout outlined expectations for feedback:

What Should Be Included In Teacher Feedback?

- Feedback should identify the goal of the teacher and the level of performance
- Feedback should be specific regarding teacher and student actions during the observations
- Feedback should be meaningful to the teacher
- Feedback should identify next steps for improvement and teacher growth

The most frequent phenomena include professional development, content knowledge and certification.

Education builds upon the acts of instruction and growth. Instruction is expected to be effective and increase student understanding and knowledge in a measurable manner. Additionally, not only is student growth measured, but teacher growth should also be measured. Teachers are to be rated across a performance scale with the expectation of growth over the year and over the years measured by observation and student performance.

School Culture. The final theme generated by document analysis was School Culture, defined as where the evaluation act occurs, influenced by the context and relationship structures developed by those involved. This theme consists of *guidance* and *climate*. This theme has a frequency count half of the next lowest theme. *Guidance* focuses on the input and support of others to improve practice such as receiving feedback. As will more clearly be illuminated within the participant analysis, even if feedback is present, it does not automatically improve teacher practice and performance. The setting and context within which it is received affects the interpretation and usefulness. The relationship between those providing and receiving feedback influences the result of the feedback.

Climate includes leadership, responsibility, expectations, culture and relationships constituting the context within which everything occurs. The SDE provided direction regarding inter-rater agreement, the degree in which two raters using the same rubric give the same rating to similar experiences, in order to increase validity and reliability among

151

evaluators. The document addresses communication and understanding between evaluators

and evaluates as well:

> Achieving IRA within a district starts with building a Common Frame of Reference
> among all educators so that the evaluation system works consistently across the
> district. All educators—evaluators and evaluates alike—must develop a common
> understanding of what excellent instruction "looks like" in the classroom.

Similar to *guidance*, the underlying relationships with the environment, peers, and leaders

influences peoples' actions and interactions in a school. These undefined components result

in the climate of a school and influence communication and understanding.

Education reform efforts, including NCLB and before that ESEA, have established

the intent to improve the American education system. In order to do so, these acts and others

have incorporated practices demanding the implementation of effective practices including

those related to evaluation.

> The SDE recognizes that along with the education community, it is our collective
> responsibility to support an evaluation system that provides formative feedback to
> those being evaluated so that performance can improve over time and the evaluation
> system can contribute to student achievement. (p. 2)

The themes of **Evaluation Process**, **Performance Rating**, **School Culture,** and **Teacher

Preparation and Performance** represent the reform efforts as they apply to the context in

which the participants work on a daily basis. From the guidelines at the federal level to the

state authors of the waiver request to district and building leadership and finally to the

teachers impacted by changes, the information flows through many channels, interpretations,

and attempts at communication before arriving with those most directly impacted by the

policy changes. Through interviewing those impacted, collecting their stories, a clearer

picture can be drawn of the understanding and perceptions that exist.

Within-case Analysis

The perceptions and lived experiences of each participant make up individual cases consisting of two distinct data sets. As previously noted, the first includes the survey data, interview transcript minus narratives, participant supplied documents, and field documents. The second data set consists of the narratives provided by the in-depth interviews and narrative prompts. Each data set were analyzed as previously described and the combination of the results from the analysis of these two data sets make up each case, organized into two parts. The first part of each case is a re-story created from narrative experiences shared and generated from an analysis using the three-dimensional inquiry space (Clandinin & Connelly, 2000). The second part of each case consists of the themes derived through the creation of interpretive codes. Each interpretive code developed from multiple descriptive codes within and defined by participants' language and experiences as shared during the survey, interview, and narratives (Grbich, 2013). Each case captures the experiences most closely associated with teacher evaluation practices and understanding of how to best measure teacher performance in an era of reform and change. The following cases of Eric, Paul, Ray, Nancy, Lance, and Aaron share their experiences and presented in no particular order as each holds equal value.

Case 1 - Eric

"Remember, everybody, even though there is no assignment due tomorrow, there will be a unit quiz. Your homework is to study and be prepared for what we have covered so far this week!" said Eric as his students packed up to leave for the day. He smiled as the last of them filed out the door, and as he watched them leave, he saw his principal waiting for them to pass before she came into the classroom. Eric held his smile in place even as he felt his

good mood begin to drain away. That hadn't always been the case when Mrs. Sherman came by his room.

When Eric began teaching, Mrs. Sherman constantly provided coaching and support. She would drop by the class routinely, offering notes of encouragement and advice for improving different parts of his lessons. His first couple of years, Eric truly relied on Mrs. Sherman, going to her when facing difficult choices about how to handle student conduct or to help find different resources for his students and lessons. But that was before her promotion. As the assistant principal, Mrs. Sherman went out of her way to help many staff members improve their practice, but as she took the head principal role this past fall, everyone slowly quit going to her for help. Rather than offering advice, she only offered judgment. When asking her how he might support one of his struggling students, Mrs. Sherman told Eric, that having a struggling student was unacceptable, and he needed to improve his instruction, rather than pointing out he was trying to do so with her guidance, he simply closed his mouth and left her office. Over the first few months of school Eric and many other teachers have slowly withdrawn from engaging with Mrs. Sherman unless absolutely necessary.

As Mrs. Sherman entered Eric's classroom, he noticed she was holding some papers in her hand, and she had a stern look upon her face. "Mr. Carlson, I've completed your evaluation for the fall semester. Here you go," she said as she handed Eric the papers in her hand.

"I didn't realize I was being evaluated this school year. I don't believe I'm on cycle for evaluation again until next year," said Eric beginning to glance at the papers she had handed him.

"As the principal, I can choose to evaluate any staff member whether they are on cycle for evaluation or not. I've chosen to evaluate you and a few others that weren't on the schedule for evaluation to make sure everyone stays on their toes," Mrs. Sherman stated, turning to leave.

"Mrs. Sherman?" said Eric, trying to gain her attention.

"Yes?" she said turning back to him with only a hint of exasperation in her voice as if he had been the one delivering the surprising news.

"I don't recognize these forms. Are you sure this is mine?" asked Eric with dread starting to creep into his voice.

"The district adopted a new teacher evaluation protocol. I stated this at the start of the school year when addressing the entire staff. I specifically said that new tool would be used with all staff being evaluated this year," stated in a clipped tone.

Eric countered, "Yeah, but you never shared the new tool with us."

"Every teacher has access to such documents on the district's website. Just because you never looked at them, doesn't mean you're not responsible," Mrs. Sherman stated and turned to leave the room again.

"But why would I have looked at them if I didn't know I was being evaluated this year?" Eric said to her back as she walked out of his classroom and turned back towards the office. Eric sank down onto his chair, wondering what had happened to her to make her change like she did.

As Eric read through his evaluation, he found he had been marked in several categories as unsatisfactory or not implementing even when it was a regular aspect of his teaching practice. Mrs. Sherman had not visited his classroom during instruction this past fall

and it was apparent in her comments. She made repeated references to needing to adhere to district and building expectations that he was already in line with regarding his instruction. As he read, his fear and despair began to grow, but at the end those emotions began to turn to resolve. Eric began making notes on the evaluation and pulling examples from his lesson plans and student work to illustrate his performance more accurately.

As he worked to defend his practice, Eric connected his current dilemma with that of students he had worked with on past projects. He realized how important it was for students to have a clear rubric and expectations when beginning larger projects and writing assignments, and how some of them must felt when receiving remarks and grades when they didn't fully understand what they were to produce to demonstrate their gained understanding and knowledge. If nothing else, this experience had certainly reinforced effective practice with students in that regard.

As Eric wrapped up his final notes regarding his evaluation and instructional practices, he reflected on how Mrs. Sherman had changed in her practice over the past few months. He thought back to his time as a new teacher, and tried to relate it to how Mrs. Sherman might be experiencing her role as the head administrator. Does she have someone to turn to for guidance and support, or is she trying to meet her supervisor's demands without knowing where to turn? It certainly wouldn't excuse her treatment of staff the past few months, but it is an interesting possible insight. Whether that is the case or not, he was meeting with her that afternoon. Closing his portfolio, Eric wondered if this was still the right place for him moving forward.

Three themes illustrate each case: **Classroom Practices**, **Culture of Evaluation**, and **Perceptions of Value**. Each theme consists of two interpretive codes giving definition to that

theme for that case. Across cases, some themes are defined by the same interpretive codes while others may share none or only one interpretive code.

Classroom Practices. For Eric, the combination of *teaching practices* and *student measures* results in the theme of Classroom Practices. These include the teacher's actions of planning and delivering instruction, managing student behaviors, and ensuring a guaranteed and viable curriculum as well as measuring student performance in a manner to inform future lesson planning and communicating to parents (Danielson, 2011).

Student measures include a variety of manners in which to determine what a student knows or how much a student has learned (Goldhaber, 2015). For some, student measures are evidence of student learning. Eric illustrated what and how he might use evidence with his students,

> I have students pick two or three things to put in a paper portfolio every 9 week period so they can reflect back to the beginning of the period and compare what I did in the 3rd nine weeks with what I did in the first nine weeks. I can show my mom or dad at conferences that I've actually done some homework. I tell them to put something in that you're proud of, put something in that your not so proud of, but you were able to come back and show growth. Maybe a pre test and a post test, homework that you just didn't understand this homework, but I did the corrections on the homework after we sat down and went through it again, and I was able to get all of my corrections done correctly.

Continuing to describe how he knows if students have mastered the content,

> I like to get kids up to the board, have them demonstrate. If they can get past the anxiety of doing that in front of all of their peers and whatever other adults are in the classroom and still be successful at solving a math problem, then I think when is just a paper pencil test, then they'll do even better because the anxiety is already gone.

When I asked how he measures student learning, the teacher responded,

> Testing, conversation, guided practice, independent practice, gradual release, how long the gradual release takes. I measured it a lot by asking students to describe what they are doing. If they can't describe what they are doing then they really don't know what they are doing, not making any progress, but if they can go through the whole,

even if they are sitting there not working on their assignment, but they can explain how to do the assignment, that's growth to me.

Noticeable is that the teacher does not focus on straight performance, but rather on students improving from a starting point and demonstrating that improvement in a variety of ways. He also references students being able to recognize that improvement through reflection and being able to communicate it to parents from evidence of their work. Within Eric's case representations of *student measures* occurs more frequently than the counterpart *teaching practices.*

Eric shared that the concept of rigor arises during conversations about his *teaching practices* with his supervisors in relation to the standards, strategies, and activities he employs in class. Rigor, as defined by the district's evaluation training presentation is the combination of complexity and student autonomy, the higher each of those are while in balance with each other the more rigorous the task and learning. During the interview, Eric shared that rigor has not been well defined by supervisors looking for it during observations, and feedback related to it is broad as exampled by this quote from feedback the teacher received following an observation. "The assignment lacked rigor and relevance. The assignment was also not an example of purposeful instruction based on state standards." Student behavior management also falls within teaching practices (Danielson, 2011, Marzano, Frontier, & Livingston, 2011). When sharing about a supervisor's visit to the classroom, the teacher said, "He has been in my room twice. Not necessarily observing anyone in particular, but watching how I handle behavior issues. It was like once in the fall and once in the spring." When talking about what may occur in the classroom if students don't learn the content the first time it is introduced, "find a different way to put that information in front of the student. Maybe it requires a different modality instead of just a

lecture. Maybe I need to have them get up and present the lecture themselves," and "probably going to suggest that you reteach it or find a different way to get that information in front of the kids," were examples of reteaching or revisiting content to strengthen and deepen student knowledge (Marzano & Toth, 2013).

My experience has shown measuring student knowledge does not create gains in student understanding, and without measuring student growth, a teacher cannot know if their efforts lead students to greater gains. Together, measuring student knowledge and using that information to inform and improve instruction, students can attain their learning goals.

Culture of Evaluation. Another theme, Culture of Evaluation, reflects the juxtaposition of those being evaluated and those doing the evaluation and the interactions occurring through the evaluation process. Many teachers fear evaluation with or without the inclusion of student growth measures (Finnegan, 2016; Flores, 2012; Morgado & Sousa 2010; Tuytens & Devos, 2009). While some recognize the potential for improving practice others feel it will not occur within an evaluative framework (Abdo, 2017; Mette, Range, Anderson, Hyidston & Nieuwenhuizen, 2015). For Eric, *supervision* and *leadership provide greater detail to the theme. Supervision* involves the relationship between the supervisor and those overseen (Blasé & Blasé, 1999, Marzano, Frontier, & Livingston, 2011). Eric emphasized the relationships he had with various administrators over the course of his career, and he also recognized how relationships could change related to circumstances.

> When she moved up to become the head principal, her whole perspective changed as she is in a different chair in at a different desk and different expectations being placed in front of her because she is driving the ship instead of following direction from another leader. That change in perspective from her lead to a change in her evaluation process that she used. We weren't changing tools of evaluation, but the way she evaluated changed and my first negative evaluation came through that change in perspective from her. Really disintegrated our relationship that we had built prior to her taking that principalship. Went through a process of fighting the evaluation and

all these negative remarks with education association, really struggled with whether I was where I needed to be based on that poor evaluation and the disintegration of our relationship led me to leave that school district.

Communication can be a key factor to a teacher's success, and that communication must flow both directions (Abdo, 2017). When describing an evaluation conference with an administrator, Eric shared,

> I told him all I'm doing is acknowledging you're just taking a snapshot. You're not telling me I'm doing a good job or bad job, this is just what you're seeing right now, and he affirmed that with me, so I would just sign off on it. There's no reason in my mind to respond with but you didn't see all this. He already knows that, and so it was more of a friendly conversation than a back and forth volleying of ammunition. I appreciated that he was able to just have a conversation with me instead of seeing it as some way to confront each other.

Not all communication was described as congenial. Reflecting on another experience, Eric said, "I had no idea what was on the assessment piece until I sat down to read through it after he had already done the assessment." Eric described a contrasting example when he felt more support. "He went and redid my evaluation and brought it back before he even gave it to the school board. And I appreciated that." Supervisors may often be tasked or expected to provide leadership, but it does not have to be limited to a supervisor.

Leadership can occur in a variety of forms from a variety of people. Teachers should receive effective instructional leadership from peers or supervisors to continue to improve their practice (Abdo, 2017; Fairman & Mette, 2017; Mette, Range, Anderson, Hvidston, Nieuwenhuizen & Doty, 2017). Eric described many people visiting his classroom, but he was unclear of their purposes at times and highlighted a point of frustration with leadership. When describing an instructional coach, he shared, "I know she's doing something on the computer, but I don't know who her documentation is for. Is it personal? Is it for somebody above her? I have no idea, but I never see any of that from her." Another time, he described

an assistant principal, "It was always fairly positive. You're doing a good job. This is what we're seeing, little room for improvement here. If you changed this, it might help." When discussing the walkthrough protocol his building leaders implemented, he shared a pragmatic outlook,

> I thought it went fairly smoothly. I know there are other people that like to bicker about everything that comes across that walkthrough piece, but you got to let some things go. It's not a criticism every time someone's in your classroom. It's an opportunity to grow on. I always tried to find something the teacher was doing well, something they could work on, and put both of them in there. Some teachers see every single thing as a negative and I'm going to argue with you about every single sentence that you put on the walkthrough. That's not what it's for. Use it to help yourself grow. That's how I approached it.

Overall, it appears Eric desires feedback and constructive criticism, and his frustration arises when it is unclear or not provided.

Eric described growth as stemming from feedback. "Well, like with my student portfolios, I would hope that there would be things that showed areas that I could improve in, also some areas that I was proud of, positives and some areas for growth to happen." Eric drew many similarities between his students' learning and him improving his practice, "Documentation. You can't show growth if you don't know where you started. Have to have a beginning point and an end point. Then we can compare the two to see if there is any growth." Also closely related in Eric's point of view is improvement, as illustrated in his description of performance evaluation. "It is supposed to be an improvement process. You have to be able take criticism. Nobody's perfect. It doesn't matter if you are on the side of the person being evaluated or the person doing the evaluating. Nobody's perfect." When asked about the purpose of evaluation, Eric said,

> Improvement. Improving the teacher's instructional style, maybe adding to the teacher's instructional style to meet more students at a different level. It's all about improvement. It's not necessarily, these are all the things your doing wrong, but if I

saw this I could say that you're maybe doing things a little bit better, looking for this, this, this, and I'm not seeing any of those today, may not have seen it while I was in the classroom, but did you do those things throughout the day. Just, pushes me to be better at what I do.

Eric consistently referenced believing all teachers could always continue to improve their practice and performance in the classroom.

Eric's view of the Culture of Evaluation rested heavily on interaction with building administrator's and other school leaders. As will be seen through the discussion of the other participants, their view of this theme carried more balance between leadership and peer interactions. Eric, alone, lent himself to the interpretive code *supervision*. In my reflection, I wonder if it might have been a result of my interviewing, past experiences that were not shared, status of relationships with peers, or might it be reflective of his core belief? Similarly, the interpretive codes present in Eric's third theme are also unique to him.

Perceptions of Value. Perceptions of Value holds how teachers see their worth within the context of their work and school. Teachers' decisions to stay or leave are greatly influenced by their relationships with administrators and by administrator's ability to manage the day-to-day working conditions (Boyd, Grossman, Ing, Lankford, Loeb &Wyckoff, 2011). Teachers often disagree with administrators regarding the accuracy of ratings assigned to performance (Finnegan, 2016). Additionally, effective tools implemented in an accurate manner consume enormous amounts of time and resources (Riordan, Lacireno-Paquet, Shakman, Bocala & Chang, 2015). Eric's Perceptions of Value consists of *development* and *challenges*. For teachers to improve, principals need to provide constructive, collaborative feedback and professional development (Matthews, 2017; Shugart, 2017). *Development* provides insight into Eric's experiences to improve his practice. Much of Eric's viewpoint reflected an absence of training. While he and from his point of view, the principal, both

recognized the need, its absence permeated Eric's perspective. "We didn't, I mean we had walkthrough protocols but I was never trained on that." When discussing a conversation with an administrator, he shared that at times he would be evaluated on aspects of his job that he did not know they were looking at because he had not been trained on the evaluation tool to be used to measure his performance. Describing a conversation with a supervisor, he stated,

> I said, I'm doing this, I'm doing this, I'm doing this, you're marking that I'm not. Here's my marker board of things that I need to put in place, things I already have in place. After I looked at your assessment, here are the things that I didn't realize that I was being assessed on, but now they're on my to do list.

Eric also shared his frustration with the training he does receive, stating, "so, I'm supposed to go get training this summer. But, just because you've been trained does not create a high qualification. So that's where my questions come from." So, while, the interpretive code is *development*, it becomes quite clear that Eric's perception includes the belief and frustration that there is a lack of development. This has certainly been a challenge for Eric, along with many other frustrations as demonstrated in the next interpretive code.

Circumstances impeding growth and improvement make up the interpretive code *challenges*. One such challenge is uncertainty, and Eric shared many areas and questions that led to having uncertainty, for example in reference to decisions made by administration regarding staff assignments for the coming year, Eric said,

> I have seen people move from a highly qualified position to an area that is brand new that has nothing to do with their highly qualified area. That's not my call, but I sometimes wonder what the justification is for that, and because I'm on the teacher side of that, I don't expect I'll ever hear. But, there are questions about why. Not just from me, but from other people around me. And I don't have an answer, so I don't create rumors. That's something that bothers me. Why would you move someone that is a highly qualified person out of their curriculum area into an area they have no experience in?

At the same time, Eric voiced frustration with not knowing what he would be teaching next year while feeling quite confident many of his colleagues already had their assignments. "It's on the schedule. I don't know that it's set in stone, but I'm pretty sure I'm not teaching this next year.

Conflict poses *challenges* as well. Eric presented himself as a team player and accepting of the school leader's direction and choices in many realms, but at the same time, he still felt frustration with supervision at times.

> I did have something to add to that. I'm not only being evaluated by my principal but also by the assistant and my instructional coach. The evaluation was handled by the principal, but I had walkthroughs or sit-throughs by all three and some days two of them came in at the same time or right behind each other and had a completely different viewpoint of what was taking place in the classroom, from each other. Assistant principal came in, did a walk through, sat for a few minutes, made his comments, sent the email, got up, walked out and the principal followed him in. They passed at the door, and the principal came in, went through his normal protocol asking a student what are you doing, what are you supposed to be doing, what are you supposed to be learning today and the difference between the two walk through end pieces that I saw in my email was like night and day. In a matter of five minutes it was like a completely different experience for both parties. Again, I'm not going to use that as ammunition. If that's what you saw, then that's what you saw. I can't fight that, there's no reason for me to fight that. It's a snapshot.

As illustrated by Eric's experience, he found himself in the middle of opposing feedback and direction. Additionally, there were the interactions with peers regarding duties and assignments as well. "I'm hearing that she doesn't necessarily get along with everybody and that may be part of the decision. I don't know. I've never really had any issues with her, but I've only been here for a year, and there may be some past history that I am unaware of." When large groups of people are together some conflict will occur, but coupled with uncertainty as a recurring focal point, and it becomes a challenge. Eric brings a point of view of multiple years of experience in education in multiple settings while having only been in his current setting for a short amount of time. His **Perceptions of Value** include *development*

even while does not feel he fully received it, and his frustrating interactions with supervisors and peers highlighted *challenges* he perceives in his work setting. Paul on the other hand has only been at Hawkins Middle School since completing his education licensure.

Case 2 - Paul

"Uhg, this is such a garbage waste of time," said Paul in frustration under his breath so that no one else could hear him. He and his fellow teachers participated in a review of the evaluation process including which years of service were on cycle for evaluation. Paul just realized his evaluation would be due. Paul could already predict what would occur. He would be told what goals to work towards. He could pick extras if he wished, but he would still be held accountable for the same goals as everybody else regardless of how well he performed in those areas. The early observations would be full of low marks with later observations full of high marks so that he and the administration could show teacher growth. Last time, Paul did not even think he received his own evaluation. The way it was worded seemed off in relation to what was included in his observations. The whole system just seemed broken and unproductive to Paul, a waste of time.

Over the next few weeks, the assistant principal, Mr. Clark, held a goal setting meeting with Paul, followed by some initial observations. Mr. Clark scheduled Paul to visit his office after school following a formal observation.

"Paul, thanks for coming by my office. I'm incredibly busy trying to follow up with everyone," greeted Mr. Clark as Paul knocked at the small office's door.

"No problem. Are you ready for me? I have some students waiting for me to review some work they were struggling with," replied Paul stepping into the small office.

"Yes, come in. As you are aware, our evaluation process requires pre-conferencing and post-conferencing for formal observations. I know we didn't get a chance to discuss the observation before I dropped in today. You don't mind signing off on our pre-conference do you as we go over the observation?" said Mr. Clark.

Paul responded, " I guess that's okay." His voice rose slightly at the end, almost as if asking a question himself.

"That's great. I appreciate it. I've just been swamped getting to everybody on my list. You know what I mean?" said Mr. Clark. "So, I've been through your room a few times informally, and you're always doing a great job. We really appreciate your work with the students. They behave great. You never send students to the office. Parents never have a problem with anything going on in your class. Overall, doing a great job. As you know, we're required to show growth over the course of the year, so even though you're doing a great job, these marks I'm about to share with you may seem a little low. That's okay. No problem. By scoring you low now, it makes it easier to show growth by scoring you higher later. That make sense?"

"Sure," said Paul.

"Great!" said Mr. Clark. "So, as you look over this paper, you'll see the goals that everyone is doing. If you selected any other goals, they will be listed after the building goals. I know most of the marks are at the starting level or just above it. That's what I was talking about before, so that we can show you grew over the course of the school year. After you read it over, sign it at the bottom."

"Okay," said Paul, signing the paper without really reading any of it. It looked the exact same as last time, he thought.

"That's great. Thanks again for your time," said Mr. Clark as he collected the signed observation form. "I'll put a copy in your school mailbox for you. You should have it some time this week. You have any questions for me?"

"Nope," said Paul as he stood to leave. He couldn't tell if he was more disappointed in the broken system or angry with himself for going along with it, such a waste of time.

Five months later, Mr. Clark visits Paul's classroom again, and asks him to come by his office at the end of the day.

"Paul, thanks for coming by on such short notice. I really appreciate it," greets Mr. Clark as Paul enters the small space. "Can you believe the year is more than half over and spring break is just around the corner? It has been a quick school year."

"Mm-hmm," mumbled Paul as he took a seat in the chair across the desk from Mr. Clark.

"Paul, I asked you to come by today so that we could finish up your evaluation. If you recall, the evaluation process calls for three formal observations with pre-conferences and post-conferences. We finished one of those at the start of the school year. I came by your classroom two different times today, and once we discuss those, we can wrap up this year's evaluation cycle. I know it's not ideal to do them both on the same day, but time just keeps slipping away from us doesn't it?" said Mr. Clark as he passed Paul a small stack of papers.

Paul held a neutral face to the best of his ability. He wanted to scream. He wanted to yell that this isn't right, that it's not the way to improve teaching or help teachers. He wanted to throw the papers on the desk and storm out. Paul mumbled agreement while keeping the frustration and disappointment from showing in his body language and voice.

"So, just like before, you're doing a great job. Your students stay in your class. The parents aren't complaining about anything. Every time someone visits your classroom, all of our expectations are met regarding using the building wide strategies and being on track with the curriculum. You're doing a great job. You'll see all of your marks went up, some satisfactory, some excellent. We're really glad to have you as part of our team. Any questions?" asked Mr. Clark as Paul shuffled through the papers.

Paul was not really looking at the papers as he moved through them. He was revisiting a conversation he began with himself last year, wondering if this was the right place for him. He loved teaching and working with the students, but this part of the work just did not feel right. Paul signed off on the pages marked with spaces for his signature, and he began to stand to leave.

"Paul, thanks again for your time. If you don't mind, could you swing by after student dismissal tomorrow, and I will have the formal written evaluation prepared for you to sign. Thanks. I appreciate it," said Mr. Clark as Paul turned to leave.

"Sure, no problem," said Paul becoming more concrete in his thinking that a new setting might be a good idea.

Classroom Practices. Paul's Classroom Practices focuses on *teaching practices* and *student measures* just as in Eric's case. Whereas Eric's focus for *teacher practices* narrowed to rigor and student behavior management, Paul emphasizes curriculum first and foremost. When sharing an experience about planning lessons for his students, Paul explained:

> When we get together to plan ELA [English language arts] along with our teacher leader, we spend most of our time creating what we will do for that week, and my complaint has been, why doesn't the district have a curriculum for us so that we can spend our time focusing on instruction and how we will give that curriculum and interactions with our students rather than spending all of our time creating that?

Shifting to instruction, Paul continued sharing about how he does not connect the evaluation process to his daily practice. "I don't think about the evaluation process when I am planning my lessons or units, or even typically how I will instruct." Paul described how he believed supervisors should discuss with teachers how to improve their practice as part of the evaluation process,

> They should ask the teacher what they saw during that lesson, how they reacted to different things that happened. Then they should ask the teacher what was one thing they thought went wrong and how they fix that for next time or in the future. What was one thing you think went well, and how will you build upon that?

Teaching practices captures Paul's lens that curriculum and associated resources should be readily provided by a school district, and the teacher's primary focus is on to effectively delivering curriculum through instruction with the supervisor weighing in to offer support and guidance on improvement of practice. At the same time, Paul's words highlight that he does not feel this typically occurs in his current setting.

Student measures, including assessment scores, provide teachers with a marker of student knowledge. When asked to describe assessments used as part of his instruction, Paul mentioned pre and post assessments repeatedly and the use of writing samples. "That's by far the easiest one, one of the smallest lengths of time that you can get though. Then, anything that is objective data, so, multiple choice is easy to use. A lot of writing is subjective. Even grading of writing can be subjective as well, even if you're using rubrics." Paul continued to describe the need to measure beyond finite measures of time and expand an examination of student growth to larger time fields.

> So, I mean, writing samples, not necessarily even just, beginning of the year to end of the year, but if you could get school wide, 6th grade writing samples and then, now their 8th graders, show them how far they have come. Trying to show it on more of a macro level than a micro level because at middle school, it's hard to see at the micro level growth at these grade levels.

Paul also acknowledged that assessment should serve a larger purpose "And we're talking like, not in the moment, but we've collected all this data and we have to go somewhere after this."

Paul acknowledged the use of data at a variety levels including at the state and district leadership levels.

> I am also part of the building leadership team, which I am just now remembering, and we have been tasked by state accreditors, thinking about how we can improve the school based around our district goal, which is improving test scores in ELA and Math by 10%.

Paul also uses data with his students, "data does help, having something tangible to show them growth does really help with this grade level." As Paul connects the use of student data with performance evaluation, he questions his supervisor's ability to do so effectively.

> It should, if we're going to take student work and we're going to use it as part of my evaluation, that would mean admin is looking at my student's work and they are evaluating me based on that. Which, in theory, would give admin a better idea of what is going on in school, a better idea of what is being taught, a better idea of what students are able to do and not able to do, which I don't think they really understand.

Paul's *student measures* captures the use of assessment and data to measure and guide improvement of a teachers' practices and of a school's implementation when used knowledgeably.

Culture of Evaluation. In contrast with Eric, Paul's second theme includes *peer interaction* as integral part of the **Culture of Evaluation**. Through peer observation, collaboration and feedback Teachers feel more confident in their work and student performance (Abdo, 2017; Klingelhutz, 2017; Mette, Range, Anderson, Hyidston & Nieuwenhuizen, 2015). Paul seeks knowledgeable peers to improve his practice.

> No. I think if I'm changing practice at all, if I'm looking at changing practice, I do that more during PLC work with my teacher leader that doesn't do evals, and with my

colleagues. Evaluation just seems like, it's something that they have to do, and they do it. I also don't view him, I don't view admin as teacher leaders, or I just couldn't imagine him telling me something about ELA and how to do it better. I don't see that coming from him. I can see that coming from my teacher leader because she has a background in ELA, and she is more knowledgeable about it. I just see his role differently than.

Paul mentions working with his PLC, professional learning community, a group of peers that collaborate how to address student needs and instructional supports, and he mentions working with his teacher leader, a staff member whose primary purposes is to coach up teachers in their instructional performance. Paul consistently views administrators as outside of improving his instructional practice.

I don't know when I'll be evaluated again. I don't see anything coming up again, at least coming up from admin that would be doing those evaluations. Now, my teacher leader, who does not do evaluations, will be more of a coach throughout that time, and we will have those conversations. But I don't see my evaluator, being admin, really having any say in what I do until I am evaluated again.

Since Paul values the guidance and opinions of his fellow teachers over his administrators, it is not surprising that he relies on peer support.

When discussing improving his practice, Paul refers to working with his PLC and teacher leader consistently, and he often counter-points a lack of faith in administrators' familiarity with the curriculum and content specific practices to push his practice up. This is highlighted by a small portion of a quote referenced earlier, "I think if I'm changing practice at all, if I'm looking at changing practice, I do that more during PLC work with my teacher leader that doesn't do evals, and with my colleagues." When receiving tasks and directives from supervisors, Paul also references connecting that work with peers. "I know he's [principal] pushing for that pre and post assessment data meeting and we can look at that during PLCs and gauge that." At the same time, Paul acknowledges that not all peer

interactions are supportive. When discussing how students are assigned to classes and class sizes are balanced, Paul identifies that some peers are not so supportive.

> I think that number of kids in the room falls more to counselors than admin, but I also do know of certain teachers that will do their best behind the scenes to move certain kids out of their rooms when they know that they don't want them, even at the expense of other teachers. And not letting them know that they are doing that.

Overall, Paul relies heavily on his peers as more knowledgeable supports and coaches to improve his practice while realizing that not all peers are acting altruistically.

The second interpretive code, *leadership*, brings the attention back from peers to those tasked with leading the school. One of the phenomena present in Paul's case are the expectations held by administrators. Paul described his school principal as, "I would say that in general, my principal is very hands off, not really wanting confrontation, like letting other people do their jobs." At the same time, Paul acknowledges that this viewpoint has changed over time, and he did not originally feel this way. "I would say that he probably treats new teachers differently than veteran teachers." Paul believed his own experiences early on were more demanding than of recent years. "I would say that he was probably, pretty hard on me that first year, as far as like, observing, more feedback, trying different things, more critical in ways." Paul believes his administrator sees him differently now. "He [the principal] would say that I have gotten better, but I have also seen him be more critical to newer teachers as well. When putting it in the context of performance evaluation, Paul shared, "In theory if you are doing everything correctly with the evaluation process, you should be meeting expectations." At one point, Paul described his role in the evaluation process,

> My role in the evaluation process, if someone is evaluating me and they give me some feedback or criticism or whatever I need to adjust what I am doing based on that. So my role is to do the best I can, but to adjust when I need to based on what they say or what I get from the evaluations.

At the same time, Paul recognizes a shift in relationship with the school principal.

> From year one to this year, even from year one to year two, it was completely different between him and I. Year one, I didn't really feel like I was expected much as a teacher, and then something switched, and I'm not sure what it is. But we've been fine ever since.

Paul presents two views of leadership in his school. The first is micromanagement, as the school principal directs teachers on specific tasks. This is demonstrated when describing the goal setting for the evaluation process.

> They could be anything, there's hundreds in the evaluation tool, whatever you call that, so we're supposed to pick targets, however our principal has the targets already pre-picked for what he is looking for from everybody. So, if he chooses three targets for evaluation, you could choose three more, in theory, be evaluated on 6 things, or you can just pick the three he has chosen and just leave it at that.

At the same time, Paul shares that the principal delegates tasks and the communication of expectations to others. Paul describes grade level team leads must run team meetings and communicate the administrator's expectations and agendas.

Paul paints a complex picture of *leadership*. At times the principal is very directive, and yet Paul feels the principal is also very hands off allowing staff to do their assigned duties. Paul also describes his relationship with the principal changing, but had previously noted he felt the principal was more demanding of first and second year teachers. Might Paul's experience have been the change that Paul noted in their relationship?

Paul's Culture of Evaluation combines the *peer interactions* and *leadership* to portray a setting where the principal establishes expectations of performance or result, but then creates space for the teachers and other staff to achieve that result through collaborative efforts. As a result, Paul identified more closely with those he is embedded in the work with and holds the belief they would be better able to evaluate his performance rather than an

administrator that Paul views as standing outside the actual work. Recognizing this point of view allows for deeper understanding of the third theme.

Perceptions of Value. In another deviation from Eric's case, Paul's Perceptions of Value held two different interpretive codes, the first of which is *power*. Newer federal and state directives regarding evaluation, achievement testing, and subject emphasis cause teachers to feel they have no control or authority in their work (Vardas, 2014). The autonomy of choice could provide teachers with a sense of power, especially newer ones to the profession, who often receive the most challenging students, likeliest to have traveling classrooms, and receive the most difficult assignments resulting in higher rates of attrition (Bieler, Holmes, & Wolfe, 2017; Harmsen, Helms-Lorenz, Maulana, & van Veen, 2018). Teachers and principals often misunderstand each other's messages (Abdo, 2017). Paul mentioned previously, he believes new and veteran teachers are viewed differently with different expectations, and this carries into his views of their power as well. "Well, in my experience, I think that, new teachers typically get the worst assignments. I think that returning teachers do get a say in where they want to go, what grade they want to teach, subjects, things like that." Administrators play a large role in affecting teacher retention by establishing working conditions and relationships with staff members (Boyd, Grossman, Ing, Lankford, Loeb &Wyckoff, 2011; Hardin-Bartley, 2014). In other school areas as well, teachers are often tasked with taking on additional roles of a variety of natures including but not limited to coaching athletics, sponsoring clubs or activities, or participating on staff committees. Paul has been approached to participate in a variety of additional duties beyond his classroom. "I don't know what would have happened if I would have said no, but I could've said no. I didn't feel any pressure to say yes." In this instance, Paul felt he

174

maintained the power and could have said no, but would he have if he was still a newer

teacher operating under the perception of different expectations? This leads to examination of

control.

Teacher control was touched on earlier when Paul described teachers, or more

accurately administrators, selecting goals for evaluations. "I do know new teachers have at

least one of their goals chosen for them." Paul went on to share about new teacher goal

setting, "For new teachers, he makes, he forces them to choose certain goals. These are the

ones you will do for new teachers. Then, after a couple of years, or maybe after one year, you

can choose your own, but they have to be in a certain domain." Paul highlights the shift

presented earlier regarding new teachers versus veteran teachers' control. Additionally, at the

end of the year, when assignments for the next school year are assigned, Paul believes this is

another area where control resides solely with the administrator,

> Someone that's teaching math right now, but getting moved, possibly to teaching
> elective courses, and rumors could be that's because their evaluations are not very
> good, and they don't want them in a tested subject for the state. Obviously math and
> ELA is what is taking precedence over everything right now.

When specifically asked if he could influence policies or practices at his school, Paul shared

that he believed he had earned the respect and opportunity to do so.

> So, I do feel like I can influence some things in my position, where I am at now,
> which is good. I feel teachers should be able to influence practices that happen in the
> entire building. I shouldn't have to be an admin in order to influence those things.

For Paul, *power* resides with and is assigned by the school principal. In his current setting,

Paul represents this as starting with no power as a newer teacher in the building, but with

time being granted opportunities to have more choice and options regarding assignments and

additional duties. At the same time, Paul recognized that poor performance could result in

teachers being assigned to teach elective classes versus the core content classes if student

performance or performance evaluation scores do not meet administrators' expectations.

Appraisal provides clarity to Paul's **Perceptions of Value** along with *power*. Hardin-

Bartley (2014) demonstrated human resources procedures and policies devalue people and

create distance between the acts being evaluated and the results that occur. With this in mind,

evaluating teacher performance separates teachers from those evaluating and tips the scales

towards judgment rather than collaboration. When previously discussing assessment and

data, Paul shared he did not believe administrators held the ability to accurately evaluate the

work of teachers due to a lack of curriculum and content knowledge. This was pointed out

specifically when talking about administrators looking at student work to evaluate teacher

performance,

> Which, in theory, would give admin a better idea of what is going on in school, a
> better idea of what is being taught, a better idea of what students are able to do and
> not able to do, which I don't think they really understand.

Paul also expressed his lack of belief in the observation components of the evaluation

process.

> I mean right now the evaluations feel like, here's your time, here's your dates, this is
> when you're being evaluated. So, it's like you're getting ready for this one big
> moment and that doesn't seem very productive to me. If it was more of a regular
> thing, seems like it would be more productive, because now you're not just looking
> for that one time that this person shows that they did all these things.

Overall, Paul explained the evaluation process as a flawed tool for generating judgment of

teacher performance or attempting to improve it.

> I don't really feel like there was a lot of coaching going on. Even if there were
> criticism, there would have been typed comments within the evaluation tool. Not a lot
> of face-to-face meeting, talking about a lot of things.

Paul pointed out that even if feedback were happening, it would be provided through

electronic means rather than through coaching, professional development, or training. This

idea has been touched upon earlier when highlighting the differences Paul perceived between

how newer and experienced teachers were treated with choice and control. This occurs

between different administrators as well.

> I was, some years, evaluated by a different person. Some years it's a vice principal, some it's a head principal. It's usually one of those two. One year, I was being evaluated by the vice principal, so he was the one doing all of my evaluations, even the informals, the head principal wasn't coming in anymore, really. It was the vice principal's job now to get a feel for what was going on in my classroom. Let's be real, the only way they can do these formal evals is by trying to get a guage of what is going on throughout the year and they do that through informal evals.

Paul emphasized informal observations and interactions influence the final result of a

teacher's evaluation more than formal observations.

Also, Paul described administrators as inaccurate regarding evaluation practices. In

addition to believing they used the informal observations more than the formal observations,

Paul shared experiences where he believed he received other people's evaluations.

> I mean throughout, not just the last one, but first one, you've got to start low for there to be any kind of gain, and then the evaluation was written and it was like He is a great teacher, blah, blah, blah. 'She' is great at classroom management. 'Her' classroom is run very well. You know, whatever it said. But the pronouns were overall female pronouns, which, was concerning because I never really asked but I'm guessing it was copy and pasted from somebody else's evaluation. Why else would all the genders be changed?

Appraising teachers accurately requires attention and effort not illustrated by Paul's

experiences.

Paul's **Perceptions of Value** as demonstrated by *power* and *appraisal* describe a

setting where experienced staff members have gained the comfort and freedom to make more

decisions for themselves while newer teachers receive more direction and demand. His

experiences also highlight the ineffectual use of evaluation to improve his practice, and he relies more heavily on peers and teacher leaders to help him improve. While Paul has only worked at Hawkins Middle School and only for a few years, another study participant, Ray, has been at Hawkins for more than 30 years.

Case 3 – Ray

"Ray? Hi, I'm Devin. The principal directed me to introduce myself to you. He said that he spoke with you about being my mentor teacher since this is my first year teaching. Do you have a few minutes?" asked Devin nervously. Devin has been nervous for months, finishing college and starting as a new teacher, in a new city, a school full of unfamiliar faces. Everyone has been really supportive and welcoming, but Devin's confidence remained low as he contemplated the challenges he would be facing.

"Hi Devin. Yeah, he spoke with me. Glad to help out any way that I can. What can I do for you?" asked Ray. He looked the new teacher over. When the principal approached him to be a mentor, he was not excited about it, but he did not have a problem doing it either. He had worked with lots of new teachers over the length of his career. The principal shared that Devin interviewed well and shared a lot of ideas for incorporating technology into lessons. That was all well and good, but Ray knew there was more to teaching than technology.

"We are supposed to set up weekly check-in meetings, and I am to share the monthly mentor-mentee meeting calendar with you that they gave us during orientation. Do you have a Google calendar or outlook I can share these to?" asked Devin pulling out his phone.

"How about we just pick a day and time to meet each week. We won't need to share a calendar for that. You can just email me the dates and times of the other meetings too. That way I'll have them. Do you have everything you need for your classroom?" Ray responded.

He used the calendar on his phone some, but he tried not to be too reliant on his phone to run his life. Devin shared that he thought he had everything they needed regarding classroom resources and materials. They talked for almost an hour about the basic operations of the school, where different things were located, routines and procedures that most teachers use or the students would be familiar.

As Devin and Ray wrapped up their conversation, Devin began to leave. He paused a few steps from the door and hesitated. "Forget something?" asked Ray. Devin turned halfway around looking more nervous than when we first came into Ray's classroom.

"Uhm, can I ask you one more thing? How do you… how do you know if it's working? How do you know if you're getting through to the kids rather than just wasting their time? How do I know if I'm a good teacher?" The questions just spill out of Devin in a rapid ramble. He had not meant to ask them. He was not sure he wanted the answers. Underneath his nervousness, these questions highlight Devin's true fears.

"All good teachers ask themselves those questions regularly. So, you're already on the right track. The thing to keep in mind is, short of hitting a student, there is almost nothing you can do in one day that you can't fix the next day. You're not going to ruin their lives with a bad lesson. Each day, you look back over what happened, and make adjustments for the next day. That's all that any of us can do. As long as you keep asking yourself that, you're on the right track. Don't worry. You'll be fine," Ray shared watching Devin begin to relax a little, nodding his head.

"Oh, okay, thanks. Thanks again for your time this afternoon. I'll check in with you on Thursday like we set up. Thank you," said Devin as he turned back to the door. He was obviously still a bit nervous, but he did not look quite as consumed by it as before.

With Devin leaving his classroom, Ray thought back to when he was first starting out as a teacher all those years ago. His first few years of teaching, he was not necessarily scared but was concerned he was going to be evaluated in a poor light. Then administrators and other teachers would come in and tell him how great he was doing. Ray did something better than they had ever seen it done before. So, then he got to the point of thinking, "well, I must be a pretty good teacher. I'm not the best teacher ever, but I must be pretty good." He had lots consultants come in when the school would adopt new initiatives. One of them was from Florida. He actually took a class with her one weekend for continuing education credits for license renewal. Ray remembered talking to her at the weekend workshop about her visits to his classroom the previous week. She had been really impressed with the students' use of the strategies so quickly. She invited Ray to join her in the consulting field during summer sessions. He ended up turning her down due to other commitments during the summer, but he could look back and see a bit of confidence boost in himself when that happened. He would have to make a point to find those opportunities for Devin, little things, early on, to help smooth out the obvious nervousness.

Ray was sure Devin was as nervous as or more so than when he first started, and he also knew that visits from the principal and assistant principal, at least initially, would only increase that problem even if the feedback were positive. Based on his own experiences, administrators were more concerned with completing their required paperwork for observations and evaluations than comforting a new teacher. It was not that the principal's did not care; they just had their attention focused elsewhere. That is why Ray agreed to be a mentor when asked. He remembered what that experience was like all those years ago, and he is glad to help others through it.

Classroom Practices. While Ray's classroom practices could have been defined by the same interpretive codes used by many other participants such as Paul's, Ray's phenomena were more specific resulting in narrower interpretive codes. Rather than looking at a range of teaching practices, Ray's actions zeroed in on *strategies*. Strategies include those techniques employed by teachers to engage students in instruction.

> What was impactful for me was when the superintendent, at least maybe, 20 years ago, he brought in a new initiative, and they showed us a bunch of teaching strategies. A lot of them were new when I was a student, but they had just kind of fell out of favor.

Ray repeatedly emphasized many classroom management strategies he would employ to maintain order and discipline when working with students.

> Yeah, things like proximity. A lot of teachers stand at the front of the room or sit at their desk. If you are moving up and down the room, that kind of keeps kids on their toes. What really keeps them on their toes is to stand behind them. They don't know exactly where you are. Something else I would do, taking kids to lunch. I guarantee you that every teacher in the building with the exception of maybe a couple of us, walk in front of the line. I walked at the back of the line. The kids couldn't see you.

Ray identified learning strategies he would employ as well. As he and his colleagues continued to learn more strategies and emphasize their use during instruction, Ray would try to create balance for his students.

> Just, there were tons of strategies that we learned. Cooperative learning, all the different things, but it was more than cooperative learning. That was one of them. That was the one that most administrators latched onto. Then they expected cooperative learning 24/7. Just like everything else, you can ruin it by overdoing it. I know one other thing we were talking about, when we first started using computers to do assessments, the kids loved them. Our test scores went way up, and they said it was because of technology. After about the fourth or fifth year of using computers, the test scores were almost back to where they had started. So, it was just because the technology was new and the kids were excited about it. I've had a couple administrators, probably more than a couple wanting everybody to use this new strategy, ask me why I don't want to use the new strategy. I said, if we're all doing it, the kids are going to get tired of it. They go to every class and that's all they do. So, I'm going to do something else in my classroom. It will be something different for them. I don't know that it made sense to any of them. It made sense to me.

Ray continued to describe the strategies he would use to engage students during his

instruction.

> A lot of questioning, I had kids, and this was a cooperative learning thing, but I had
> seen it done before we had even heard of cooperative learning. I used inside/outside
> circles. I used buddies. I just called it, told the kids, get with your buddies, study
> buddies is what I called it. Then it was, not peer tutoring, but some cooperative
> learning name, kind of like a peer coach, like cooperative learning. You made sure
> you had one really high kid, what you perceived was a high kid, and maybe a low kid,
> a couple in the middle, male/female mix, racial mix, kind of impossible to do
> completely. You can kind of do it. I would get the kids to help other kids.

Ray's emphasis on classroom strategies solidified his viewpoint of teaching practices.

Similarly, he held narrow focus regarding measuring student performance.

Assessment, for Ray, acted as the focal point for discussing measuring student

knowledge. Ray shared that as his experience grew and he knew more about the variables

influencing students' lives, he began to adjust and make changes to how he measured

students' knowledge and understanding.

> When I first taught, I 100% based it on their unit test. Then, over the years, I got to
> thinking, that's not a fair way to do it. The kids might be having a bad day. Whenever
> they took the test, you would have a kid that would get 80s and 90s and then on one
> test get a 50. It wasn't because they didn't study for that test or didn't know the
> information, probably not. If they got 80s and 90s on the first five, they are probably
> going to do the same. They were probably having a bad day. You know as much as I
> do, maybe more, about all the different family situations, they could be going
> through. They got countless stories kids have told me about their family life. So, and
> it wasn't the way I did it later, probably wasn't a good way to understand their
> knowledge, but if they were putting in the effort, most kids are going to get it if
> they're trying. And there are some kids, no matter what you do, they're not going to
> get it, whether they don't want to or they just don't have the ability or whatever you
> want to call it.

Ray continued to expand on his understanding and role regarding assessments, including

standardized assessments such as those used for state assessments.

> Standardized testing has a purpose. But the problem with standardized testing, in my
> opinion, is that it doesn't show you what they know. It shows you what they don't
> know. Because, in history for example, they might know 500 things about the Civil

War, but the assessment only asks five questions about it, and they only get two out of those five right, and it looks like they don't know much about the Civil War. Know what I'm saying? It doesn't show what they know. It shows what they don't know.

Ray's **Classroom Practices** adopted the goal of finding equilibrium for his students learning experiences. As teachers incorporate, Ray utilized them as the other teachers do, but as he felt the students become flooded with their use, he backed off and found alternatives for them so as not to overload the students. Similarly, with assessments and measuring student learning, Ray recognized an imbalance regarding scores reflecting student learning and then sought ways to create balance for the students. While Ray attempted to create balance in the culture of his classroom and practice, that is not how he described evaluation experiences.

      Culture of Evaluation. Beginning with Paul and continuing through Ray and all of the remaining participants, the interpretive codes providing clarity to each one's Culture of Evaluation involves *peer interaction* and *leadership*. Ray's tenure at Hawkins Middle School incorporates more than 30 years establishing himself as a school leader through longevity, so, understandably, he established relationships and *peer interactions* with many different colleagues as well as administrators over the years. Some of those relationships represented trust. "I don't know what expectations he had of me. He knew me from when he taught there before. He knew I was going to show up and do what I'm supposed to do. I guess that's what he expected." While others, it was a matter of reliability. "It was kind of like the squeaky wheel gets the grease. I didn't ask for much, but when asked for something, I pretty much got it. The things I asked for were things for sports. Make sure I had updated equipment, stuff like that." Ray recognized relationships with some colleagues were much more contentious and difficult.

      I've seen administrators push people out of a building just because they didn't like them, had nothing to do with anything but their personality. The admin didn't like

them, then they would figure out a way to get rid of them, multiple different ways on that.

Ray described administrators took action against peers, sometimes upsetting him when he felt it was inappropriate regarding their performance or duty assignments.

> One of them I'm thinking of, he didn't care for this teacher. He really didn't fire her, he found another principal that agreed to take her. So, he kind of pushed her to another building. From what I know, she was a really good teacher. It was a personality thing. That was my first principal that did that. Then, another principal wanted to get rid of somebody. This was horrible, but this guy shouldn't have been hired. She got kids to make up stories about him. So, instead of going through all that, he resigned. He went to another district.

Bringing it back to Ray's personal experiences, even in the role of a school leader as a more tenured staff member, he shared that his relationship with every administrator has been different.

> I learned fairly quickly, the only thing I can influence is what I do in my classroom with my kids. And I always wanted to change things with administrators or whatever, and some of them I've felt like I could go talk to them and they would listen to me and some would actually take my advice or my suggestions. Some of them wouldn't or didn't, but I knew, the only thing I could really control was my classroom. And I'm not talking about discipline either. I'm talking about instruction and strategies and things like that.

With each administrator, Ray's relationships with previous administrators influenced his relationships with the new ones.

Another aspect of peer interaction involved responsibility. Ray described responsibilities as following through on personal and professional commitments. "I think it's just a personal thing if you have responsibilities or if you have pride in your job, you're going to do it whether they're looking over your shoulder or not." He recognized that not all of his fellow teachers and colleagues maintain the same level of work ethic that he described in himself. "Some people, they have to push them and stay on top of them all the time." Ray defined himself as taking pride in his work and his students' accomplishments regardless of

recognition from others. "For me, it didn't matter if I was going to be evaluated or not, I was

going to do what I thought was the best I could do." With this sentiment, it is obvious that

Ray holds high expectations of himself, and he expects his school leader to do the same.

Ray illustrated *leadership* in a variety of areas, at the same time, he recognized that

administrators require teachers to do many things including adopting specific strategies or

practices in the classrooms or school procedures. A portion of a quote shared previously

highlights an example of administrators demanding teachers to adopt specific practices.

"That was the one that most administrators latched onto. Then they expected cooperative

learning 24/7." When discussing performance evaluation, Ray identified administrators

demanded better performance rather than coaching it up in teachers.

> I believe the purpose for many administrators is to get it documented so that they can say they've done their job and the reason I say that is that a lot of our teachers, our young teachers, experienced teachers, it doesn't matter, really need help. The administrators don't go to their room and help them out. They just say, you need to improve this, you need to do that, they make suggestions, and I know the TLs [teacher leaders] are supposed to do that, but they don't either.

Ray identified one of the highest demands on teachers, to raise test scores, as coming from all

areas such as the media, the state, parents, and administrators. Ballou and Springer (2015)

conducted multiple analyses regarding errors in the design and implementation of evaluation

systems, and they determined a statistical significance suggesting some teachers alter

assessments scores used for their accountability.

> Like our analysis of roster verification, the results suggest that teachers are taking advantage of opportunities presented by the system to improve their own measured performance, in the one case by dropping from their rosters students who will not harm their value-added scores and in the other by providing assistance to students taking tests under their supervision.

With demand comes pressure to perform. Ray implied he had heard such stories occurring in

his district, but he declined to share them, as he did not witness them directly "That's

185

probably where a lot of teachers end up cheating on assessments and administrators pushing teachers to cheat on assessments. Got to get those test scores up, and they got to get their merit pay."

Pressure, as Ray identified earlier, was applied by some administrators attempting to force teachers to leave or take on new assignments and duties. "I know that a couple of administrators, they used it to get rid of teachers rather than trying to help them. They used it to put pressure on them to send them out the door." The pressure does not always rely on teacher performance as sometimes it may originate in personality conflicts. "I've seen administrators push people out of a building just because they didn't like them, had nothing to do with anything but their personality." Ray's years of experience has allowed him to witness a variety of school leaders and approaches, each wielding different expectations for individuals and groups of staff members.

Ray's longevity in education and at Hawkins Middle School provided an opportunity to experience Culture of Evaluation lead by a variety of district and school leaders allowing his perspective and understanding to change and develop over his career. While Ray shared experiences that were somewhat positive or neutral such as peer relationships with administrators that allowed him freedom and trust, many more of the shared experiences were negative in nature highlighting poor management and high stakes environments. This is important to keep in mind as the third theme is introduced.

Perceptions of Value. Over the course of Ray's long career, his Perceptions of Value evolved along with his own sense of self-value and assessment. The evolution occurred within a context of *standardization* as evaluation protocols dictated what would be measured, by whom, when, and how often. Consistent, constructive procedures including routine

observation and discussion lead to improved teacher performance (Reinhorn, 2015; Shugart, 2017). Ray describes his first few years being observed and his performance evaluated as intimidating, but with experience and success, he began to feel more confident and comfortable.

> My first few years of teaching, I was, I don't want to say scared to death of them, but I was concerned I was going to be evaluated in a poor light. Then administrators and other teachers would come in and tell me how great I was doing. I did something better than they had ever seen it done before. So, then I got to the point well, must be pretty good teacher, I'm not the best teacher ever, but I must be pretty good.

During his many years at Hawkins Middle School, Ray interacted with a variety of consultants and instructional coaches. Ray described an interchange with such a coach that was quite impressed by his performance in the classroom.

> She was, like a lot of teachers will have a timer and you've got to tell kids four or five times, get over there and get over there. I didn't do that. Show the kids a couple times how to do it, and it was upon them to do it from now on and that's what I expected. I think that's something, if you have expectations of the kids and they know it, they're going to try to live up to it.

Experiences like these increased Ray's confidence, and he worried less about his performance when being evaluated. Even when told to sign off on evaluations from supervisors that had never been in his class, he was comfortable doing so.

> It's not going to be the current principal. Probably my, I don't remember, it was probably my second or third principal that I had. He just came in with the evaluation and said sign this. I said, can I read it first? And he goes, sure. I read it, and it was what I expected him to say about me, so I signed it. But he didn't evaluate me.

Ray described the district and school's standardized evaluation protocol related to observations.

> Some people, they had to do two evaluations in the fall and one in the spring. For some people it was one in the fall and one in the spring, I think. I think it had something to do with how many years they were in the district. New teachers had to do two and two or two and one. I think new admin are supposed to evaluate everyone in the building.

Ray referenced how often he should have been observed when describing his experiences, but each time it was to highlight how his experiences deviated from the process rather than fitting into it, such as the administrator providing an evaluation without visiting his classroom. It is not surprising Ray did not invest heavily when selecting goals for his later evaluations, as his experiences had trained him that they would not carry any weight or consequence based on his performance. Ray did not invest much interest or effort with his most recent selection. "Some teachers' goal was, mine was to use technology, and I used technology. I just put that down as a goal for them to have something to look at. I certainly didn't use it like some of the younger teachers use it." Ray also did not describe the evaluation process or experience as very dynamic. When describing his last evaluation conference, he shared, "She [assistant principal] came to my classroom, took notes, and went to her office, and she told me what she thought, and I signed it. But I don't remember anything specific." As an experienced teacher, the actual protocol that should have been in place was to select at least three goals from 60 elements grouped into four domains. Then from the time the goals were selected until February, Ray should have had a minimum of three formal observations of at least 30 minutes each with a pre and a post conference for each observation followed by a formal summative evaluation summary conference requiring his signature. Based on Ray's shared experiences, it is safe to say the protocol was not strictly followed. The purpose of the protocol is to allow administrators to accurately and consistently evaluate teacher performance.

Ray's experiences illustrated that even a standardized protocol does not necessarily receive standardized implementation let alone consistent performance evaluation. Repeated

dissonant experiences may lead to a negative school environment, which takes us to the next aspect of Ray's Perceptions of Value.

Toxic carries a strong connotation of poison, danger, hazardous or foul. Much like in a science lab, the presence of toxic material is not in itself harmful provided it is contained, isolated, and disposed of properly, and the same time, if it spills out, it can contaminate and ruin everything it comes into contact. Beginning with Ray and carrying through the remaining participants, their Perceptions of Value contain *toxic* components. Administrators influence teacher retention, day-to-day working conditions, and evaluation experiences while at times ignoring proven practices such as mentorship resulting in greater harm and less benefit. (Abdo, 2017; Boyd, Grossman, Ing, Lankford, Loeb &Wyckoff, 2011; Smith & Spatariu, 2015). As noted in the previous theme, while Ray has shared some positive and many neutral experiences, many more components reflect negative experiences, which influence his perspective of performance evaluation experiences. To highlight a few critical concerns shared by Ray's experiences, it is apparent the previously discussed standardization has not been present. First, Ray recognized that administrators are often unfamiliar with instruction, so they are not able to provide coaching and guidance.

> The problem is, some administrators don't understand instruction, so they really can't help anybody because they don't know it themselves. One of our admin came from a business background, and then they became a business teacher, and eventually became admin. They really didn't understand instruction. The reason they became a business teacher, it was in high demand, so they were being pursued, kind of like science and technology now. There is such a lack of it, they will take anybody so that they don't have a sub.

In addition to not having a deep understanding of instruction, Ray shared that in his experience administrators often fail to perform their evaluation practices successfully, and at times it does not appear as if any effort is invested on behalf of the administrator.

I think new admin are supposed to evaluate everyone in the building. I know an administrator that didn't. She wrote up the same evaluation for several people and a couple of the male teachers got, "she does a good job." Could you change that to he please. I'm sure she wasn't the only that did it. It was not surprising that she did it.

These illustrations coupled with those shared as part of previous experiences culminate in a variety of complaints that could justify a cry of unfairness on behalf of all involved.

Ray identified experiences that he felt were unfair in nature such as some teachers seeming to get special treatment.

Some people would leave early, come in late, not get subs, things like that, and some people would complain about things, and okay, but other people would complain and he [the principal] would be on their case about it, "all you want to do is complain." Then they complained a lot more. But I think that was expected of them.

Even when it is not specific individuals involved, there are those school practices that are seemingly put in place to support student learning, but from teacher perspectives appear unfair in nature such as intentional class size disparity.

I think, science and social studies usually had larger class sizes because Math and ELA were always considered the priorities so they would have a 2 to 1 ratio with the teachers for Math and ELA or maybe a 1.5 to 1 teacher ratio. They would have more sections, so less kids per section, sometimes that didn't work out either, depending on where a kid had to be slotted. One class would be 40 and another would be 20. It usually wasn't that way in math or science now. Science is the big one now, trying to keep their class sizes small.

Unfair experiences seemed so pervasive to Ray, that even the thought of some new practices generate an assumption of unfairness.

I don't know that I would make any changes in the [evaluation] process as long as it's being done legitimate. Because I know, this might be off topic, but performance pay, I am not a fan of that. I know administrators say they would never do it, but a teacher they like, they could give them all the best kids. They would perform well on the state assessment. Teachers they didn't like could get the low kids and they wouldn't perform well, so who is going to get the merit pay. I'm all for bonuses and getting money, but there's got to be a better way than that.

Ray's Perceptions of Value holds a knowledge of what practices should be in place, but the inability of administrators and past experiences result in a negative, *toxic* lens through which new experiences occur.

As I received Ray's experiences, encountering his frustration with the status quo of the larger evaluation system, he held an underlying tone of resignation, acceptance that this was how it has always been and always would be regarding evaluation. Ray did not rail against the inability of administrators. He did not condemn the evaluation process as a whole. Much of his experiences were delivered as a matter of fact and the implication being that this was everyone's experience. At the same time, even though he accepted this as common practice, his dissatisfaction remained clear. While this might lean towards a stereotype of a veteran teacher, another participant has had double-digit years of experience in education with Hilltop School District at multiple schools and has had almost a decade of experience at Hawkins Middle School, and she provides a somewhat different point of view.

Case 4 – Nancy

Who does this guy think he is? Nancy struggles to keep her mouth closed as her assistant principal talks to her about her classroom and students. She felt her jaw begin to drop multiple times not believing what she was hearing. This guy, who entered the educational field because he wanted to coach sports, kept droning on about how he thinks the classroom could be better because he talked to one of her non-English speaking students and that student could not tell him the main point of the lesson. He also spoke with a student who had spent the last month in the office more than in the classroom, and the assistant principal knows that because he was the one dealing with the student's discipline. He just keeps talking about how great he was in the classroom without giving any clear direction or advice.

By the end, Nancy's jaw hurt from holding her mouth closed. She did not say anything only nodded, until he finally left, and she said, "bye."

Nancy closes the classroom door behind the assistant principal, turns away and silently mouths a scream. Listening to that man creates equal parts sickness and rage. Twenty-five minutes that she will never get back and nothing to show for them other than he puffing up his ego a bit more. The only compliment ever given reflected a team performance that included a teacher he favored, and likely, the compliment was meant more for her than the whole group. Nancy internally talks herself down the entire drive home that evening in order to let it go before entering her home.

Only a few years before, Nancy recalls working with a truly effective principal. They butted heads at times too, but every time they worked together, she walked away with something she could use that same day or the next. Mr. Timmons taught in the same classrooms Nancy eventually taught as well, a mix of students of different backgrounds, cultures, and even languages with the ever growing English as a second language population. He knew what worked with students and what fell short of productive. Not only did Mr. Timmons know what he was talking about, he could also coach. He led productive staff professional development. He modeled lessons in the classrooms with students. He coached teachers in and out of the classroom about more effective practice. He did not talk about himself; he spoke about students, their needs, and how to address them. She still used many of the strategies she learned from him. When he visited her classroom, she always knew she would gain something whether she liked it or not. Nancy misses those conversations and growing opportunities.

Looking back, Nancy recognizes the rarity of those conversations and a productive principal-teacher relationship. Most of her principals and assistant principals, during her career, managed the school well, good people, doing their best. She enjoyed working with most of them. She changed schools at times looking for new opportunities to explore and to challenge herself. The new settings, colleagues and students have provided growing and learning opportunities. Nancy tackles new content subjects and stretches her professional skills often, and when necessary, leans on her peers to provide input and work through challenges with her. Unfortunately, administrators other than Mr. Timmons, have had little to contribute. One of her frustrations occurs regularly. In her career as a classroom teacher, consistently, her evaluations and coaching assignments land with the assistant principal that has no background in the subjects she teaches. Their comments and feedback almost always come back to her relationships with students which she knows is a strength and is recognized by her supervisors, but that does not push her to grow nor offer her the avenues to pursue her practice.

Her supervisors rarely have the content knowledge or insight into the curriculum to guide and direct her performance. They talk about strategies in a general sense. They talk about language and vocabulary used in a superficial manner. They point out she is aligned with the curriculum, and the students are participating.

Nancy recalls a recent observation the assistant principal held in her classroom. The students were taking a test online, and as the students completed the test, she could see the results immediately. Nancy called students up to her desk to talk to them about their scores, give a quick lesson on something missed and send them back with practice problems targeting specifically the area where the student did not perform as

193

well. She did this from her desk at the back of the classroom and could observe the entire class as she called students back to talk to her. The assistant principal came in for about five minutes, walked around the classroom and the students were testing and Nancy called students back to her. Later that day, the assistant principal spoke with her about the lesson. Nancy felt more talked to rather than with about the lesson. He directed her to monitor the students by walking around the classroom, so she can see what they are doing as they do the work. He shared that this way she would know what feedback to give them after they were done. Nancy had to hold in a scream of frustration that day as well. She tried to explain how she was monitoring and differentiating feedback as the students competed the task, but he just repeated his original direction.

As she pulls into her driveway, Nancy shakes off the frustrations and resentments building from them. She knows she provides quality teaching to her students and is confident in her work. While she misses the development gained from working with Mr. Timmons, she knows, using what she gained from him, she provides quality service to her students, and that her current administrators may or may not recognize it. But she knows, and she is fine with that.

Classroom Practices. Nancy's case aligns with Eric and Paul's regarding **Classroom Practice** as it focuses on *teaching practices* and *student measures*. For Nancy the phenomena contributing to *teaching practices* incorporates the development and learning of those practices along with content knowledge. Nancy firmly shared that strong content knowledge is needed to coach and guide a teacher of that content, and an administrator without that background has little to no credibility.

So, to me, coming in and evaluating me, it really had no bearing on me as a teacher as far as am I doing a good job or am I not doing a good job, because to be quite honest, he [the principal] couldn't tell me that because he didn't know my content.

Consistently, Nancy identified content knowledge as a key factor in understanding, supporting, coaching and evaluating classroom teachers.

I would almost say, if it's even possible, which I don't think it is because it would put peers too much on the same, but I think that we are able to evaluate each other better than someone who is not in the classroom. People in the same content evaluate each other better than someone not in the classroom or haven't taught that particular content.

Nancy continued to emphasize content knowledge as vitally necessary to effective coaching and evaluation of classroom teachers, suggesting fellow teachers in the same content would be more effective than administrators.

I had a teacher leader, coaches, to be quite honest I think we should almost be evaluated by our instructional coaches because they have, well at least they have in the past, had that same content area or they have taught in it.

Some school districts have found success training teachers as evaluators (Fairman & Mette, 2017). Still others have adopted a specific model of teachers evaluating teachers called peer assisted review (PAR) (Smith, 2015). Content knowledge represents one aspect of a teacher's actions in the classroom while pedagogy and behavior managements represent other aspects.

When discussing points in her career when she felt she learned from administrators, she repeatedly referred to a principal from a school she worked at prior to her current assignment. She described that administrator, not unsurprisingly, as having strong content knowledge and stated she believed he had gone back into the classroom after leaving the principal role. "I would probably say that during that time period, I grew the most as a teacher, even though he and I had some conflict." Nancy shared that this administrator was in her classroom often, and he spoke with her often about what he saw in the classroom and

ways to improve her practice. "I still learned a ton from him. I learned, he would come in and he would talk about anchor charts, and I had no idea what an anchor chart was before that." Even when not engaged with the principal, Nancy would consider his insights important and look for ways to emulate others that had his approval regarding their practice. "I would look at the things he would like about her, and I would be like, okay, this is what I need to do, to change." Nancy carried her learning with her to new schools and assignments and throughout her practice. "When I said something about the vice principal, and about the praise, giving praise, not overly, but giving praise as needed, I learned that there at that school. The anchor charts are huge. I still use them to this day." Much of Nancy's *teaching practices* developed through her pursuit to improve her performance.

INancy expanded on how she improves her performance through the use of *student measures*. She identified a variety of methods to measure student understanding, including standardized testing, exit slips, and pre/post assessments. She also recognized that some had more emphasis than others.

> Well, the standardized testing is of course a little more high stakes; so definitely reading and math have more of a pressure put on them to make sure they are doing what they are supposed to be doing with the standards and whatever the school has chosen.

In addition to the pressure some high stakes assessments created, there was the added pressure of conflict present between Nancy and her administrator at times that even successful student scores could not overcome. "She was extremely critical, and the only praise I got from her was finally when we took the state assessment and my kids absolutely rocked it in math." Ultimately, Nancy feels she succumbed to pressure like many teachers do, running out of time, things have to be cut.

196

I would collect exit slips. All of these things would show a picture of where they started. We did do pretests at the beginning of the year. As time crunch happened, pretests are wonderful, they're a grand idea, but when you start crunching, trying to throw something in to 90 hours that take 500 hours, then you start cutting things.

Only so much can be cut though because one of the points of assessments is to measure student growth. Nancy referred to using standardized test results for one measure of student growth, but at the same time recognized that one-shot tests do not always provide the most accurate picture of a student's learning and potential.

We of course have the [NWEA] MAP test that I really like, so that I can see from year to year where they've grown and into the year where they've grown and even have that broken down into the different number sense, algebra, geometry, and data, so that I can see if they've grown in the areas that we've actually gone over. But I also understand those aren't foolproof either. If they're having a bad day, they can definitely go backwards.

With this in mind, other less formal methods of student learning can be employed, many of which can occur throughout daily tasks and activities.

I rely mostly on my teacher observations as far as my student growth. Walking around, asking questions, looking at their work, seeing what they, within the talking of the group, if they are truly understanding it or not, and of course, I use everything that I do in my classroom as a way to see where they're at and monitoring they're growth and they're understanding.

When discussing what goals she strives for to know students have learned what she intended to teach them, Nancy points out that they are a little bit different for every student. She recognized the second language learners may be more focused on social language acquisition while special education students would be focused on individualized goals.

So, I have been looking for everything under the sun for growth from him. Did he ask me for a pencil, or did I have to figure out that he needed a pencil? Did he go get his iPad without me? Having to figure out did he need his iPad? Did he open up and get on the app by himself without me having to come and get him logged in? Even those tiny growths, it has nothing to do with my content, but he has everything to do with the success of my class running smoothly.

Even when referencing her students that do not receive a specialized curriculum, Nancy

points out that different students start in different places even when all are trying to reach the

same goal.

> What I might expect from one student, when I'm walking around and looking at their
> paperwork, I may not have that same expectation with another student. I may just be
> excited that they can do that first step on their own and without having to ask for help
> with the second step. Whereas another student, I might expect them to get all the way
> through the process or algorithm without having to ask questions and then I can give
> them a more complicated problem.

While individual growth is still growth, required state assessments still focus on reaching

certain levels of achievement rather than demonstrating individual improvement.

Nancy's Classroom Practices consisted of a focus on content knowledge and proven

strategies learned from peers and administrators while measuring student improvement

individually through a variety of assessments and tools. Throughout out our interaction,

Nancy portrayed confidence in her actions and abilities, but when addressing the act of

evaluation, she expressed displeasure. She felt she performed well overall, but when being

observed or evaluated, she described negative experiences unreflective of her abilities as an

educator.

Culture of Evaluation. From Nancy's experience, peer interaction and support

offered opportunities for improvement while administrative support, holding that same

potential, often resulted in empty words and actions. These two viewpoints will be explored

more in their respective interpretive codes. The first, *peer interaction*, spotlights the

opportunities and challenges present when engaged with fellow teachers. Nancy repeatedly

emphasized the benefits of engaging with peers, especially those teaching in the same

content. She relished the opportunity to share ideas and discuss what was working and what

was not as successful. During past practices, she shared how beneficial it was to be able to

build peer observations into their regular weekly practices.

> I thought that was one of the incredibly positive and impactful, being observed, not
> necessarily evaluated in a formal way, but just be, hey these are the things you're
> doing great and we always left little notes for them, but also for ourselves, how can
> we take it back to our own classrooms.

Nancy stated her belief fellow teachers with content knowledge would be more effective

evaluators than administrators without that content knowledge. "People in the same content

evaluate each other better than someone not in the classroom or haven't taught that particular

content." At the same time, not all peer collaboration achieves equal benefit. At one point

Nancy shared frustration with an administrator and pointed out that even when her students

demonstrated success, the principal attributed that success to the other teacher working with

the students rather than herself. "They did math and reading but I had additional support in

reading so therefore she felt like it was because of the additional support that they did well."

With another administrator, Nancy encountered misperceptions of effort and worth when

operating in a collaborative situation with peers. In this case, it resulted in a strained

relationship with an administrator for multiple years, all because of peer miscommunication.

"One colleague was like, oh my goodness, I went and told him this, this, and this, but he

must have thought I meant this, this, and this about you." Even having encountered some

negative results from peer interaction, Nancy recognized the power of working with fellow

teachers. When asked what steps would benefit a teacher whose students were struggling, her

first suggestion was to partner the struggling teacher with a peer mentor in the same content,

"maybe some mentoring from another teacher." Rather than receive coaching support from a

principal, Nancy consistently looked to peers for support rather than to her school leaders.

Nancy's allegiance leans away from principals and towards peers due to negative experiences with *leadership*. As previously highlighted, if an administrator lacks content knowledge in the same field as a teacher, Nancy felt they lacked the ability to provide insight and guidance for improvement. As a result, feedback from administrators rang hollow for Nancy resulting in a waste of both her and the administrator's time.

> He could tell me things like, I appreciate that, I like that you know how to give praise, and that you give praise easily in your classroom and it's something that slips off your tongue and just happens, but to me, again, it didn't mean a whole lot to because he was a gym teacher before so, he couldn't really help me with my content coming in.

Nancy repeatedly supported the practice of peer observation and discussion over that of administrators observing and providing feedback.

> I think that people don't understand how important that is to look at other teachers. I think that's more important to me than having somebody who has never taught my content come in my classroom and tell me that hey, you went from developing to applying.

Nancy has reached a point where the thought of being observed or evaluated no longer matters to her because she feels it carries so little value. "When they tell me I'm being observed, then I know. I don't care. It doesn't bother me." She described being evaluated like exchanging emails. "So he has come in and done that before and posted it, and I would go online and read his comments and respond to them." Nancy described it again.

> They come in. They observe. They put it on the online thing. I fill it out and I say I agree or disagree with it, and write comments. That happens. Like I said, I have no idea when it's going to happen.

Even though Nancy said she does not know when she will be observed, she does describe her current principal as having clear expectations related to classroom visits. "The way things are put on there are nice and clear as far as, this is what I'm doing, this is what he is going to come in and look for." Part of Nancy's confidence or ambivalence towards being observed

may result from knowing what will be looked for during any observation. "Expectations that we are following the domains, and that we are doing the school's, I don't even know what to call them now, the school's expectations that we've chosen as a school to follow." In addition to knowing what is expected during an observation, Nancy also believed evaluations and observations can be effective tools if used correctly.

> I think the best evaluations, and I just have to put this out there, are the evaluations that truly let a teacher know, what you see in the positive and what you can work on. If you don't give any feedback whatsoever, I don't know what you've seen in my classroom. I don't know what you're looking for. I don't if you've found what you're looking for. That truly is the best thing a teacher can get, no matter if they're a first year teacher or a thirty-year veteran. You've got to at least hear from the person that has come in and watched you.

Nancy spoke very highly of one principal in particular sharing multiple experiences where she felt her practice improved; not surprisingly, he had experience with teaching the same content as Nancy. "He would write on there the positives and then write on there some things he would like to see worked on, and I know he evaluated everyone every year because he was just constantly in the classrooms." These types of experiences were outnumbered by those where administrators provided feedback asking the teacher to do things she felt were counterproductive to the lessons she was providing. "So, he was like, maybe next time you can be up moving around monitoring, and I was like, I had to write on the thing, this was a test, and I monitored on my computer." The culture described by Nancy relies heavily on *peer interaction* and tolerates *leadership* action.

Perceptions of Value. Nancy's relationships with peers and principals likely stemmed from her perceptions of those individuals and the roles those positions played in her daily work experience. Nancy's Perceptions of Value did not present her career as mired in a series of *toxic* environments, there were aspects present in most of the settings she described

that contained negative influences on the rest.. As an example, Nancy described biased

treatment, poor work relationships, and favoritism.

Many schools heavily favored math and reading instruction because in recent history,

these subjects received the most attention related to state assessments. Nancy described these

subjects and teachers of these subjects receiving additional pressure and demand while non-

tested subjects are unmolested by such attention.

> Well, the standardized testing is of course a little more high stakes, so definitely
> reading and math have more of a pressure put on them to make sure they are doing
> what they are supposed to be doing with the standards and whatever the school has
> chosen. As you get further away from the testing, such as with our encore, it's
> definitely not, I don't get looked at all for the encore class. Period. I don't even turn
> in lesson plans. He [the principal] could care less.

At the same time, some of this may be the result of poor professional relationships. "Her

evaluations were incredibly unprofessional, in fact, my second evaluation was, wow, we see

Mrs. Nancy can teach math." Whether as a result of poor relationships, poor evaluations, or

poor student scores, Nancy's perception was that some teachers are favored and protected by

school leaders, and those teachers received easier assignments and more recognition.

> My colleague, I feel, in the last five years, since the principal has been there, ends up
> with the higher class of when we would level them. She always ends up with higher
> kids, and I never understood why. One time, the vice principal said that I accidently
> got a class that was higher, and I said what's wrong with me getting a class that's
> higher? Why was that supposed to go to her? I don't understand that.

On the other hand, Nancy stated favoritism directly resulted from non-classroom acts that

have nothing to do with student performance.

> I have a teacher that gets a lot of recognition because she volunteers for everything.
> She likes to do all the extra stuff, and she's seen as a really good teacher because of it.
> But we don't know how she is, inside a classroom or how her classroom is. It's
> mainly because the admin is like, yeah, she is volunteering for this. So, we don't have
> to go around and ask somebody to do it. I concentrate more on my classroom, so I
> don't necessarily play that game that well, because I don't need anything extra. I need
> to just focus on my classroom and how that's doing and that's all the time I have, is

for that. So, I think some people look at a teacher and say hey they're a really good teacher because they do all this extra stuff versus, hey she's concentrating on her classroom doing what she needs to do, more of a silent teacher that maybe people don't necessarily see what they're doing and how well they're doing.

These experiences, while maybe only lasting a passing moment, linger with Nancy providing a *toxic* lens to her perception of other interactions.

At the same time, Nancy also pointed out that she recognized in conflict, she played a part as well, and the end result can still be positive. "I would probably say that during that time period I grew the most as a teacher, even though he and I had some conflict because what I feel like was misunderstandings about his ideas about my personality." Even if she perceived an experience as negative in the moment, upon reflection she may find it as beneficial now.

So, anyway, even though, some of the evaluations, especially in the beginning, I really had to look at and say well yeah I don't do that, I probably could spend some time, you know he did do an excellent job of evaluations now that think back to it.

Other experiences do not age as gracefully, and they continue to carry strong emotions when Nancy recalls them. Nancy became obviously upset sharing an experience, that at the close of one school year, her principal decided to reward those teachers whose classes had been the most successful on the state assessments taken that spring.

And it was really hard watching that because I know there was three math teachers and only one of them, her classes made it. It's really hard watching that because you don't get dealt, you don't have say in who you're dealt. So, you might have a bunch of really high kids, but just got lucky enough to get them, and you might have a classroom full of all 1s but you've watched them grow on their MAP testing, but they aren't 6th grade level so they're never going to show that on the state assessment. They're always going to show a 1. But you might have had huge growth. That was the only time she did that, and I think she actually got in trouble for it. I think some people called down town and said this is so not fair, and I hope that I never see that again in my teaching career period.

Even in the best of settings, incidents occur that leave a bad taste in the mouths of some. As Nancy reviewed her experiences, she shared from multiple schools and administrators, but she carried those varied experiences with her each year and into each new encounter providing *toxic* shade to the lens through which she views each new experience.

Nancy shared a list of experiences where it appeared that those in charge at a given time seemed incapable of communicating or lacked knowledge themselves. These experiences occurred across a variety of settings, and as a result *incompetence* further defines Nancy's Perceptions of Value. Nancy stated she rarely, if ever knew when she would be observed and evaluated. "Like I said, I have no idea when it's going to happen." Even when she would be observed, she believed their feedback was moot due to a lack of content knowledge. "Because to be quite honest, he couldn't tell me that because he didn't know my content." When discussing how principals decide assignments and communicate them to staff, she threw her hands up in the air sharing she has been surprised by her own assignments from year to year. "Again, I didn't know I was going to have that. I would assume that the principal just looks at it and decides what works best for the school and the student." Nancy also recognized her own lack of knowledge when it comes to describing peer practices and their effectiveness. "Once again, it goes by what have people seen because if you think about it, we really have no idea what goes on in somebody else's classroom. So, unless they are doing things outside of the classroom that are wonderful then you really have no idea." Individually, each experience lent itself to being waved off as an oversight or inconsequential, but when collected and paired with previous negative experiences, it is not surprising that a teacher may lose trust in her school leaders. Teachers lose trust in administrators, evaluation systems, and grow increasingly frustrated with poorly

implemented or changing evaluation tools (Braslow, 2016; Bush, 2017; Byford, 2018; Satty, 2016).

Another example of losing trust in administrators occurred during Nancy's most recent evaluation cycle. She described how the individual charged with completing her evaluation, regardless of her performance during observation, intentionally scored her low early in the year and then higher as the year went so that it would look like she demonstrated growth. While this benefited Nancy over the course of the year, she shared the manipulation of the evaluation system continued to detract from her trust of that supervisor and the district.

> They never want you to be the top thing because then you don't have anywhere else to grow, so you're usually stuck at what I call level 3 and 4, and you always put 3 down on paper so that they can get you to 4 by the end of the year, if that makes sense. By the end of the evaluation, and what it has done for we teachers is we just go through and, I don't even know if I should share this or not, but we just go through and mark that we're clear over here on the left hand side for everything, so that we can pick something we want to choose and make sure we grow on it. So how does that show between what I can actually do as a teacher and the evaluation is that I feel like I'm continually growing, and I feel like I'm clear to the right end of it. But, really, there's not a way to show that kind of growth on the evaluation scale because, nobody wants, the administrators don't want you at clear to the right. They want you to be able to show, they got to be able to show you're still growing, or they've got to show growth. There is no way to do that once you are clear to the right.

Incompetence captured a description of a system that continually undercut itself through such manipulative practices. Even if given the benefit of the doubt in each instance described by Nancy, the collective result painted the image of a system where the teacher did not trust the system to move forward successfully unless changes and improvements to practices occured. Nancy's Perceptions of Value held a *toxic* view of an *incompetent* system.

From my own interactions with teachers, I wish I could describe Nancy's experiences as unique, but many teachers I have worked with over the years have expressed similar encounters with administrators and evaluation practices. I mentored a new teacher who

received her first evaluation from an administrator that had scored her on the lowest third of the rating scale in all categories in the fall, and on the highest third of the rating scale in all categories in the spring without having observed in her classroom all school year. My mentee was furious following both evaluations, more upset by the false nature of the action than the actual score.

Case 5 – Lance

"I've got to be honest with you, I've about had it! This is a joke. I'm just tired of it, tired of all of it. I mean, what's the point? Really, tell me, what's the point? I've been there longer than almost everybody. I come in everyday, never miss a day, rarely miss a day; I can count on one hand the number of days I've missed in the last ten years. I am one of the first people there, and I'm one of the last ones to leave each day. I'm prepared for my classes, and my students leave my room having learned something they didn't have when they came in that day. So, you tell me, what's the point of bringing in these outside consultants that are supposed to tell us how to really do our job. I can't believe it. I just can't believe it." Lance vented to his colleagues over drinks after work one day in the fall. The school principal had shared that afternoon during the staff meeting that 'guests' would be coming in, at the district's request, to visit with most of the teachers about improving their practice, modeling lessons, providing feedback, and adopting new teaching strategies. While many of the schools were also participating, it just felt like a punishment to Lance. He felt as if the district and principal were telling him personally that he was not doing a good job. No one actually said that, but Lance certainly interpreted the new practice that way.

Over the following months, Gary, the consultant assigned to Lance's content and its teachers, met with Lance individually every other week and as a building content group the opposite weeks.

"Lance, how are you today?" asked Gary as he came in during Lance's planning period. "You ready to look at the upcoming unit?"

"Sure," replied Lance restraining himself from rolling his eyes. "Let me finish putting away this student work from last period, and I need to do a quick set up for the next period." A few minutes later, "what was it you wanted to look at?"

"The next unit, have you started preparing for it?" replied Gary.

"Yeah, I can show you what I have. Here it is," said Lance sliding his laptop over to Gary to share what he had put together.

"That's really good, better than I was going to suggest. Can I have copies of that so that I can share them?" asked Gary.

"Sure, no problem," said Lance.

Over the next few months, this pattern repeats over and over again. Lance feels the consultant is just taking his work and sharing it out, never really providing support to anyone. Lance brings this concern to his principal, and she says she will look into it, but Lance never hears anymore about it. Lance's frustration continues to grow, and he stops sharing his work with Gary. He still meets with him per the district expectation, but he doesn't provide copies of his lessons or materials any longer. Lance notices Gary spends less time with him and cancels some of their meetings to address concerns in other rooms. The absence of the consultant is appreciated by Lance who does not miss having to sit through the in person and conference calls. The entire experience left Lance with a bitter taste in his mouth.

As it turned out Lance's experience was not unique and the following year, the contract with the consultants was not renewed. Building administrators were tasked to perform similar consultancy structures with their teachers rather than the former observation feedback cycles previously adopted. He described an evaluation session

"Welcome class, let's begin our lesson today. Let's start by writing down everything you know about the topic on the board. I will be coming around looking at what you write down to inform me of what you already know and what you still need to learn. This is a quiet activity, so eyes on your notebooks and mouths silent. Thank you," said Lance as his class began. As he moved around the room, the principal and assistant principal came in together with their laptops. They took seats next to each other in the back, talking quietly to each other.

"Looking good students. I'm seeing a lot of good answers and insight. You already know a lot about this topic. That's good and means we will likely get to the fun part of the lesson quicker. As you are working, know that I am looking for five key details. If you have more than that, put stars next to the most important five. If you have less than five, then try to think what the others may be," directed Lance as he continued to walk around the room looking over student shoulders.

The principal waved Lance over to ask him a question, "what are the students writing?"

"They're writing down what they already know on the topic as an anticipatory activity," replied Lance.

"So what are the five things they are looking for?" asked the principal.

"I don't want to say right now because they will hear me and write that and eliminate the opportunity for me to learn what they know," answered Lance.

The principal stopped him from walking away again, "why didn't you write the details on the board so that all the students can be sure to get them?"

Taking a deep, calming breath, Lance answered, "I am trying to see what they already know. If I write them on the board, then they would just be copying them down. It doesn't show me what I am trying to find out."

As Lance continued his lesson, each step was analyzed and suggestions made in the moment that contradicted the intention of that part of the lesson. At least Gary just took my work instead of challenging and trying to change everything I do thought Lance. He thought, sometimes, he needs to be careful what he wishes for because it may not be better than what he had.

Classroom Practices. Much of Lance's case follows the same path as Nancy's. Each theme holds the same interpretive codes present in Nancy's. Similarly, Lance's *teaching practices* held a strong focus on content knowledge also. Lance's conversation focused on content knowledge and curriculum. From his point of view, in order for someone to teach a subject or observe and provide feedback, then an underlying knowledge of the content and curriculum is necessary. Content knowledge frequently dominated Lance's language. He highlighted his perception that his content requires daily interaction to follow along at the pace he pushes his students.

> My content is a little bit different animal. We had teachers in the district that knew what they were doing in it. But, you have to understand, and I don't hold it against the administrators, and that's in any curriculum, but you know, what we do in it, unless you are in their everyday, it's kind of tough to follow what's going on. It just is.

Lance emphasized that his comfort with the material originated with his experience and contributions to creating it for the district.

> So, if an administrator wants to know, hey what are you going to cover next semester, I could tell him because I wrote it. So they knew, and I guess more than anything, an administrator was never blindsided when they walked into my room. They knew they have a pretty good expectation of what they are going to see.

When reflecting on supports he had received during his career, Lance pointed out repeatedly a former teacher that had become an administrator consistently understood and was able to provide guidance and coaching.

> We would go through what I was doing in class, and any problems I was having, anything I needed as far as material. It was really good for me, because she [assistant principal] was a science teacher. I mean, she had a really good understanding of what I was doing, so I didn't have to explain a lot. We just kind of went through any discipline problems, anything like that, but she was in my room, maybe once every two weeks. She would come in for five or ten minutes and get a feel for what we were doing and then she would leave.

In a similar vein to content knowledge, curriculum was also a frequent focus. One of Lance's first comments during the interview related to his duties and responsibilities was, "number one, that you follow the district scope and sequence. The district has a basic curriculum, and that we follow it." Everything related to Lance's instruction began with the scope and sequence provided by his district, and which he reported he assisted in creating. "I wrote their curriculum. I know what we're going to do two weeks down the road. I knew how I was going to test them on it." When describing interactions with administrators, it continued to be a primary point of concern. "Her focus was on classroom management, and following scope and sequence. Was I following the guidelines of the district as far as curriculum, and classroom management? I took care of business in the classroom." Additionally, when describing his goals for performance evaluation, the curriculum continued to be a primary point of focus. "I identified the two things that I was going to work on in class for my

210

evaluation, and that's what we went through for my evaluation. For me, classroom

management, and curriculum, content were the most important for me. That was it." Lance

maintained his focus remained on those emphasized points in the curriculum. "We had

certain emphases on our scope and sequence. Make sure you hit this, this, and this. Test them

on these things, certain things. And that's what I would hit, and that's what most of my

assignments are on." Consistently and repeatedly Lance stated his work with students began

with the curriculum and his mastery of the content, and with that, he could teach any student.

For Lance, *student measures* consisted of assignments and assessment. Still stemming

from his content knowledge, Lance emphasized aligning his measures of student mastery

through his daily assignments. "The daily assignments were what we did every day. Because

I designed the assignments according to what the district wanted, the guidelines of the

district. And that's pretty much how I made assignments." Lance also expressed confidence

in measuring student understanding through questioning as much as unit tests.

> If I ask a kid a question and they know the answer, that's how I judged whether or not they knew something. Test wise, not unless I made it up myself, and we went over it the day before, and I would go over it on the overhead or some way and say, this is what you're going to see. This is what I'm going to cover, and so they had an idea.

Ultimately, Lance expressed measuring a student's growth took time and comparison of

assignments or activities.

> I measured a kid's success from day one to the end of the semester by how well they did on assignments everyday, and how well they improved on assignments everyday. And that, more than anything, because I could take a kid, first day of school, and an assignment they did, and where we moved from there to the end of the semester, hey this is what you did the first day, this is what you're doing now, regardless of what they did on a test.

Overall, he disapproved of the use of larger, standardized tests for measuring student

understanding. "To me, the worst way we can measure student growth is to give them a test.

In our district any way." At the same time, Lance recognized that school and district

expectations did not allow him to dismiss or ignore more traditional tests or assessments.

> Every nine weeks, we get tested about what they did that nine weeks. Those were
> pretty important. And we have to move relatively quickly. I mean every day, we have
> to get stuff done. To meet our goal at the end of the nine weeks, that's the test. We
> just have to.

Lance identified the importance of test anxiety and working with students not to worry about

a one-shot exam. He coached students to recognize a variety of methods to measure their

progress.

> I said, students learn in French, I teach in German, the test is in English. So you
> know, I would tell kids this is just a test. Take it. Do your best, but understand there
> are other measures of your success besides taking this test.

In his colorful way, Lance summarized the need to measures students multiple ways in order

to truly understand their mastery of content. Lance's Classroom Practices entailed a heavy

focus on content and curriculum delivered at a steady pace and measured through

assignments rather than formal assessments.

Culture of Evaluation. *Leadership* and *peer interaction* further detail Lance's

Culture of Evaluation. Lance described working closely with content peers, those teachers

teaching the same content but with different students in different grades. He met with them

regularly throughout the school year; they visited each other's classrooms to provide

feedback and support and at times covered each other's classes when it was needed.

> I think in [our] department, we really worked at giving each other pats on the back.
> We talk about those days when we would have meetings, and we would say hey, nice
> job with what you're doing. More than anything else, we were there for each other.
> Another teacher came a little bit later on, and he was starting to work in too. We just
> got along really well. But as far as from the district or from the building, I guess they
> were there, but for us, it is more within the curriculum, more within the curriculum.

Even though they incorporated very different teaching styles and approaches, Lance

and his peers recognized benefits from collaborating and learning from each other. This

is exampled when Lance described going into a peer's classroom that incorporated a

very different management approach with students.

> Disorganization, kids doing all kinds of stuff. Talking to each other, just, not paying attention but somehow or other, when it came to doing whatever he needed to do to get them to understand what was on the test, they did it. Because, I mean, his tests were, his district scores were just hard to explain, but they were good every month. I mean I saw them every month. They were just good.

Lance described the building wide practice of nominating peers for monthly recognition

as a mostly superficial practice, but at the same time he acknowledged the importance

of being recognized by peers.

> With [them], and our meetings that was our recognition, more so than anything else, and that was just for me. Other teachers may say hey, you know what, I want somebody to recognize me every week. And that's fine, but for me, it just wasn't.

The strength and peer reliance was evident in much of Lance's description, which led to

alignment of understanding, support and communication within his *peer interactions*. Much

of the alignment originated from being in each other's classrooms for learning and support

purposes.

> We were pretty much in alignment with what we were expecting, and I was in his class several times, I was in the other class. I even subbed for him on occasion. So I knew what was going on in their classes too.

Lance described professional development meetings with content peers as an incredibly

beneficial time to learn from each other and provide feedback to one and other.

> Believe it or not, district meetings. District meetings when we would have other teachers come and talk about what they did in the classroom. More than anything else, when somebody came in and said, hey, this is what I'm doing in class and they bring stuff, that, that was good. That really was.

Lance explained he led his content group of teachers in the school similar to a department chair.

> I coordinated all of our meetings between 6th, 7th, and 8th grade. Any emails that came from the district to me, we would discuss any emails that came across, and then, we would just talk about what we are doing. What we were covering that month, that week, what units we were all in, any questions that any of them had.

Overall, Lance believed strongly in the interaction and support of peers to ensure his own effective practice and supporting each other in their content. At times, Lance felt a similar alignment with the *leaderhsip* team, but not with all of them.

Lance, having had a long career in education in multiple school districts, has had multiple principals and assistant principals provide *leadership*. He spoke of one of them incredibly highly, repeatedly. "She was there enough times. She knew what I was doing. Well not only that, she was in my room too. She would come in, just walk around and sit and look at the kids, observe, and then write me a note. She was really good." In response to how often she visited, Lance replied, "She was in my room, maybe once every two weeks. She would come in for five or ten minutes and get a feel for what we were doing and then she would leave." When asked what she would look for when she visited, Lance shared, "Her focus was on classroom management, and following scope and sequence. Was I following the guidelines of the district as far as curriculum, and classroom management. I took care of business in the classroom." He went on to add, "you've got to be able to manage the classroom, and every nine weeks, we get tested about what they did that nine weeks." To add to the description of his work with this particular administrator, Lance stated, "I knew what she was expecting. I identified the two things that I was going to work on in class for my evaluation, and that's what we went through for my evaluation."

When asked about other administrators he had worked with, he explained that some were okay and knew their business, but many were lost when it came to his classrooms and content. "An administrator needs to be in that classroom at least once every two weeks for 15 to 20 minutes. And then, I think, have a knowledge beforehand of what that teacher is doing that week." Lance emphasized more than once that to truly know how effective a teacher is, the supervisor has to be present and engaged. "I think administrators have to take the time to be in the classroom. More than anything else, the administrator has to take the time to be in that classroom, more than one time to do an evaluation." Administrators invested in collaborative, constructive feedback based on observations and conversations with teachers increase student performance (Liang & Akiba, 2015; Reinhorn, 2015; Shugart, 2017). Lance described interactions with some of his less than favorite supervisors as disappointing.

> They never came to me and said what are you doing? What are you covering this week? They just never did. As a matter of fact, I got a feeling I know why. She told me one time when she was in the room, "I didn't like science at all." Well to me that doesn't mean anything, but you know, you don't need to tell me that. Because I've had a lot of kids that didn't like science that got great grades in my class.

Leadership varied based on the leader with one stellar example, a sampling of acceptable, and quite a few subpar interactions as described by Lance. For him, the Culture of Evaluation varied based on peers and leaders. The more access to peers in his content and leaders that had strong content knowledge, the more Lance felt successful and flourishing. When those preferred interactions were less accessible, Lance described the culture as much more negative, which connected third theme.

Perceptions of Value. As noted before Lance's case follow the same interpretive codes as Nancy. While each encountered unique experiences and interactions in their careers, it led them to similar points of view holding a negative connotation of their setting. The *toxic*

lens with which Lance interacts with his work environment stems initially from a sense of

unfairness. An example of this would be when Lance described his colleagues view of

evaluation practices. He shared his colleagues believe that rather than the quality of

performance being evaluated, that personal feelings or interactions influenced the results.

"They've said they're unfair. That it was personal. It wasn't business. It wasn't professional.

It was personal." Lance went on to explain why he thought his colleagues believed it was

personal.

> I think it's because they for some reason or another, their teaching became a personal thing with the administrator and not a professional thing. And the administrator, because of that personal relationship that they didn't like, that's how they got evaluated. And I'm thinking of one teacher in particular, actually, two, probably three, and they all quit. Well, they didn't, they just left.

From Lance's perspective, he ruminated quite a bit on his perception of unfairness residing in

class sizes. Lance explained that since reading and math were both tested contents and

received a lot of attention, that those subjects often had more staff and smaller class sizes.

> No rhyme or reason. None. I understand that at the 8th grade level, if you have one 8th grade science teacher, you're going to see every kid. But I couldn't understand why you needed, and I understand the importance of math, but I couldn't understand why you needed four math teachers. When our math scores were in the, and I got to look at math scores too, they were in the dumper for us every year. So, I couldn't understand why I'm seeing 160 kids and the math teacher is only seeing 35 or 40. And my test scores were higher, were better comparatively speaking than math scores. And I'm busting my butt every day for 10 hours.

Lance felt strongly about this topic, revisiting it throughout his interview. "I mean to me, it's

totally unrealistically unfair to me and others as science and social studies teachers. I should

have been paid twice as much. I saw twice as many kids." He shared that he spoke to one of

his principals about the disparity in class sizes at one point, but was quite dissatisfied with

the answer.

He said, hey, that's just the way we've set it up, and that's what the district wants to do. So, I said fine, that's when I said my days were numbered at that point because you know I got there at six in the morning and those teachers would roll in at 7:40 and leave at 3 and I'm going, I'm here at 6 and I leave at 5. I'm grading papers. I'm making tests. I'm running off assignments. You're done. I was, a good example, I'm in the copy room one morning, I'm running 160 copies for two days. A math teacher is running 160 copies for four teachers for a week. That was the difference. So, when you look at testing wise, if somebody comes in and says it's merit based performance, you can't do that. I'm seeing 160 kids a day. There seeing 45-50. We have language arts teachers seeing 30 kids a day and that got to hand pick the kids they had to teach. I could give you names.

Lance further illustrated his perception of unfairness when describing that some content subjects not only have fewer students, but the furniture and classrooms favored the more heavily tested content subjects as well.

I was in a meeting after that, and there was a language arts teacher who shall remain nameless, who said, what we need to do, there was 16 kids in that class, and I was in there watching because I was allowed to go in because it was on my plan time, and I watched her do her presentation on the writing, and this teacher said well, there were 16 kids in there. We should narrow that down, we should maybe have four kids in there, and I'm looking around and I put up my hand, and I looked at [the principal], and I said, hey how about if we do this, how about if you put thirty kids in that classroom instead of four and half those kids have their back turn to you because they're at a table where they can't see the board? How about you do that instead? And see how that turns out?

Sprinkled throughout the conversation were such comments as, "Not four times, not 20 or 30 kids. I got a 160 that I got that I have to meet the needs of. And, you don't do it that way."

Class sizes were not the only perception of unfairness though. When mentioning a consultant that was brought in to discuss improving instruction, Lance shared that he believed he coached the consultant more than he was being coached.

He said, hey can I borrow this? Could I take this? Could I get a copy of this? And I'm thinking, wait a second, you're here to give me stuff, and here I'm giving you stuff. What do you do? And I told him that, you're of no use to me because, I'm telling you how to run your class and your supposed to tell me how to run my class.

Lance looked for balance in roles and responsibilities. He expected class sizes to be balanced. He expected consultants to coach rather than being coached, and he believed that evaluations should remain impartial separate from personal relationships. When these points tipped to imbalance, his perception was that they were unfair.

Lance referred to the consultants brought in to work with multiple teachers including him as a waste of time and money. He felt strongly that he was providing them with more coaching than he was receiving and that as a result, they made his job more cumbersome and difficult than it already was.

> It was a waste of time, but it was a waste of time as far as I was concerned because I had to do a lot of work. He would come in once a month, and I had to do a lot of work to get ready for him, but as far as me knowing what I had to do to score well on those tests, like I said, I gave him stuff. So, I knew more than he did about what I was supposed to be doing.

Lance would have preferred to have never been included in the work with the consultants. "It was a waste of time and money. The guy who came to evaluate me, who was supposed to come in and give me lessons and show me how to do things better, I ended up giving him my assignments." Working with the outside consultants was not the only thing Lance described as a waste. "For me, after a certain number of years in the district, evaluations are a, I don't want to say a waste of time, but they are an exercise in futility." The waste and unfairness described by Lance illustrated the *toxic* Perceptions of Value tinting his viewpoint.

The second interpretive code providing clarity for this theme was *incompetence*. When discussing content knowledge, he identified one assistant principal in particular as quite knowledgeable and competent, but overall, he described most of his supervisors as clueless when it came to observing or evaluating his classroom.

> They were clueless. I'll put it that way. They had absolutely no clue what I did. None. Number one is because, I have a real suspicion that it was way above their heads. I

mean, what I taught, I had other teachers coming in, I don't know what you're doing in here, but how do you get these kids to understand this stuff at this level? I have administrators, they had no clue what I did, how I did it, they just didn't. Mr. Thomas does, and Ms. Reilly really knew. She knew what I did. Mr. Jones and Ms. Smith had less of a clue than the others.

Lance described, when in in his classroom, he felt he was in charge regardless of who came into the room. If an administrator came into his classroom while he was teaching, he expected them to meet his expectations, but he rarely felt that they did.

I pretty much dictated to her, pretty much chapter and verse this is, I had to explain it later, lay it out to her as far as what I did, and how I did it. Same with Ms. Smith. Can I say this on tape? I kicked her out of my room one day. Yep, well, I gave her a choice. I said, you can either close that laptop and be quiet and stop your talking or you can leave, because I don't tolerate it from my students, I'm not going to tolerate it from you. They closed their laptops and walked out. I was perfectly fine with it.

Lance believed that many administrators avoided his classroom, and they intentionally skipped visiting his classroom when doing walkthroughs; they avoided him.

They just never did. As a matter of fact, I got a feeling I know what Ms. Smith told me one time when she was in the room, she said, I didn't like science at all. Well to me that doesn't mean anything, but you know, you don't need to tell me that because I've had a lot of kids that didn't like science that got great grades in my class.

Lance described multiple individual administrators as avoiding him and his classroom, which he chalked up to their inability to understand the content he was teaching the students. "Neither Ms. Smith nor Ms. Henry ever came to me and said what are you doing? What are you covering this week? They just never did." As a result, Lance felt avoided because of their cluelessness, which influenced Lance's perception of incompetence in some of his supervisors.

Lance felt strongly about the toxic, cluelessness presented in his Perceptions of Value. At times, he acknowledged principals who were engaged, present, and positively contributing to his work and practice, but these were mentioned in passing as he would go

into descriptions of the negative experiences that provided most of his point of view. His words were stronger and included more emotion when sharing the *toxic* experiences, and similarly his words were more disappointed and frustrated when describing cluelessness. For Lance, the negative perceptions outweighed any positive experiences he had. Another participant, Aaron, not surprisingly encountered some of the same experiences, but holds a slightly different perception.

Case 6 – Aaron

"Hey, Aaron, just a heads up, it looks like the administrative team is making rounds today," said Mary, an experienced teacher, on her way to the copy room.

"Thanks Mary. I'll be sure to shine my shoes and prepare for inspection," said Aaron, only slightly rolling his eyes and containing his sarcasm. Aaron wondered why so many of his colleagues made such a big deal about classroom observations and walkthroughs – all part of the school dance. The principal shares the cookie cutter goals that everyone will adhere to throughout the year. As the principal or school leaders visit the classrooms, teachers put on a show trying to demonstrate they're hitting the goals by doing things they not normally do every day. The classroom visitors check their boxes on their forms. Then the visitors go back to their offices and email out ratings and scores, and they might put some generic note with it. If our checked boxes aren't the right ones, we're told to put on a better show next time. We're not actually told to put on a better show, but that's what it really means because we don't actually teach the way they want to see it. No one does.

Aaron continues to swim in these thoughts as his third period class begins to file into his room. Aaron stands at the door and greets each student by name as they enter the classroom. For everything in the curriculum, Aaron considers personal interaction to be one

of the most important teaching points. Students do not know how to engage and interact with real people anymore. Students spend so much of their time engaged with a screen or online content, that interacting with real people has become a foreign concept. It doesn't show up on the administrators' checklists either. Most of the concepts emphasized in this class are not on their checklists, fostering creativity, asking questions of the world around them, engaging in thoughtful conversation with peers, the soft skills that are slowly becoming a buzz in the education world but have yet to be defined well enough for checklists. Immediately behind the last student to enter, the assistant principal walks in to join the class. Aaron maintains his smile even as he notices the clipboard and attached checklist in the assistant principal's hand.

Aaron makes an exaggerated point of highlighting different strategies and parts of his lesson that most closely align with what he knows is on the checklist without completely abandoning his intended lesson structure. It's not like the checklist has bad strategies or truly limiting options, but it emphasizes language that is at times awkward in the lesson. It's not natural for the students or Aaron to use, so it always makes the lesson seem artificial and less productive. The assistant principal only stayed through the intro to the lesson. After about seven minutes, he quietly left the classroom, leaving a post-it note on the teacher's desk that read, "good job!" A few minutes later the instructional coach came through with a similar, slightly different checklist, observing lessons. Again, Aaron stayed mostly true to his intended lesson, but put on a show of emphasizing the name of the strategies the students were using to discuss the content. After a few minutes, the coach left, to be followed by the head principal a few moments later. She never came all the way into the classroom. She opened the door, watching from the threshold, typed some notes into her phone, then stepped back and closed the door. Aaron thought she must have downloaded a checklist app and was

trying it out. He then wondered if he would get any useful notes or feedback from any of his visitors today. Just as he had this thought, Mary, the teacher that had told him to expect observations today, walked in carrying a clipboard.

When possible, the principal asks teachers on their plan times to join in the parade of visitors to the different classrooms. Aaron recites an internal prayer of thanks. If nothing else comes of today, he knows he will be able to talk to Mary and receive some useful feedback into his practice. While confident in his teaching, Aaron knows that his lessons could always use some tweaking, and he doesn't always see it in reflection. He can't point to a training, professional development session, or even a conversation that dramatically changed his practice, but he knows that he improved his teaching through experience and conversation with other teachers. Aaron looks forward to his conversation with Mary, knowing she will have more input that, "good job."

Mary hung back as the students filed out of the classroom at the end of the period. She makes dramatic checking gestures on her clipboard as she walks to the front of the room. She and Aaron have talked before about the merry-go-round of visitors on these days and how unproductive they seem. They both smile at the inside joke.

"I have to set up for my next class. Want to talk about your lesson closing at lunch, or do you just want me to add my checklist copies with the others?" Mary said grinning, knowing Aaron's answer.

"I'll see you at lunch," said Aaron.

Classroom Practices. Like Ray's Classroom Practices, Aaron's case includes an interpretive code that could have fit into another participant's teaching practices, but his language focuses on a narrow component of teaching practices that it resulted in a unique

interpretive code, *expertise*. Students of teachers given opportunity to develop mastery of a curriculum rather than attempting to master multiple demonstrate modest statistical gains (Goldhaber, Cowan & Walch, 2013). Aaron emphasized the importance of *expertise* related to curriculum and content knowledge. He felt strongly that he embodied expertise and contributed to the district by having participated in creating curriculum for his content. "I am also teaching a course, the curriculum of which, I've kind of created over the last few years." At the same time, Aaron recognized that his principal expected him to have expertise and to deliver the curriculum as part of his core duties. "From there his expectations are obviously that I'm going to uphold district mandates and curriculum standards with fidelity." A detractor from teachers performing at a high level, according to Aaron, is movement from subject to subject or grade to grade. He states that movement prevents teachers from reaching *expertise*.

> Teacher turnover is too high from going from one subject to another to another, from this grade level to that one, and it never lets teachers get settled in and really master the content their trying to teach, and the same thing's true if we're constantly changing our scope and sequence as a district. It's happened four times in six years.

While recognizing such movement inhibits the development of expertise, Aaron acknowledged that there are times that it is beneficial such as when a teacher moves up to the next grade to stay with the same students.

> You're sacrificing a little bit of expertise for relationship, and relationship, I think, personally, is a bigger key to student growth and success than expertise. So, if you are looping and you're trying stick with your kids to me that makes more sense, but from what I've witnessed, it's more random, and I don't know what's going on to be honest, why people are constantly moving in our district or in our building even. But I've never, one time, had the same team from year to year.

In addition to curriculum and content knowledge, Aaron expressed the benefits of finding schedule balance among classes, contents, and teachers. He described times when certain

contents received fewer students and larger amounts of the instructional day resulting in a

large disparity among staff and student instruction.

> This is the third year we've been on a balanced schedule where everyone gets to teach the same amount of time and then as a result gets the same number of kids. In years past, we had, and I think it's the same for all middle schools, was that you get 80 minutes for math and 80 minutes for ELA and 40 minutes for science and social studies. What that means is, it used to be that we taught 40 minutes if you were social studies and science teacher and you'd have twice the number of kids and sometimes you'd have 40 kids in a classroom and you'd have 180-200 kids in a given school year.

Aaron explained this balance allows teachers in all content areas to know their students and

how to meet their students' needs more effectively.

Unlike Ray, the second interpretive code matches that of multiple participants, as it

broadens back out to capture a variety of concepts related to *student measures*. Aaron

recognized many ways to measure student performance, and a teacher must identify what

they are measuring before determining how.

> So, when I start talking about what ways can you measure your student growth, it depends on what the actual student growth measure is. Is it academic? Is it soft skills? But I should be able to see my kids better able to perform.

The tools he used to measure different types of student gains depended on the types of gains

being targeted. "What are you going to put the focus on? If you're asking me academically,

quizzes and tests and projects and did they meet the criteria. If you're asking me from a

holistic standpoint, it's a lot more." When focusing on academic growth, Aaron pointed out

that not only should a teacher see achievement on individual activities, but over time,

teachers need to look for growth in how the students presented their ideas as well.

> So, you know, quizzes, projects, conversations, generally just work, in general, like what is the product that is being produced, you should see beyond just the straight academic sense, see it in terms of growth, more than did they answer this question correctly or not, but you should see growth in written work.

While Aaron acknowledged teachers should be monitoring student growth in writing or

interaction with content and peers, he also recognized teachers do not necessarily record

every piece of data or observation on every student, as that would be overwhelming. Rather,

according to Aaron, teachers should make mental notes to refer to as they assess student

work to monitor if the individual students make appropriate adequate progress in their

communication of knowledge about the content.

> One of the ways I would evaluate this would be looking at from the beginning of the year to the end of the year, can you follow instructions and how well have you been able to take look at my instructions and follow them with fidelity, right. I'm not taking data on that, but I am getting a sense in my head, seeing what this kid does from the beginning of the year to the end of the year, how often is this kid completing things on time. Am I having to constantly come over to this kid at the beginning of the year and saying, hey, have you gotten this done, x, y, z and so forth? Towards at the end of the year, for the most part maybe, I'm only going over once or twice. All of those things are things, that as an educator, you are assessing on a day-to-day basis, essentially trying to evaluate your kids and see that they're growing, but you're not writing these things down, right.

In addition to individual teachers examining students' test scores or demonstrations of

knowledge, Aaron pointed out the need for teachers to come together to discuss student

needs and the evidence of student growth. This allows teachers to examine student

development more deeply and problem solve how to serve students that may have stalled out.

Additionally, it becomes a point of emphasis to discuss the students as a larger whole and

how their needs might be met so as not to become mired in a single statistic such as an

assessment score.

> We come together and we talk about it like hey I felt like this kid and this kid improved, but these two kids didn't, and here are the things I tried to do throughout the year to address that but I never really saw growth. And maybe he would have this same feedback, or maybe he would differ, and we would try to come up with solution or at least a consensus on how we interpreted the data, right. I feel like that would be the approach. Because if you're simply looking at test scores, well then we're not, we're not really focused on student growth. We're not really focused on educator growth. We're just focused on checking boxes. And what's good academically isn't

225

necessarily so good for the overall growth a kid, or the overall long-term outcomes for that particular student either.

Aaron's **Classroom Practices** pinpoints the need for teacher *expertise* of content, curriculum, students, and their needs. At the same time, *student measures* must include more than a classroom assessment or quiz and should incorporate holistic growth of students in all aspects of their academic life including soft skills such as peer interaction, collaboration, and communication verbally, in writing, and presentation. Aaron described his classrooms' culture as gatherings of students to explore their world and develop a better understanding of their connections to it.

Culture of Evaluation. Aaron's descriptions and experiences of the Culture of Evaluation do not mirror those he described of his classrooms. The interpretive codes providing greater clarity to Aaron's theme match those of multiple other participants. As described during the discussion of student measures, Aaron recognized the importance of teachers and colleagues talking to each other to improve and refine practice, *peer interaction*. Aaron, when asked about talking with others to improve his practice stated he could not recall a life changing, career altering conversation, but over the course of his career he has had many conversations with peers that have contributed to his development incrementally.

> Specifically, I can't recall, like an exact instance, but I do appreciate the opportunity to go and discuss my goals with another person because it always gives me the opportunity to see another point of view and it's one of those things where I use the evaluation process as much as I can to inform my instruction.

More so than speaking with supervisors, Aaron described peers as being those he often seeks feedback from due to their availability, accessibility, and his desire to learn through conversations.

> Well, because I'm always asking people questions, and I'm always going to people and saying hey what do you think about this? Does this make sense to you? So, it's

hard for me to pick out any one single instance or impactful moment because personally I'm doing it so often, and I'm not trying to make myself seem like some, you know, self-improvement guru, but I'm just constantly, I just never feel sure about what I'm doing regardless as to the fact that I know I'm a good educator. I just never feel sure about it. So, I'm always going to people and asking them about what I'm doing and how they feel about it, and really, honestly, I'm one of those people who don't tell me what to think, don't tell me what to do, but your feedback is appreciated, and I'll take as much feedback as you're willing to give. So, that's why it's hard to answer your question. I can't think of anything specific.

In addition to seeking feedback from peers, feedback from *leadership* is an expected part of professional practice and the evaluation process.

Effective leadership includes feedback, and Aaron described the evaluation process present at the school as including a self-assessment questionnaire, teacher selected goals, and the administrator providing feedback related to those goals over the course of the school year.

You also discuss how what you're doing is related to the evaluation scale that they have you fill out every year, the questionnaire, it's like a million and one questions on all the things you think you're doing well in the classroom and not doing well. So, wherever you mark as your personal goals then in those meetings you're taking a look at where you are at applying this or at developing on this and so forth. So the idea is to figure out where you're most needed, where youarea of most needed of improvement is as both assessed by you and your administrator and then you check throughout your year based on those goals.

Aaron stated feedback from observations stems from the identified goals, but it also goes through the administrator's filter of interpretation. "The feedback on those are based on both your administrator's expectations and your administrator's interpretations of said goals." Aaron described the feedback process as being very informal and somewhat superficial. His descriptions do not include depth or impact to practice so much as being able to say a process was followed much like a formality.

Back to those reviews, you know after I go through this, this, and this, here's a process, or here's a bunch of things for you to look at online, check to make sure that you look at my stuff and we'll meet about it later. That's kind of how it typically

goes. If it's an informal, just walkthrough like pop in for five minutes, quick whatever, it's generally a just a quick little online, hey this is what I observed and then you just check to make sure that okay, yeah, I saw that. So, that's kind of like a snapshot.

Feedback stemming from observations, which Aaron described as a very formal process that limits meaningful conversation about teacher practice. He expressed a desire for less formal, more meaningful conversations rather than a formal process.

> I think, just overall, once a week or once bi-weekly informal observation, take a few notes as you walk out the door, what did you see, what did you get out of what I'm doing here and have a conversation and talk about, and less of a cookie cutter fashion of what my goals are because maybe [the evaluation tool] doesn't necessarily encompass or correctly describe my pedagogical goals as an instructor.

Aaron later described how he might alter the evaluation process to be more inclusive of compiled observational data rather than responses after single, brief visits.

> I think the observation process needs to be a continual process that is less formalized, and I think what needs to occur is that perhaps we have a quarterly review or semi-annual review where we review the year's worth or the half year's worth compiled observational data.

Aaron described interactions with supervisors could be more targeted, more impactful rather than to perform what seemed to be assigned performances for display. "I mean at this point in time, especially in this particular district, it's, I'll be coming in this week to take a look for this, this, and this, so make sure I can see that in your classroom." From Aaron's perspective, the Culture of Evaluation lies shallow as *peer interactions* and *leadership* have the potential for large and lasting impact on student growth and teacher implementation, but due to formality and shallow formal conversations the potential remains inert.

I recognize the complaints present in Aaron's descriptions of evaluation practices as rigid and shallow in formality. From my own perspective as an administrator charged with evaluating staff, those years when many staff members must be evaluated according to policy

guidelines differ greatly from those years when only a handful of staff are on cycle for evaluation. When great numbers of teachers must be pushed through the evaluation process, scheduling, timing, meeting all requirements of the process becomes quite time consuming and the casualty becomes the richness of the conversation about practice and improvement. When fewer teachers are identified as on cycle, more time can be given to each to provide richer feedback and additional opportunities for observation and discussion. As typically occurs, in my experience, the number of available evaluators regularly falls short of what is needed to enrich the evaluation process.

Perceptions of Value. With the described culture in mind, it is not a surprise to find the third theme defined as *toxic* and a *ritual.* Formality filled Aaron's descriptions of leadership actions connected to evaluation procedures. When describing how teachers' select their goals for the evaluation process, he felt they were disconnected and represent one form of formality.

> Focus on that as opposed to these cookie cutter goals that don't necessarily align to the personal goals of the educator, that don't necessarily, they're not personal. They're a box to check. That's currently what it feels like. So, I don't know. That kind of gets you, I mean, I think that would be a better process.

On more than one occasion, Aaron described goal setting and the evaluation process as looking the exact same for every person rather than tailored to strengths, weaknesses, or needs. "The cookie cutter process makes it feel less serious, less personal and really, honestly, just makes it feel like a waste of time." Aaron again described the evaluation process as a structure that prevents administrators from helping teachers improve rather than as a tool supporting that improvement.

> Makes it feel like this is just something that I have to do, as opposed to, hey, I'm an administrator who wants to make sure you're meeting your own goals, what are you trying to do? What are you trying to create in your class? Get to understand me, don't

229

just ask me questions and check a box. Because at the end of the day those boxes are never really going to give you a good idea of what I'm trying to create or do.

Ultimately, Aaron described the evaluation process as a tool designed to make sure administrators are paying attention to their staff members and what they may be doing in the classrooms.

> I mean, the purpose of evaluation, I'm assuming is to have some level of, to just have some base level of competence, and I think essentially that's what most of it is to just making sure the person that you have in the classroom isn't completely incompetent at their job. Is it a process that I think is geared towards an individual improvement, not necessarily. I think that's what its goals are, but do I think that's what it's doing, no, not necessarily, so I think it's mostly just there as a way to make sure that administrators aren't missing glaring issues in personnel.

When discussing whether he believed he can influence or change the problems he described, Aaron expressed teachers need to take action at a level higher than building administrators because the decisions selecting what evaluation tools to use and how to use them occur at the school board level.

> We're talking about checking boxes here. Like, okay, you want a four rating system. What the hell does that even mean, right? So, those are the types of things that I feel we need to be focused on trying to impact our policy decisions made at the school board level. That means getting in those meetings. That means e-mailing the people above us, in a nice way, and not in an accusatory way.

In the end Aaron's **Perceptions of Value** aligned almost completely with it being a meaningless ritual that wastes time and creates frustration. Aaron's interpretations of the evaluation ritual contributed to this a *toxic* viewpoint.

Feelings of frustrations contribute to a *toxic* lens. Multiple times Aaron expressed his frustration through word choice, tone, and directly identifying those things causing him to be upset. When explaining his perspective of evaluation goals, he believed there were other ways to know how well or how effective teachers were performing their duties.

> You know there is a lot of things that the evaluation process could be used for as a pulse check, not just on some, pardon my French, some bullshit goal that's been, you know, thrown out there, but as a real legitimate way to see where the people in your building are and how they're performing and how they're feeling and how your policies are affecting them.

He also expressed frustration when talking about staff movement and assignments from year to year, "Just in my own building, witnessing from what I see, we're moving around way too god damn much, pardon my French." Also, class size was a point of anger, which Aaron expressed that he almost resigned when class sizes were remaining too large.

> To whereas before, the average was 35, I think is what I had in each class period. And I was ready to be done at that point. I was literally like, I think the only reason we got that changed was because me and a few other teachers were like, we are gone if this doesn't happen. It was just too much.

Overall, Aaron explained much of his frustration regarded evaluation and expectations from leadership. All teachers receive the same acknowledgement and compensation regardless of effort and outcome. Whether a teacher truly invested and really showed student growth or a teacher just shows up for a paycheck, the system sees them and treats them the same.

> There's no consideration for ability, for effort, for production, for growth, none of that matters. It doesn't matter as you know, it doesn't matter if you're best teacher in the building or the worst teacher in the building, the only thing that matters is how long have you been teaching and were you here when we were able to get our raises or not. So, your effort does not matter, and it takes a lot to get fired. It takes a lot to get fired. So, really, there's no expectation, I mean you can set the expectations, but there is no structure in place to ensure that I'm legitimately doing a good job on a day to day basis, and because there's no incentive structure, and because the evaluation process is what it is because you can't be in there every waking minute to make sure someone's doing a good job, well you know the state of education as it is today.

Aaron continued to express frustration in that the systems in place perpetuate mediocrity even when all of the demand is for improved practice.

> I'm the minority. I'm not saying there's not educators out there that do their best, and I'm not saying there's educators out there that are not doing a good job, but when you don't have an incentive structure put in place for those high achievers, unless those high achievers are really there because they love that specific thing, you're not going

to have high achievers, because why the hell would I put that effort out. I'm one of the few people who do, and I'm sure that you're aware that there's probably five in a building that are willing to do all the above and extra, and that's generous.

While he mostly focused on how the system impacted teacher efforts and performance, Aaron expressed frustration and disappointment, believing the system impactied students in a similar fashion. By forcing the staff into mediocre performance, students and families similarly accept such mediocrity.

> To me, that's a recipe for exactly what we have now, a lot of complacency, a lot of apathy, a lot of depression, a lot of lack of focus, a lot of lack of purpose and a whole generation of kids sitting around wondering what the fuck they're supposed to do with their lives because no one's ever taught them anything that mattered to them. Because all we were focused on was the academic goals, to again, which 30% or less of our kids are ever going to actually need anyways.

Ultimately, Aaron expressed frustration with most aspects of his current setting at one point or another either because they change too much or not changing at all in relation to his beliefs of what would be best for the students and the system.

> Another aspect contributing to his toxic outlook, Aaron explained he had to change how he teaches when being evaluated because the evaluation tool does not accurately capture his methods when interpreted by supervisors.

> So, already, I have to change the way my typical day-to-day lessons are based on the fact that I know I'm being evaluated. So, is it an accurate assessment of how I operate on a day to day based on these goals, no, it's not. Is it an accurate assessment of can I meet these goals, then yes it is.

Educators share Aaron's frustrations across the country as they are repeatedly tasked with accomplishing more while be provided fewer resources (Giroux, 2013).

> Where teachers do enter the debate, they are objects of educational reforms that reduce them to the status of high-level technicians carrying out dictates and objectives decided by experts far removed from the everyday realities of classroom life. (Giroux, 2013, p. 462)

Another area of change holding a toxic perspective for Aaron is the assignment of duties

from year to year. Aaron shared his grade level team changed every year even when the

teachers remain with the school. In addition to colleagues changing grade levels, at times, his

colleagues are assigned to teach different subjects as well. Aaron believed these changes

negatively impact a teacher and a school's performance, and yet they happen, according to

him, every year.

> The teacher ultimately knows what they are able to do. And, you know, (a) you're
> trying to push the teacher outside their box, that's one thing, but it's another thing
> saying, hey, next year, you're going to teach this grade level because it works out
> better numbers that way. Well, okay, what is our ultimate goal. Well it's student
> growth. It's making sure our kids are doing good, not necessarily that we can fill a
> classroom. So, if we're really looking at what's going to be impactful towards the
> best outcomes for our kids, we need to make sure that we're not pushing teachers into
> situations where they're going to be overloaded with work or they're going to be
> outside their expertise if they don't feel comfortable with it.

When asked to change grade levels himself, Aaron stated he refused. He knew many

colleagues had been asked and many had agreed. He was not sure how many had refused, but

he felt he had to refuse in order to maintain his level of quality and expertise as a teacher.

> The shifting. I can't, I can't speak to the conversations an administrator had with
> another educator and their motives and you know all that, so I can't speak to those
> specifics. But I can say from my own. I've been asked, but I firmly said no.

The movement of teachers from content to content and from grade level to grade level clearly

troubled Aaron.

> I've heard conversations where teachers were like, oh, I'm teaching this next year.
> Oh, why did you do that? Oh, because so and so wanted me to do it. It's like, so
> there's that. And then there is like the voluntary moves. But we shift around too
> much. I mean, in the last six years that I've been in this building, I might be one of
> the only teachers who has not started teaching either a new grade level or a new
> subject. I might be one of the only

Aaron believed the changing of teachers between contents and grade levels negatively

impacted student performance as well.

Why can't you guys figure this out. Oh, because you were teaching 8th graders last year and now you're teaching 6th graders, and you're totally forgot that they don't know half the stuff you think they know. Right, so, these are decisions that, I mean, again, I feel like we do it too much now.

In the end, Aaron expressed recognition that for positive change to occur, it begins with teachers knowing and doing what is right for themselves and the students to improve.

We have to take it upon ourselves to improve and we have to take this as our responsibility to improve the system, and until we do that, it's not going to happen, and right now I don't really see that urgency. Some educators have that urgency, but I don't really see the urgency to improve. I don't know how you feel, but I don't see it.

Change and frustration created a toxic point of view. Aaron's **Perceptions of Value** include piles of frustration with formality, change, and a shallow evaluation system and recognition that the evaluation process acts as a hollow ritual unable to contribute productively to the school or its stakeholders.

Cross Case Analysis

Each participant shared their sacred and cover stories and hopefully, their secret stories, as well as perspectives of a shared career providing their unique points of view and interpretation of the events they encountered individually and as a group (Clandinin & Connelly, 2000). Examining each case supported within-case and provided the opportunity to look through that participant's lens. Stepping back and looking at the collection of these individual voices allowed for greater depth and understanding. Tables 3 and 4 highlight individual understanding from the participants, and lay the groundwork for cross case analysis. Table 3 consists of listing each participant's themes, interpretive codes and combined frequency counts that contributed to each. Table 4 provides each interpretive code, the participants that identified with that interpretive code, and the sum of frequencies that contributed to each interpretive code across all participants.

Table 3

Themes and Interpretive Codes for Individual Cases

Participant	Theme	Interpretive Code	Frequency
Eric	Classroom Practices	Teaching Practices	24
Eric	Classroom Practices	Student Measures	34
Eric	Culture	Supervision	68
Eric	Culture	Leadership	71
Eric	Perceptions	Development	39
Eric	Perceptions	Challenging	36
Paul	Classroom Practices	Teaching Practices	10
Paul	Classroom Practices	Student Measures	32
Paul	Culture	Peer Interaction	21
Paul	Culture	Leadership	54
Paul	Perceptions	Power	43
Paul	Perceptions	Appraisal	39
Ray	Classroom Practices	Strategies	32
Ray	Classroom Practices	Assessment	23
Ray	Culture	Leadership	40
Ray	Culture	Peer Interaction	63
Ray	Perceptions	Standardized	49
Ray	Perceptions	Toxic	98
Nancy	Classroom Practices	Teaching Practices	43
Nancy	Classroom Practices	Student Measures	43
Nancy	Culture	Peer Interaction	75
Nancy	Culture	Leadership	85
Nancy	Perceptions	Toxic	74
Nancy	Perceptions	Incompetence	47
Lance	Classroom Practices	Teaching Practices	88
Lance	Classroom Practices	Student Measures	39
Lance	Culture	Peer Interaction	91
Lance	Culture	Leadership	61
Lance	Perceptions	Toxic	45
Lance	Perceptions	Incompetence	43
Aaron	Classroom Practices	Student Measures	80
Aaron	Classroom Practices	Expertise	32
Aaron	Culture	Peer Interaction	37
Aaron	Culture	Leadership	38
Aaron	Perceptions	Toxic	60
Aaron	Perceptions	Ritual	35

Table 4

Themes and Interpretive Codes for Cross Case Analysis

Interpretive code	Participants	Combined Frequency
Appraisal	Paul	39
Assessment	Ray	23
Challenging	Eric	36
Development	Eric	39
Expertise	Aaron	32
Incompetence	Nancy, Lance	90
Leadership	Eric, Paul, Ray, Nancy, Lance, Aaron	349
Peer Interaction	Paul, Ray, Nancy, Lance, Aaron	287
Power	Paul	43
Ritual	Aaron	35
Standardized	Ray	49
Strategies	Ray	32
Student Measures	Eric, Paul, Nancy, Lance, Aaron	228
Supervision	Eric	68
Teaching Practices	Eric, Paul, Nancy, Lance	165
Toxic	Ray, Nancy, Lance, Aaron	277

Summary of Findings

As I consider the phenomenon from the perspective of cross case analysis, I imagine a campfire at night. The fire represents shared experiences related to the phenomena with different variables such as wind, ash, feelings and past encounters associated with campfire experiences, temperature of the evening, smoke, and distance from the fire linking to the variables of the actual shared experiences. Even though each participant is present and experiencing the same campfire, each also continues to have a unique experience in relation to the campfire based on the multitude of variables. In speaking to each person, I gain greater understanding of their campfire experience and what the fire illuminated around him or her, but by talking to all of them, I gain a greater understanding of the entire evening and what the fire illuminated for all of them

Three themes clarified in the light of the shared experience, and all of them were bound by the themes found during the document analysis. Classroom Practices included those events and activities associated with the typical classroom. Four of the participants shared both interpretive codes, a fifth participant matched one of the interpretive codes as can be seen in table 4. This fifth participant's second interpretive code and both of the sixth participant's interpretive codes reflected more narrowly defined components of the interpretive codes presented by the other participants noticeable in table 3. From the broad perspective, all six participants focused on similar concepts and experiences in the classrooms including teaching strategies, instruction, assessment, professional development, and other forms of measuring student learning. While multiple themes gleaned from document analysis could connect, the strongest links tie to Evaluation Process and Teacher Preparation and Performance. The Evaluation Process includes the observation of Classroom Practices in action while Teacher Preparation and Performance involves the teachers' experiences, pedagogy and continued professional development.

The theme of Culture of Evaluation presented, in some regards, even greater alignment among the participants. Table 4 also highlights six of the participants represented the same interpretive code, and five of those six aligned with the same second interpretive code. The lone participant that deviated comparatively doubled down on the first interpretive code by demonstrating more specific language to one aspect of that initial interpretive code as noted in table 3. The Culture of Evaluation most closely aligned with School Culture from the document analysis. A close examination of the interpretive codes and their examples from the data draw comparisons to leadership, peer support, interaction among colleagues,

responsibility and many other descriptors of the school setting and how participants interact within it.

The theme of Perceptions of Value highlighted the greatest deviation among interpretive codes and invested focus of the participants. Four participants shared the same interpretive code and two of those four shared another interpretive code noted in table 4. Beyond that, each participant had unique interpretive codes to illustrate their Perceptions of Value, which can be seen in table 3. From document analysis, the most closely aligned theme was Performance Rating. Both capture the unique, individual knowledge and capability present in each participant.

Stepping back and looking at the scene of the campfire from a distance, it amazes that such similarity in experiences, presence, and interaction with the phenomenon can generate such a variety of perceptions and realities. Each participant stepped up to the campfire experience with their own expectations and understanding of what the experience would be and what would be gained from it, and as a result, each participant walks away from the shared experience with their own perceptions, values, and beliefs about the experience.

Conclusion

Education reform efforts and federal policy changes demand increases in the quality of teachers, leading to the adoption of teacher evaluation tools that incorporate student growth measures as a significant component of a teacher's job performance rating (ESSA, 2015; USDE, 2010, Thomsen, 2014). By examining teachers' perceptions of the inclusion of student growth measures in summative evaluations, this study provided a glimpse into the viewpoint of some of those individuals most affected by such changes. As a whole, each recognized the necessity of performance evaluation as well as connection it may have to

student performance. At the same time, each participant illustrated how past evaluation systems have failed and their expectation that any new adopted practices would do the same. Most of them had ideas and suggestions for how to improve their practice connected with evaluation, yet none of them aligned with the evaluation tool and processes adopted by their district and school. For this reason, this study provides a valuable resource to those faced with making such decisions moving forward as it captures and presents the voice of those who have not been heard.

Chapter five will answer the research questions, illuminate the discussion of the study's place in the field of related research, and further research needed to explore the topic.

CHAPTER 5

SYNTHESIS

This study opened with a series of questions regarding how you might know if someone effectively performs their duties. The questions and examples began broadly referring to any profession, but they eventually narrowed, addressing how a principal might know if students have learned what they need to learn from their teacher. At the core of my study remains the question, how do we know if a teacher effectively teaches his or her students.

Chapter one introduced the context of the study. Educational reform efforts prompted changes in teacher evaluation practices to include student growth measures as a significant part of performance evaluation. Chapter two explored the literature underlying the study with a close examination of the history of teacher evaluation, leadership practices associated with teacher evaluation, teacher power, and the nature of student growth measures. Chapter three outlined the process for this narrative case study including acquisition of participants and analysis of data. Chapter four detailed the findings of the study including detailed analysis of each case and a cross case analysis. Finally, chapter five provides responses to the research questions, implications of the findings, future research needed and reflections of my journey.

Answering the Research Questions

The primary research question, identified in order to bring clarity to teachers' perspectives and understanding related to performance evaluation, centralizes a topic of conversation currently contained in educational reform efforts. The primary question being: How do teachers perceive the use of student growth measures in performance evaluations for core content teachers at an urban mid-western middle school? Three sub-questions provide

guidance for analysis: (1) What do teachers understand about the context of job performance evaluation? (2) How do teachers define student growth when measuring their own performance? (3) How do teachers describe performance evaluation? Beginning with the sub-questions, each will be answered by examining the evidence provided by the individual participants across multiple data sources, and then it will be discussed collectively looking at the evidence across cases. Once the sub-questions have been analyzed, the primary research question will be discussed in a similar fashion.

What do teachers understand about the context of job performance evaluation?

Each participant demonstrated a basic understanding of the school's performance evaluation process, which included goal setting, classroom observations, pre and post conferences about those observations, and a summative evaluation, which is submitted to human resources. Many of the participants varied on some of the details such as the number of expected classroom observations and the vocabulary used to describe different parts of the evaluation process. Greater variance occurred when describing the purpose and role of performance evaluation. The State Department of Education (SDE) described effective teacher evaluation systems as improving instructional quality and promoting student academic growth (2019). Part of the theme Performance Rating from the document analysis, every district in the state has been required to adopt an evaluation protocol that includes a minimum of four possible ratings for teachers related to their performance. Also, the collection of observational data contributes to one part of the summary rating, while student performance data is supposed to be included as the other part in the teacher's summary performance rating, but at this time, this school district has not been implementing that practice (State Department of Education, 2019).

For Eric, performance evaluation was very cut and dry. Administrators evaluated teachers according to the adopted tool as part of the interpretive code *supervision*. He recognized that administrators could only catch a glimpse during such observations, and he believed his administrators recognized that fact as well. "It is supposed to be an improvement process. You have to be able take criticism. Nobody's perfect. It doesn't matter if you are on the side of the person being evaluated or the person doing the evaluating." Paul, within the code *appraisal*, felt that many aspects may be useful at times, but most points of performance an administrator may be looking for do not occur on a daily basis. "It's just, when you know an administrator is coming in you need to make sure that they will see that happen in those 10-15 minutes that they will be in there." Paul knew the process would occur, he just does not believe anything useful will come from it. Consistency and fairness hold value for Ray as he believes tying financial incentive or reward into the evaluation process would result in an increase in favoritism or preferential treatment based on relationships. "I don't know that I would make any changes in the process as long as it's being done legitimate." Ultimately, Ray expressed the belief that evaluations provide evidence for administrators to show they have worked with teachers even when that work has not resulted in improved teacher performance highlighted in the code *toxic*. "Like I said, some admins, not all of them, some think its just to get documentation because they have to present it to their superior, but it should be to help them get better, to help teachers improve." Nancy would agree the evaluation process and purpose ring hollow from her experience. "Well, I think the purpose is to make sure that we have good teachers in the classroom and they're doing what they're supposed to, and yes, I did use air quotes." The air quotes mentioned were around 'good

teacher' and 'supposed to.' Her experiences leaned towards evaluations being more relationship based than ability based shown in her examples of *toxic* perceptions.

Lance, like Nancy, felt strongly about whether or not his evaluators knew his content. Outside of one assistant principal and a few others he described as okay, Lance spoke very poorly of his supervisors ability to understand the work he did with students let alone evaluate or provide feedback on it as demonstrated in the interpretive code *incompetence*. "They were clueless. I'll put it that way. They had absolutely no clue what I did. None. Number one is because, I have a real suspicion, that it was way above their heads." Without the appropriate credentials or respect, Lance dismissed evaluation as a waste of time. Aaron examined evaluation with a more balanced perspective. While he recognized the need to put on a show for those that observe him, he also recognized that a teacher should be able to flex into that performance when they know they are being observed exemplifying the code *ritual*. "So, is it an accurate assessment of how I operate on a day to day based on these goals, no, it's not. Is it an accurate assessment of can I meet these goals, then yes it is." Aaron recognized potential value in evaluation to improve teacher practice, but he believed it needed to be less formal and more continual.

While the district begins adopting new practices to meet expectations, the teachers of Hawkins Middle School illustrate their lack of faith in the evaluation process as a whole. Eric, alone among these participants, viewed the evaluation process as a straightforward tool and expectation, the remaining participants painted the performance evaluation process as hollow, flawed, ineptly administered and predominantly pointless. Aaron recognized evaluation could be a productive tool if it was radically changed, but the changes he desired do not align with the adopted state requirements.

How do teachers define student growth when measuring their own performance?

New state requirements connected to performance evaluation include student performance measures. While Hilltop School District has not incorporated this practice yet, it is likely to occur in the next few years as the district continues to align with state expectations. SDE's documents outline the use of at least three student data points contributing to the calculation of a final summative rating included in the interpretive code *score* (2019). Districts have the autonomy to select which data points to incorporate as well as to define levels of student performance. These student growth measures in addition to the summary rating generated through the observation protocols equals the final summative rating.

While some of the participants acknowledged standardized assessments as one tool to measure student growth, consistently, the participants defined their own performance through other tools. In discussion of code *student measures*, Eric communicated his belief portfolios of student work provide an accurate demonstration of student growth and contribute evidence to the work he and the student have accomplished.

> I have students pick two or three things to put in a paper portfolio every nine week period so they can reflect back to the beginning of the period and compare what I did in the third nine weeks with what I did in the first nine weeks.

Eric also highlighted talking to his students, asking questions, gradual release of responsibility, and asking students to do work on the white board at the front of the classroom in front of peers.

Paul emphasized the need to expand beyond limited time fields and suggested looking at comparisons of student work across multiple years as a portion of his discussion of the code *student measures*. "So, I mean, writing samples, not necessarily even just, beginning of the

year to end of the year, but if you could get school wide, 6th grade writing samples and then, now their 8th graders, show them how far they have come." Paul stated he believes student work provides powerful evidence of student performance, but he thinks administrators incapable of processing and understanding the student work as a reflection of teacher performance.

Ray over the course of his career moved from basing all grades of students on standardized unit tests to believing they fail to provide evidence of student learning at all. He expressed strongly that such exams illustrate what students do not know rather than what they have gained in knowledge and skill. Ray's interpretive code deviated from the other participant's in connection to measuring student progress; his was much more narrow in scope as it focused on *assessment* rather than the other participant's *student measures*. He concedes formal, standardized assessments will remain a large part of classroom instruction and measuring performance, but he recognizes the need to monitor student conversation, interaction with each other and the content to more accurately gauge student acquisition of knowledge and skill.

Nancy, also, accepted the role one-shot, high-stakes assessments currently play in the field of education. Her theme of Classroom Practices provided evidence of her belief in pre and post tests provided more accurate information, but she also admitted when time begins to become limited the pre-test becomes the first item cut from the lesson plan making it a null proposition to use as a growth measure. "I rely mostly on my teacher observations as far as my student growth. Walking around, asking questions, looking at their work, seeing what they, within the talking of the group, if they are truly understanding it or not." Much like

Ray, Nancy gauges student understanding on their interactions with the content rather than a formal assessment tool.

Lance holds an equal aversion to formalized or unit tests. "If I ask a kid a question and they know the answer, that's how I judged whether or not they knew something. Test wise, not unless I made it up myself, and we went over it the day before." He expressed multiple times as captured in the discussion of his *student measures*, not holding faith in formal assessments or standardized tests. Describing the need to compare student performance across time, Lance focused on daily assignments and comparing them from the start of the school year to the work completed at the end. "…this is what you did the first day, this is what your doing now, regardless of what they did on a test."

Aaron's theme of Classroom Practices shared the same sentiment; student work from the start of the school year to the end of the school year provides a clearer image of student growth and teacher contribution. He also acknowledges teachers must prepare students beyond the content such as incorporating soft skills such as presentation, communication, effort, and collaboration, and he expressed the belief teachers should monitor student growth in these areas as well. "I should be able to see my kids better able to perform."

I must admit, when reflecting on this research question in particular, at the very beginning of my program, designing the study, I naively believed teachers would align to a theory or concept of measuring student growth and teacher growth in accordance with their content such as ELA teachers wishing to measure with writing samples over time, science teachers requiring student projects and experiments evaluated with rubrics, math teachers incorporating a computational formula to calculate growth, and so on. Over the course of my research, this notion became broken up, loosened from my bias. In examining teacher

responses from multiple contents, even as a small sample size, it clearly illustrates teachers

wish to measure students and be measured by their work and efforts in the way that best

represents their growth. For many of the participants this included some capturing of student

knowledge and skill at the start of the school year and comparing it to a similar sample at the

end of the year. Paul suggested the samples occur across multiple years. Lance and Nancy

expressed the time frame could be condensed to multiple smaller chunks with pre and post

assessments wrapped around curriculum units or nine week sections of time. All of the

participants included measuring some form of subjective, teacher judgment of knowledge

acquisition based on talking to students, questioning them, and listening to them talk to each

other as they engaged in the content.

How do teachers describe performance evaluation?

From the documents provided by the school district, specific timelines and deadlines

establish when different events and stages of the evaluation process occur providing a clear

picture of the theme from document analysis, Evaluation Process. Additionally, the district

provided training schedules and presentation materials from administrator trainings on the

topic of teacher evaluation outlining expectations, purpose, strategies, and communication

structures to incorporate. Consistently, across all participants, the submission deadlines are

the only aspect of their experiences aligned with the district materials. Throughout

participant narratives and descriptions, the performance evaluation process devolves into an

inaccurate, incompetently managed, bureaucratic farce.

Paul expressed frustration throughout the entire process. Unless, he knows someone

is coming to observe him related to the evaluation process, he does not take it into

consideration in his daily practice. Paul described the evaluation process as lacking utility

and occurs more as performing for a single person audience rather than effectively determining a teacher's quality or ability resulting in a sense of powerlessness captured in Paul's theme Perceptions of Value. "It's like you're getting ready for this one big moment and that doesn't seem very productive to me." Similarly, Ray put little value in the evaluation process seeing it as more reflective of people's personalities than of performance. "I think it's just a personal thing if you have responsibilities or if you have pride in your job, you're going to do it whether they're looking over your shoulder or not." Ray described goal setting as having to put something down for them to look for when they come into the classroom, and it gave an example of the *standardized* process evaluation was viewed as. At its best Ray viewed evaluation has something that occurred separate from the work of teaching students.

Nancy's Perceptions of Value included a description of one of her more negative experiences as humiliating. "Her evaluations were incredibly unprofessional." Nancy went on to describe those performance evaluations as being full of sarcasm and reflective of unwarranted criticism. Nancy also repeatedly noted the competency of those who were supposed to evaluate her leading to the development of *incompetence* as an interpretive code. She remarked often that they did not know her content or even know enough about teaching that she could use their feedback. "Because to be quite honest, he couldn't tell me that because he didn't know my content." For Nancy, performance evaluation adds to the list of things she has to do even though it does not provide a benefit for her or her students. Aaron described an alternative process of ongoing conversations prompting thinking, reflection, and improved practice rather than judged performances. "I think, essentially, that's what most of it is, to just making sure the person that you have in the classroom isn't completely

248

incompetent at their job." With his perspective, Aaron set the bar quite low for the purpose of performance evaluation describing it as a *ritual*.

Eric held a different point of view of performance evaluation, for the most part accepting it at face value and focusing on those points of performance he could influence. When Eric encountered evaluations from administrators that overreached their observations or did not communicate what they would be scoring him on, he self-advocated. Eric expressed confidence speaking for himself when encountering such situations. He described his supervisors as being clear and in alignment with expressed expectations that resulted in a fair summative evaluation. "I think it's clearly laid out in front of us and makes it easy to achieve those expectations that are in front of us." While Eric did not care for the multiple tools and forms required throughout the performance evaluation process, his frustration captured in the code *challenges,* he expressed acceptance of its role in his work.

Like, Eric, Lance accepted performance evaluation as part of his regular practice and work provided it came from someone incredibly familiar with his content and teaching practices. If not, then Lance quickly aligned with many other participants describing the entire process as a waste of time and effort highlighting his connection to the code *incompetence*. Lance spoke at length, very positively, of one assistant principal that knew his content, came to his class regularly, and often discussed his work with him in a way that he felt was useful and productive. "We would go through what I was doing in class, and any problems I was having, anything I needed as far as material. It was really good for me."

Strong emotions entwined throughout the participants descriptions of performance evaluation. Nancy's expressed caused her to choke up and take a moment to reset herself. Aaron raised his voice and began punching the air with a clenched fist. Lance began striking

the table with a pointed finger to accentuate his points. Ray emphasized his points with longer pauses between statements so they could sink in. Eric took longer pauses before responding to prompts in order to collect the correct words, and Paul's voice became stern and terse when on this topic. At many other points, in each of the participants' interviews, they laughed, made light of some of events in their narratives, spoke with a smile, enjoying the opportunity to speak of their experiences. Unfailingly, even those participants that accepted performance evaluation as a necessary part of their work spoke differently when describing their experiences directly with evaluation.

How do teachers perceive the use of student growth measures in performance evaluations for core content teachers at an urban mid-western middle school?

Currently, student growth measures remain separate from performance evaluation in the Hilltop School District, but as the district aligns with state expectations, change is likely. Up to this point in time, participants painted a bleak picture of evaluation experiences. A *toxic* lens for many precludes any discussion of evaluation practices and at best, these practices represent a *standardized ritual* with many *challenges*. The addition of *student measures* the participants do not include as accurate measures of student learning does not endear an already shaky evaluation process. Eric accepts the evaluation tool as a snapshot of his practice (*supervision*), but just as he challenged his supervisor's comments, he may challenge the validity of the assessments selected for inclusion in the process. Paul already views the evaluation process as a fabricated piece of performance art, the data just becomes another piece of the set dressing (*appraisal*). Ray blatantly stated he believes administrators would manipulate class rosters to favor those teachers in the good graces of administration and to apply pressure to those who are not (*toxic*). Nancy expressed a belief class rosters are biased (*toxic*), and because of that any system incorporating their data would be skewed

250

(*incompetence*). Also, her traumatic experience when an administrator rewarded some teachers based on their scores with biased rosters caused her to express she would leave the district and maybe the profession if they system moved that direction. Lance stated that any evaluation from someone other than a master of his content is meaningless to him regardless of whether his students' data was a contributing factor (*incompetence*). If a personnel decision occurred because of low scores, he would be ready to leave.

Aaron focused on the potential of such a system, but for it to work, according to him, it would require a quasi-experimental setup where a sampling of students were selected and monitored throughout the school year by the teacher and the principal with adjustments in instruction occurring along the way (*expertise*). Each participant believed the incorporation of student work and performance could provide insight into teacher performance (*student measures*). Eric suggested portfolios spanning the course of the school year with a variety of samples of work. Nancy recommended pre and post test results along with daily work and exit slips. Ray invites the use of general trend data based on daily assignments and quizzes over short periods of time to monitor retention. Lance also recommended the use of daily assignments connected directly to the content and across modalities as created by the teacher. Finally, Paul recommended a sampling of pre and post test results with writing samples as well as looking over a combination of short periods of time and very larger periods of time that could include multiple years.

Implications of Findings and Recommendation

Every child should have access to effective teachers and principals, and in attempts to ensure this, legislation has been adopted in most states to incorporate student achievement measures into summative evaluations of teachers and principals as well as create recruitment

251

and reward opportunities for filling difficult educational assignments and generating students' success (USDE, 2010). Rather than focusing on evaluation of teachers, the emphasis should be on the development of teachers. Principals consistently demonstrate poor skills and limited capacity to have difficult learning conversations with teachers and parents, thus limiting their effectiveness to promote a positive learning climate (Le Fevre & Robinson, 2014). Substantial effects occur to student learning when teachers receive one-to-one coaching to improve instruction (Biancarosa, Bryk, & Dexter, 2010). By incorporating the characteristics detailed by Blasé and Blasé (2004) such as participating in instructional conferences, promoting professional growth, and encouraging critical reflection, principals can directly influence the quality of instruction.

Effective evaluation skills impact the process very much, becoming a coaching conversation; the teacher leaves the conversation with recognition of successes as well as next steps to take to continue to improve (Mathews, 2017; Raymond, 2017). Through the course of classroom observations, the principal will note trends and patterns needed to be addressed through professional development on a larger scale. To this end, the instructional leader generates or enlists the support of others to provide professional development learning opportunities for the staff either during in-service hours or outside the duty day (Coggshall, Rasmussen, Colton, Milton & Jacques, 2012; DuFour & Mattos, 2013). This becomes a constant, repetitious cycle of observation, conversation, support, data collection, analysis, action in delivery of professional learning for larger groups of staff members (Abdo, 2017; Mette, Range, Anderson, Hyidston & Nieuwenhuizen, 2015). Through the course of this work, the administrator may recognize individual needs or more extreme trends in needs for development. Whereas these may not be addressed in whole staff development, it still falls to

the leader to address the needs of every teacher's growth and learning to more readily meet the needs of every student in those classes. Through this cycle, the principal directly affects instructional delivery resulting in indirect influence in student achievement due to improved instruction (Clifford, Behrstock-Sherratt & Fetters, 2012).

How this might look in day-to-day practice includes a variety of actions. First, the principal must be present where instruction is occurring, in the classrooms (Ovando & Ramirez). Only by being in the classrooms, can the principal know for certain what the instruction looks like, levels of student engagement generated, and assess the rigor of the work expected of students (Grissom, Loeb & Master, 2013; Seashore Louis, Dretzke & Wahlstrom, 2010). Following classroom visits and observations, which should be long enough for a principal to have a confidence in understanding of what instruction is happening, the principal and teacher should hold a conference to speak of the lesson and student learning. This conference should not be an attempt lay down judgment resulting in raising teacher anxiety and limiting cognition, understanding and development. Rather, the purpose is to prompt critical reflection within the teacher, in order for the teacher to recognize changes needed to improve practice.

As the participants in the study shared, previous encounters with evaluation focused on formality, judgment, or personality conflicts. A contributing factor to the similarities in experiences of the participants could stem from shared racial and cultural backgrounds, so recommendations based on these results may be skewed to serving those of the same racial and cultural backgrounds and provides opportunities for further research to be discussed later in this chapter. Rather than holding the teachers accountable by test scores, administrators should be held accountable for teacher development (Coggshall, Rasmussen, Colton, Milton

& Jacques, 2012; Porter et al., 2010; Sebastian & Allensworth, 2012). Collecting data or evidence related to teacher development may increase student achievement more than incorporating it into teacher evaluation. A principal's focus should be on teacher development, and in order to do so, principal's must improve their instructional leadership and practices including coaching conversations, developing professional development to meet the needs of teachers and students, participate in repeated cycles of coaching conversations and create job embedded professional development opportunities for teachers.

Future Research

This study provided thick, rich description of the phenomenon of teacher evaluation for six core content middle school teachers in a district approaching the implementation of including student growth measures as part of teachers' summative performance evaluation. The scope and sampling for this study were quite narrow. The initial wave of education reform requiring the adoption of this evaluation practice (NCLB, RttT) has reduced to a strong suggestion rather than the previous requirement (ESSA) (USDE, 2015). While the federal wave of change pulls back, the changes to state statutes and policies remain largely in effect (SDE, 2019). So even though the initial push is gone, the lasting effects remain in the form of state requirements of student growth measures in teacher evaluation unless a new wave of change occurs pushing these practices out of action. Until such an event occurs, future research should focus on how these policy adoptions affect the educational community through studying a variety of topics through their connection to evaluation practices:

- Perspectives of teachers of different races, cultures, and backgrounds
- Teacher attrition and retention
- School climate and culture
- Student performance at various levels
- Student performance post-secondary
- Principal attrition and retention

254

- Teacher development
- Teacher recruitment
- Fidelity and accuracy of implementation
- Longitudinal impacts
- Influence on diverse populations of students
- Parent perceptions

Areas for potential future research remain broad in scope. Building from this study locally, follow the school district's adoption of student growth measures in evaluation, following up with these participants and open lanes of understanding as they go through the transition. Additionally, exploring the perspectives of participants within the same district but across multiple grade levels such as elementary and high school could add unique lens from which to explore the phenomenon. Another avenue to explore could be to incorporate, with fidelity, the instructional observation protocol already a part of the evaluation process with the inclusion of teacher identified student data in comparison to the district and state selected data to determine potential differences in results.

Final Reflections

I began my journey six years ago, exploring topics, settling on teacher performance evaluation, convinced I already knew what the results would be at the conclusion. At the time, the adoption of student growth measures began to be discussed in relation to recent legislation, and not surprisingly, the SDE adopted the practice while providing an ample amount of time for districts to adjust policies and adopt appropriate evaluation tools to be in compliance with the new state policies. My belief at the time, rooted in value-added metrics, district data, and my own experiences, aligned with the new policies, incorporating student data into teacher evaluation would be more accurate.

My own experiences as a teacher up to this point mirrored those of the participants of this study, multiple poor experiences with administrators writing evaluations without visiting

the classroom, observation feedback based more on relationship with the supervisor than the lesson observed, or no feedback at all. As an administrator myself, I strived to be accurate, timely, supportive, and productive while encountering administrative colleagues often slouching into the bad actions that upset so many teachers. I must be honest, there were times, when overwhelmed by the volume of evaluations to completed by a deadline, I am sure some of my teachers chalked up my performance summaries to their own resume of poor evaluation experiences.

As I dug more deeply into the research, repeatedly I discovered that regardless of the metrics incorporated into evaluation ratings, a large portion of the evaluation remains the same, often failing the observation-feedback cycle that becomes so bogged down and cumbersome in time and effort. Equally, for every study I could find that illustrated the potential to accurately disaggregate teacher performance from student measures, there were multiple others that illustrated poor computational formulas or ineffective collection procedures resulting in little to no correlation between teacher ratings and student performance. All systems at some point relied on the same individuals that should be leading the observation-feedback cycles; there were no ways around that component, resulting in continued ineffective practices.

Developing the literature review, I improved in reading and understanding a variety of forms, methods, and models of research. As I dug into qualitative studies, I considered the different theoretical frameworks from which the researcher pursued the topic and contemplated how other methods might provide additional insight. I began investigating how different lenses such as heuristic, narratological, or ethnography might alter a root structure

such as case study to provide unique insight into the phenomenon. Narrowing my own selection down to narrative case study, I was ready to plan and develop my study.

Once I was able to engage the participants of the study, they shared their sacred stories, cover stories, and some, possibly shared their secret stories as well. Their experiences and mine were the same, the same frustrations, the same fears, and the same concerns. Then I recognized my bias, and I stall. I asked myself if I was guarding against injecting myself into the study too much. I used the strategies identified including reflexivity, member checking, and a peer support to monitor myself. I maintained a journal and personal memos to capture my thinking for further reflection.

Analyzing the data, I was hesitant. How do I make sure I find all the answers buried in the massive collection of raw data? Going into the process, I feared the narrative inquiry would be the most difficult and the thematic analysis would flow more easily, but as I spent more time with the data, I found the narratives more naturally compelling, more intuitive and inviting while the coding and organization of the discrete components for multiple cycles of coding seemed at times like I was forcing topics or concepts together. I would have to step back and allow myself to mentally reset to look for the natural connections between the concepts shared by the participants. The follow up connections with participants were rewarding, to receive their confirmation of the transcripts and re-storied narratives or to dig back into the original conversation for more clarity.

Finally, it was time to put the findings together and report out new understandings. Whereas the analysis portion felt like I drowned in puzzles, the reporting stage felt more like a storyteller, sharing a fable with those he cares about. This topic that rang close to my own secret stories, created an opportunity for me to explore the participants and my own sacred

stories, cover stories, and secret stories. I feel like I have shared an important part of my professional and personal life as a researcher reporting the results of the study. I am both exhilarated by the idea of doing so again and intimidated about investing in such a commitment that can consume such large portions of my life.

In summary, I believe I started like many first time researchers, idealistic, like my work will change the world. As my understanding and capability increased, so did my perspective. While some studies may have the power to change the world, most research, studies the changes in the world, trying to find understanding and purpose behind them.

APPENDIX A

PARTICIPANT SURVEY

Current building
Years at current building
Years in district
Current content/subject taught
Other content/subjects taught in previous years
Race/Ethnicity
Languages spoken other than English
Do you live in the school's attendance area?
Do you live in the district's attendance area?
Were you on evaluation cycle this school year?
- If yes, did the principal or an assistant principal conduct the evaluation?
- If no, when was your last evaluation cycle?
  - Did the principal or assistant principal conduct your last evaluation?

Briefly describe your district's/school's evaluation process.
Briefly describe your most recent evaluation process and/or other evaluation processes you have encountered.
How many evaluation cycles have you participated in while at your current school?
How many evaluation cycles have you participated in while in your current district?
How many evaluation cycles have you participated in during your career as an educator?
Using one of the following descriptors, please respond to the following statements: Strongly Disagree, Disagree, Agree, Strongly Agree
- Evaluation is an effective method to improve teaching
- Evaluations accurately capture daily teaching practices
- The evaluation process is followed to fidelity at my school
- Student Growth Measures (high stakes assessments) accurately reflect teacher performance
- Student Growth Measures would be an accurate tool to act as a base for performance based pay or merit pay
- Evaluation results would be an accurate tool to act as a base for performance based pay or merit pay

Would you be willing to participate in a follow-up interview to provide more information about your evaluation experiences during your career? If yes, please provide your name, e-mail address, and phone number for future contact.

INTERVIEW GUIDE

- How do teachers perceive the use of student growth measures in evaluations used for personnel decisions for core content teachers at an urban midwestern middle school?

  o What do teachers understand about the context of job performance evaluation?

    ▪ Describer your role and responsibilities.

      • Any other responsibilities associated with your core duties?

    ▪ What expectations does your supervisor hold related to your duties and responsibilities?

      • What variations or differences are there in these expectations from teacher to teacher, if any?

    ▪ Describe your district/school's evaluation process.

    ▪ What is the relationship between the evaluation process and your ability to meet your supervisor's expectations?

    ▪ What is the relationship between the evaluation process and your ability to perform your duties and fulfill your responsibilities?

    ▪ What changes, if any, would you make in the evaluation process to strengthen alignment between your duties, responsibilities, expectations, performance, and evaluation?

  o In what ways are job performance evaluation described?

    ▪ Describe an experience you have had with job performance evaluation in the field of education.

      • Any others? (repeat up to 3 times)

- How would you describe your role in these processes?

- How would you describe the purpose of those evaluations?

- What other purposes could they have served?

- How have any previous evaluation experiences influenced your performance or practices?

o How do teachers define student growth when measuring their own performance?

- What ways can you measure student growth?

  - Any others specific to your academic content?

- Which of these do you believe to be accurate and measurable?

- What evidence could be collected to demonstrate student growth?

- What might a teacher do if the evidence did not demonstrate student growth?

- What might a supervisor do if the evidence did not demonstrate student growth?

- If evidence of student growth in your content was a part of your performance evaluation, what should that look like?

  - What would be the most important aspects of this inclusion?

    o Why would these take priority?

o How do teachers describe personnel decisions?

- How are decisions regarding teacher assignments and duties made?

- How are building assignments and class loads assigned?

- How are salaries, stipend, and other compensation determined?

- What opportunities are available for advancement or recognition?

  - How could you obtain these if you desired?

- What other personnel decisions have you encountered in your experiences?

- What changes might you make to how these decisions are made?

- What do you feel you could influence regarding any of these decisions, if you wished?

  - How might you do so?

APPENDIX C

NARRATIVE PROMPTS

There will be four narrative prompts. For each, all or some of the additional detail prompts may be incorporated to capture thick, rich description of the events being discussed.

Primary prompts (4)

- Please describe the events of your most recent evaluation experience.
- Please describe the events of your earliest evaluation experience.
- Please describe the events of another memorable or impactful evaluation experience.
- Please describe the events of any other evaluation experiences you are willing to share.

Detail prompts to be used if necessary to generate thick, rich description of the events.

- Can you describe where the events took place (i.e. classroom, office, school…)?
- Can you describe other individuals directly involved in the experience?
- What was your relationship or prior experiences with the persons involved?
- What else was occurring in your life at work and/or at home during this time?
- Following the experience, did any changes occur (in the classroom, relationships, or school)?
- What thoughts/feelings did you experience during this interaction?
- How does this experience relate to other similar experiences?

REFERENCES

Abdo, H., (2017). *Lessons learned from a comprehensive teacher evaluation system: An instrumental case study* (Doctoral dissertation). Rutgers University, New Brunswick, NJ.

Adams, D. W. (1995). *Education for extinction: American Indians and the boarding school experience (1875-1928).* Lawrence, KS: University Press of Kansas.

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, *39*(1), 54-76.

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS EVAAS) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives*, *20*(12). Retrieved from http://epaa.asu.edu/ojs/article/view/1096

Anderson, J. E. (1988). *The education of blacks in the south, 1860-1935*. Chapel Hill, NC: The University North Carolina Press.

Anzia, S. F., & Moe, T. M. (2014). Collective bargaining, transfer rights, and disadvantaged schools. *Educational Evaluation and Policy Analysis*, *36*(1), 83-111.

Arp, J. S. B. (2012). Case studies of teacher satisfaction with three plans for evaluation and supervision. (Doctoral dissertation). The University of Alabama, Tuscaloosa, AL. Retrieved from https://ir.ua.edu/handle/123456789/1677

Ayres, L., Kavanaugh, K., & Knafl, K. A. (2003). Within-case and across-case approaches to qualitative data analysis. *Qualitative Health Research*, *13*(6), 871-883.

Auriacombe, C. J., & Schurink, E. (2012). Conceptualising qualitative research through a spiral of meaning-making. *Administratio Publica*, *20*(3), 144-169.

Babchuk, W. A., & Badiee, M. (2011). Realizing the potential of qualitative designs in education: Guidelines for research and practice. In *Proceedings of the 29th Annual Midwest Research Research-to-Practice Conference in Adult, Continuing and Community Education*, 25-31.

Baird, T., & Heinen, E. (2015). Elevating teacher voice: Democracy, political action, and professional engagement. *Education in a Democracy*, *7,*135-154.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., ... & Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers. EPI Briefing Paper# 278. *Economic Policy Institute*.

Baker, B.D., Oluwole, J., & Green, P.C. III (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the race-to-the-top era. *Education Policy Analysis Archives, 21*(5). Retrieved from http://epaa.asu.edu/ojs/article/view/1298

Ballou, D., & Springer, M. G., (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher 44*(2), 77-86.

Baron, E. J. (2018). *The effect of teachers' unions on student achievement: Evidence from Wisconsin's Act 10* (Doctoral dissertation). Florida State University, Tallahassee, FL.

Beach, J. M. (2007). The ideology of the American dream: Two competing philosophies in education, 1776-2006. *Educational Studies*, *41*(2), 148-164.

Bell, L. T. (2018). *Predicting teacher attitudes toward education reform initiatives* (Doctoral

    dissertation). Lake Forest College, Lake Forest, IL.

Belmont Report (1979). *The Belmont Report: Ethical principles and guidelines for the*

    protection of human subjects of research. Retrieved from

    hhs.gov/ohrp/humansubjects/guidance/belmont.html

Biber, B. (1958). Problems of values and measures in evaluation of teaching. *Educational*

    Leadership, 15(4), 213-217.

Bieler, D., Holmes, S., & Wolfe, E. W. (2017). Patterns in the initial teaching assignments of

    secondary English teachers: Implications for teacher agency and retention. *The New*

    Educator, *13*(1), 22-40.

Blase, J., & Blase, J. (1999). Principals' instructional leadership and teacher development:

    Teachers' perspectives. *Educational Administration Quarterly*, *35*(3), 349-378.

Blazar, D. (2015). Grade assignments and the teacher pipeline: A low-cost lever to improve

    student achievement? *Educational Researcher*, *44*(4), 213-227.

Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. (2011). Qualitative research in

    management: A decade of progress. *Journal of Management Studies*, *48*(8), 1866-

    1891.

Blumberg, A. (1985). Where we came from: Notes on supervision in the 1840's. *Journal of*

    Curriculum and Supervision, 1(1), 56-65.

Blumberg, A., & Jonas, R. S. (1987). Permitting access: The teacher's control over

    supervision. *Supervision in Context. Educational Leadership, 44*(8), 58-62.

Boeije (2010). *Analysis in qualitative research*. Thousand Oaks, CA: Sage Publications.

Bogdan, R., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theories and methods*. Boston, MA: Pearson A & B.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27-40.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, *48*(2), 303-333.

Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). Examining District Guidance to Schools on Teacher Evaluation Policies in the Midwest Region. Issues & Answers. REL 2007-No. 030. *Regional Educational Laboratory Midwest*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/edlabs.

Braslow, D. A. (2016). *Teachers' perspectives on evaluation: Exploring the effects of new evaluation policies on teaching* (Doctoral dissertation). Harvard University, Cambridge, MA. Retrieved from http://nrs.harvard.edu/urn-3:HUL.InstRepos:33797240

Brunner, E., Hyman, J., & Ju, A. (2018). *School finance reforms, teachers' unions, and the allocation of school resources* (No. 2018-11). *Department of Economics Working Paper Series*, University of Connecticut, Hartford, CT.

Burch, P., Steinberg, M., & Donovan, J. (2007). Supplemental educational services and NCLB: Policy assumptions, market practices, emerging issues. *Educational Evaluation and Policy Analysis*, *29*(2), 115-133.

Bush, C. D. (2017). Teacher perceptions about new evaluation model implementations (Doctoral dissertation), Northcentral University, Scottsdale, AZ.

Byford, A. L. (2018). Teacher and administrator perceptions of the teacher excellence and support system (TESS) in Arkansas (Doctoral dissertation). University of Arkansas, Fayetteville, AR.

Campbell, T. F. (2013). Teacher supervision and evaluation: A case study of administrators' and teachers' perceptions of mini observations (Doctoral dissertation). Northeastern University, Boston, MA.

Carbaugh, B. G., Marzano, R. J., & Toth, M. D. (2015). *School leadership for results*. West Palm Beach, FL: Learning Sciences International.

Carlson, J. A. (2010) Avoiding traps in member checking. *The Qualitative Report*, *15*(5), 1102-1113.

Case, K. H. (2016). *Teacher perceptions of the impact of the North Carolina teacher evaluation process on teacher effectiveness, professional growth, and attitudes toward teaching* (Doctoral dissertation). Gardner-Webb University, Boiling Springs, NC.

Cawelti, G., & Reavis, C. (1980). How well are we providing instructional improvement services? *Educational Leadership, 38*(3), 236-240.

Clandinin, D. J. (2013). *Engaging in narrative inquiry*. Walnut Creek, CA: Left Coast Press.

Clandinin, D. J., & Connelly, F. M. (1988). Studying teachers' knowledge of classrooms: Collaborative research, ethics, and the negotiation of narrative. *The Journal of Educational Thought (JET)*, *22*(2A), 269-282.

Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research*. San Francisco: Jossey-Bass Publishers.

Clayton, C. (2013). Understanding current reforms to evaluate teachers: A literature review on teacher evaluation across the career span (Doctoral dissertation). Pace University, New York, NY.

Colby, S. A., Bradshaw, L. K., & Joyner, R. L. (2002). *Teacher evaluation: A review of the literature. Proceedings from Annual Meeting of the American Educational Research Association*, New Orleans, LA.

Coleman, E. (1945). The "supervisory visit." *Educational Leadership*, *2*(4), 164-167.

Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, *116*(1), 1-32.

Connally, K., & Tooley, M. (2016). Beyond ratings: Re-envisioning state teacher evaluation systems as tools for professional growth. *New America*. Washington DC.

Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher*, *19*(5), 2-14. doi:10.3102/0013189x019005002

Connelly, F. M., & Clandinin, D. J. (1999). *Shaping a professional identity: Stories of educational practice*. New York: Teachers College Press.

Corcoran, S. P. (2010). Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice. Education Policy for Action Series. *Annenberg Institute for School Reform.* Brown University, Providence, RI.

Coulson, A. J. (2010). The effects of teachers unions on American education. *Cato Journal*, *30*(1), 155-170.

Cowen, J. M., & Fowles, J. (2013). Same contract, different day? An analysis of teacher

bargaining agreements in Louisville since 1979. *Teachers College Record*, *115*(5), 1-

30.

Cowen, J. M., & Strunk, K. O. (2015). The impact of teachers' unions on educational

outcomes: What we know and what we need to learn. *Economics of Education*

Review, *48*, 208-223.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five*

approaches (3rd ed.). Los Angeles, CA: SAGE.

Creswell, J. W.; Hanson, W. E.; Plano Clark, V. L. & Morales, A. (2007). Qualitative

research designs: Selection and implementation. *The Counseling Psychologist, 35*(2),

236-264.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing*

among five approaches. Los Angeles, CA: Sage publications.

Crowley, G. R., & Beaulier, S. A. (2018). Public-sector unions and government policy:

Reexamining the effects of political contributions and collective bargaining

rights. *Public Finance Review*, *46*(3), 454-485.

Cuban, L. (1988). *The managerial imperative and the practice of leadership in schools*.

Albany, NY: Suny Press.

Cubberley, E. P. (1916). *Public school administration: A statement of the fundamental*

principles underlying the organization and administration of public education.

Houghton Mifflin Company.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*.

Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, *68*(4), 35-39.

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

Darling-Hammond, L. (2010). Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching. *Center for American Progress*, Washington DC.

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

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8-15.

Davis, D. R., Ellett, C. D., & Annunziata, J. (2002). Teacher evaluation, leadership and learning organizations. *Journal of Personnel Evaluation in Education*, *16*(4), 287-301.

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, *34*(2), 267-297.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research*. Los Angeles, CA: Sage.

Dewey, J. (1903). Democracy in education. *The elementary school teacher*, *4*(4), 193-204.

Dewey, J. (1938). *Experience and education*. New York: Macmillan.

Donaldson, M. L. (2012). Teachers' perspectives on evaluation reform. *Center for American Progress*. Washington DC.

Donaldson, M. L., & Papay, J. P. (2012). Reforming teacher evaluation: One district's story. *Center for American Progress*. Washington DC

Doty, J. E. (2018). Teacher performance evaluation and professional growth in the era of "educator effectiveness" in Maine. University of Maine, Orono, ME.

Drake, T. A., Goldring, E., Grissom, J. A., Cannata, M., Neumerski, C., Rubin, M., & Schuermann, P. (2015). Development or dismissal? Exploring principals' use of teacher effectiveness data. *Improving teacher evaluation systems: Making the most of multiple measures*, 116-130.

Duffett, A., Farkas, S., Rothertham, A. J., & Silva, E. (2008). Waiting to be won over: Teachers speak on the profession, unions, and reform. *Education Sector*, Washington DC.

Elam, N. P. (2017). *The impact of the Ohio teacher evaluation system on principals' approaches and perceptions toward evaluation* (Doctoral dissertation). Miami University, Miami, FL.

Ellingson, L. L. (2009). *Engaging crystallization in qualitative research: An introduction*. Los Angeles, CA: Sage.

Ellingson, L. L. (2014). "The truth must dazzle gradually" Enriching relationship research using a crystallization framework. *Journal of Social and Personal relationships*, *31*(4), 442-450.

ESSA (2015). Every Student Succeeds Act of 2015, pub. l. no. 114-95 § 114 stat. 1177 (2015-2016).

Fairman, J., & Mette, I. (2017). *Working toward implementation of performance evaluation and professional growth (PE/PG) systems in Maine school districts*. University of Maine, Maine Education Policy Research Institute.

Fielding, J., Fielding, N., & Hughes, G. (2013). Opening up open-ended survey data using qualitative software. *Quality & Quantity*, 47(6), 3261-3276.

Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of public economics*, *91*(5), 901-914.

Finnegan, R. S. (2016). *Teacher and principal perceptions of a new evaluation program for teachers*. College of William and Mary, Williamsburg, VA.

Flavin, P., & Hartney, M. T. (2015). When government subsidizes its own: Collective bargaining laws as agents of political mobilization. *American Journal of Political Science*, *59*(4), 896-911.

Forte, E. (2010). Examining the assumptions underlying the NCLB federal accountability policy on school improvement. *Educational Psychologist*, *45*(2), 76-88.

Fowles, J., & Cowen, J. (2015). In the union now: Understanding public sector union membership. *Administration & Society*, *47*(5), 574-595.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. New York: McGraw-Hill Humanities/Social Sciences/Languages.

Fuhrman, S. H. (2010). Tying teacher evaluation to student achievement. *Education Week*, *29*(28), 32-33.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco: Jossey-Bass

Fullan, M. (2006). Leading professional learning: Think ''system'' and not ''individual school'' if the goal is to fundamentally change the culture of schools. *School Administrator, 63*(10), 10–15.

Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York: Teachers College Press.

Frandsen, B. R. (2016). The effects of collective bargaining rights on public employee compensation: Evidence from teachers, firefighters, and police. *ILR Review*, *69*(1), 84-112.

Fryer Jr, R. G., Levitt, S. D., List, J., & Sadoff, S. (2012). *Enhancing the efficacy of teacher incentives through loss aversion: A field experiment* (No. w18237). National Bureau of Economic Research. Washington DC.

Giroux, H. A. (2013). Neoliberalism's war against teachers in dark times. *Cultural Studies? Critical Methodologies*, *13*(6), 458-468. DOI:10.1177/1532708613503769

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British dental journal*, *204*(6), 291.

Glazerman, S., Chiang, H., Wellington, A., Constantine, J., & Player, D. (2011). Impacts of performance pay under the teacher incentive fund: Study design report. *Mathematica Policy Research, Inc.* Washington DC.

Glazerman, S., & Seifullah, A. (2012). An evaluation of the Chicago teacher advancement program (Chicago TAP) after four years. Final report. *Mathematica Policy Research, Inc.* Washington DC.

Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A

research synthesis. *National Comprehensive Center for Teacher Quality*. Washington

DC.

Goe, L., Holdheide, L., & Miller, T. (2014). Practical guide to designing comprehensive

teacher evaluation systems: A tool to assist in the development of teacher evaluation

systems. *Center on Great Teachers and Leaders*. Washington DC.

Goldhaber, D., (2015). Exploring the potential of value-added performance measures to

affect the quality of the teacher workforce. *Educational Researcher 44*(2), 77-86.

Goldhaber, D., Cowan, J., & Walch, J. (2013). Is a good elementary teacher always good?

Assessing teacher performance estimates across subjects. *Economics of Education

Review*, *36*, 216-228.

Goldhaber, D., Lavery, L., & Theobald, R. (2016). Inconvenient truth? Do collective

bargaining agreements help explain the mobility of teachers within school

districts?. *Journal of Policy Analysis and Management*, *35*(4), 848-880.

Goldhaber, D., & Walch, J. (2012). Strategic pay reform: A student outcomes-based

evaluation of Denver's ProComp teacher pay initiative. *Economics of Education

Review*, *31*(6), 1067-1083.

Goldstein, D. (2014). *The teacher wars: A history of America's most embattled profession*.

New York: Anchor Books.

Goldstein, J., & Noguera, P. A. (2006). A thoughtful approach to teacher evaluation.

Educational Leadership, *63*(6), 31.

Grbich, C. (2013). *Qualitative data analysis: An introduction*. London: SAGE Publications.

Gregoire, S. A. (2009). *Effectiveness of principals as evaluators of teachers*. University of

    Minnesota, Minneapolis, MN.

Grimmett, P. P. (1981). Clinical supervision and teacher thought processes. *Canadian*

    Journal of Education/Revue canadienne de l'education, *6*(4), 23-39.

Guthery, S. (2018). The influence of teacher unionization on educational outcomes: A

    summarization of the research, popular methodologies and gaps in the literature. *The*

    William & Mary Educational Review, *5*(1), 14.

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school

    improvement: An analysis of the evidence. *Educational Assessment, Evaluation and*

    Accountability, *26*(1), 5-28.

Han, E. S. (2015). *The myth of unions' overprotection of bad teachers: Evidence from the*

    district-teacher matched panel data on teacher turnover.1Wellesley College,

    Wellesley, MA.

Hardin-Bartley, S. L. (2014). *Strategic management of human Capital: A critique of urban*

    education human resources policies and procedures. University of Missouri-St.

    Louis, St. Louis, MO.

Harmsen, R., Helms-Lorenz, M., Maulana, R., & van Veen, K. (2018). The relationship

    between beginning teachers' stress causes, stress responses, teaching behaviour and

    attrition. *Teachers and Teaching*, *24*(6), 626-643.

Harper, B. L. (2016). *A qualitative study of principal perceptions of performance evaluation*

    in Ohio (Doctoral dissertation). Kent State University, Kent, OH.

Harris, D. N., & Herrington, C. D., (2015). The use of teacher value-added measures in

    schools: New evidence, unanswered questions, and future prospects. *Educational*

    Researcher 44(2), 71-76.

Harrison, H., Birks, M., Franklin, R., & Mills, J. (2017). Case study research: Foundations

    and methodological orientations. *Qualitative Social Research, 18*(1).

    doi:http://dx.doi.org/10.17169/fqs-18.1.2655

Harrell, M., & Bradley, M. (2009). *Data collection methods. Semi-structured interviews and*

    focus groups. Rand National Defense Research Institute. Santa Monica, CA.

Hartney, M., & Flavin, P. (2011). From the schoolhouse to the statehouse: Teacher union

    political activism and US state education reform policy. *State Politics & Policy*

    Quarterly, *11*(3), 251-268.

Hayes, D. G., & Wood, C. (2011). Infusing qualitative traditions in counseling research

    designs. *Journal of Counseling & Development*, *89*(3), 288-295.

Hazi, H. M., & Arredondo Rucinski, D. (2009). Teacher evaluation as a policy target for

    improved student learning: A fifty-state review of statute and regulatory action since

    NCLB. *Education Policy Analysis Archives/Archivos Analíticos de Políticas*

    Educativas, *17*(5).

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.

Hiatt, D. B. (1994). Parent involvement in American public schools: An historical

    perspective 1642-1994. *The School Community Journal*, *4*(2), 27-38.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating

    teacher value-added scores. *American Educational Research Journal*, *48*(3), 794-831.

Holley, K., & Colyar, J. (2012). Under construction: How narrative elements shape

    qualitative research. *Theory Into Practice*, *51*(2), 114-121.

Hosford, P. L. (1984). *Using what we know about teaching*. Association for Supervision and

    Curriculum Development: Alexandria, VA.

Hunter, M. (1980). Six types of supervisory conferences. *Educational Leadership, 37*(5),

    408-12.

Ingle, W. K., Willis, C., & Fritz, J. (2015). Collective bargaining agreement provisions in the

    wake of Ohio teacher evaluation system legislation. *Educational Policy*, *29*(1), 18-50.

Jacoby, D. (2011). Teacher unionization in school governance. *Educational policy*, *25*(5),

    762-783.

Jansen, H. (2010). The logic of qualitative survey research and its position in the field of

    social research methods. *Qualitative Social Research 11*(2).

Johnson, S. M., (2015). Will VAMS reinforce the walls of the egg-crate school?. *Educational*

    Researcher 44(2), 117-126.

Jung, D. I., & Avolio, B. J. (1999). Effects of leadership style and followers' cultural

    orientation on performance in group and individual task conditions. *Academy of*

    Management Journal, *42*(2), 208-218.

Kafka, J. (2009). The principalship in historical perspective. *Peabody Journal of Education*,

    84(3), 318-330.

Kantor, H., & Lowe, R. (2004). Reflections on history and quality education. *Educational*

    Researcher, 33(5), 6-10.

Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, *45*(1), 34-70.

Kimball, S. M., White, B., & Milanowski, A. T. (2003). Preliminary analysis of criterion-related validity of the teacher performance evaluation system in Washoe County. *Consortium on Policy Research in Education*, Madison, WI.

Kent, S. I. (2001). Supervision of student teachers: Practices of cooperating teachers prepared in a clinical supervision course. *Journal of Curriculum and Supervision*, *16*(3), 228-44.

Klingelhutz, J. O. (2017). *Perceived effects of peer observation on collaboration between teachers at a Minnesota high school.* St. Cloud State University, St. Cloud, MN.

Kober, N. (2007). Why we still need public schools: Public education for the common good. *Center on Education Policy*. Washington DC.

Koedel, C., & Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and policy*, *6*(1), 18-42.

Koski, W. S. (2012). Teacher collective bargaining, teacher quality, and the teacher quality gap: Toward a policy analytic framework. *Harvard Law & Policy Review*, *6*, 67-90.

Koski, W. S., & Horng, E. L. (2014). Keeping sight of the forest through the trees: Response to "Collective bargaining, transfer rights, and disadvantaged schools". *Educational Evaluation and Policy Analysis*, *36*(1), 112-119.

Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as

evaluators? A case study of principals' views and experiences. *Educational*

Administration Quarterly, *52*(5), 711-753.

Kramer, M. F. (2016). *Teacher And administrator perspectives on new york state's teacher*

evaluation system. University of New England, Biddeford, ME.

Kress, S., Zechmann, S., & Schmitten, J. M. (2011). When performance matters: The past,

present, and future of consequential accountability in public education. *Harvard*

Journal on Legislation, *48*, 185-234.

Labaree, D. F. (2011). Consuming the public school. *Educational Theory*, *61*(4), 381-394.

Ladd, H. F. (2011). Teachers' perceptions of their working conditions how predictive of

planned and actual teacher movement? *Educational Evaluation and Policy Analysis*,

33(2), 235-261.

Lee, L. (2011). What did the teachers think? Teachers' responses to the use of value-added

modeling as a tool for evaluating teacher effectiveness. *Journal of Urban Learning,*

Teaching, and Research, *7*, 97-103.

Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a conception of how school leadership

influences student learning. *Educational Administration Quarterly*, *46*(5), 671-706.

Lerch, R. D. (1980). The clinical model: The optimum approach to supervision. *The Clearing*

House, 53(5), 238-240.

Levinson, M., & Theisen-Homer, V. (2015). No justice, no teachers: Theorizing less-unjust

teacher firings in Los Angeles Unified. *Theory and Research in Education*, *13*(2),

139-154.

Lewis, H., & Leps, J. M. (1946) When principals supervise. *Educational Leadership. 3*(4), 160-163.

Liang, G., & Akiba, M. (2015). Teacher evaluation, performance-related pay, and constructivist instruction. *Educational Policy, 29*(2), 375-401.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: Sage publications

Marshall, C., & Rossman, G. B. (2014). *Designing qualitative research*. Thousand Oaks, CA: Sage publications.

Marshall, K. (2012). Fine-tuning teacher evaluation. *Educational Leadership, 70*(3), 50-53.

Marzano, R. J. (2012). Teacher evaluation. *Educational Leadership, 70*(3), 14-19.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: ASCD.

Marzano, R. J., & Toth, M. (2013). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement*. Alexandria, VA: ASCD.

Mathers, C., & Oliva, M. (2008). Improving instruction through effective teacher evaluation: Options for states and districts. TQ Research & Policy Brief. *National comprehensive center for teacher quality*. Washington DC.

Mathews, C. (2017). *Use of teacher evaluation for improving instruction: a mixed methods investigation in high-poverty New Jersey charter high schools* (Doctoral dissertation). Rutgers University, New Brunswick, NJ.

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). Thousand Oaks, CA: Sage Publications.

McHaney, J. H., & Impey, W. D. (1992). Strategies for analyzing and evaluating teaching

    effectiveness using a clinical supervision model. In Proceedings of *Annual Meeting of*

    the Mid-South Educational Research Association. Knoxville, TN.

Merriam, S. & Tisdell, E. (2015). *Qualitative research: A guide to design and*

    implementation (4th edition). San Francisco, CA: Josey-Bass.

Mette, I. M., Range, B. G., Anderson, J., Hvidston, D. J., & Nieuwenhuizen, L. (2015).

    Teachers' perceptions of teacher supervision and evaluation: A reflection of school

    improvement practices in the age of reform. *Education Leadership Review*, *16*(1), 16-

    30.

Mette, I. M., Range, B. G., Anderson, J., Hvidston, D. J., Nieuwenhuizen, L., & Doty, J.

    (2017). The wicked problem of the intersection between supervision and evaluation.

    International Electronic Journal of Elementary Education, *9*(3), 709-724.

Milanowski, A. (2011a). Strategic measures of teacher performance. *Phi Delta Kappan*,

    92(7), 19-25.

Milanowski, A. T. (2011b). Validity research on teacher evaluation systems based on the

    framework for teaching. *Consortium for Policy Research in Education*. Madison, WI.

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods*

    sourcebook. London: SAGE Publications.

Morgan, G. B., Hodge, K. J., Trepinksi, T. M., & Anderson, L. W. (2014). The stability of

    teacher performance and effectiveness: Implications for policies concerning teacher

    evaluation. *Education Policy Analysis Archives*, *2*(95). Retrieved from

    http://dx.doi.org/10.14507/epaa.v22n95.2014

Moore, B. (2009). Emotional intelligence for school administrators: a priority for school reform? *American Secondary Education*, *37*(3), 20-28.

Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation: The case of the missing clothes? *Educational Researcher*, *42*(6), 349-354.

Nelson, J. A. (2012). *Effects of teacher evaluations on teacher effectiveness and student achievement* (Doctoral dissertation). North Michigan University, Marquette, MI.

Nixon, A., Packard, A. L., & Dam, M. (2011). School principals and teacher contract non-renewal. *International Journal of Educational Leadership Preparation*, *6*(3), 3-20.

Odden, A., & Kelly, J. A. (2008). Strategic management of human capital in public education. *Consortium for Policy Research in Education.* Madison, WI.

Olsen, B., & Sexton, D. (2009). Threat rigidity, school reform, and how teachers view their work inside current education policy contexts. *American Educational Research Journal*, *46*(1), 9-44.

Ovando, M. N., & Ramirez Jr, A. (2007). Principals' instructional leadership within a teacher performance appraisal system: Enhancing students' academic success. *Journal of Personnel Evaluation in Education*, *20*(1-2), 85-110.

Paglayan, A. S. (2018). Public-sector unions and the size of government. *American Journal of Political Science 62*(4).

Paige, M. (2012). Using VAM in high-stakes employment decisions. *Phi Delta Kappan*, *94*(3), 29-32.

Paige, M. (2013). Applying the paradox theory: A law and policy analysis of collective bargaining rights and teacher evaluation reform form selected states. *Brigham Young Education & Law Journal*, *2013*(1), 21-45.

Palys, T., & Atchison, C. (2012). Qualitative research in the digital era: Obstacles and

opportunities. *International Journal of Qualitative Methods*, *11*(4), 352-367.

Papay, J. (2012). Refocusing the debate: Assessing the purposes and tools of teacher

evaluation. *Harvard Educational Review*, *82*(1), 123-141.

Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Thousand

Oaks, CA: Sage Publications.

Pazey, B. L., Heilig, J. V., Cole, H. A., & Sumbera, M. (2015). The more things change, the

more they stay the same: Comparing special education students' experiences of

accountability reform across two decades. *The Urban Review*, *47*(3), 365-392.

Pogodzinski, B., & Jones, N. (2014). Exploring novice teachers' attitudes and behaviors

regarding teacher unionism. *Educational Policy*, *28*(4), 491-515.

Popham, J., & DeSander, M. (2014). Will the courts save teachers? *Educational Leadership*,

71(5), 55-58.

Porter, A. C., Linn, R. L., & Trimble, C. S. (2005). The effects of state decisions about

NCLB adequate yearly progress targets. *Educational Measurement: Issues and

Practice*, *24*(4), 32-39.

Raymond, J. K. (2017). *Evaluating one public school district's teacher evaluation program

and its implementation: A qualitative case study*. Saint Mary's University of

Minnesota, Winona, MN.

Reavis, C. A. (1978). Research in review/clinical supervision: A review of the research.

Educational Leadership, 35(7), 580-4.

Reinhorn, S. K. (2015). *Teachers' views of school-based professional learning in six high-

performing, high-poverty, urban schools* (Doctoral dissertation). Harvard University,

Cambridge, MA. Retrieved from http://nrs.harvard.edu/urn-
3:HUL.InstRepos:16461036

Reinhorn, S. K., Johnson, S. M., & Simon, N. S. (2016). Investing in development: Six high-performing, high-poverty schools implement the Massachusetts teacher evaluation policy. *Educational Evaluation and Policy Analysis*, *39*(3), 383-406.

Reissman, C. K. (2008). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage Publications.

Richardson, L., & St Pierre, E. (2008). A method of inquiry. *Collecting and interpreting qualitative materials* N. K. Denzin (Ed). Los Angeles: Sage Publications.

Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational administration quarterly*, *44*(5), 635-674.

Rockoff, J. E., & Speroni, C. (2011). Subjective and objective evaluations of teacher effectiveness: Evidence from New York City. *Labour Economics*, *18*(5), 687-696.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4-36.

Ross, J. A., & Gray, P. (2006). Transformational leadership and teacher commitment to organizational values: The mediating effects of collective teacher efficacy. *School Effectiveness and School Improvement*, *17*(2), 179-199.

Rousmaniere, K. (2007). Presidential address: Go to the principal's office: Toward a social history of the school principal in North America. *History of Education Quarterly*, *47*(1), 1-22.

Riordan, J., Lacireno-Paquet, N., Shakman, K., Bocala, C., & Chang, Q. (2015). Redesigning

   teacher evaluation: Lessons from a pilot implementation. REL 2015-030. *Regional*

   Educational Laboratory Northeast & Islands. Washington DC.

Rubinstein, S. A., & McCarthy, J. E. (2016). Union–management partnerships, teacher

   collaboration, and student performance. *ILR Review*, *69*(5), 1114-1132.

Saldaña, J. (2013). *The coding manual for qualitative researchers.* London: SAGE

   publications.

Simons, Helen (2009). *Case study research in practice*. Los Angeles, CA: Sage.

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation in*

   Chicago: Lessons learned from classroom observations, principal-teacher

   conferences, and district implementation. Research report. Consortium on Chicago

   School Research.

Taylor, F. W. (1914). *The principles of scientific management*. Harper.

Satty, P. (2016). *Pay for my performance: teachers' experiences with reform-based*

   evaluation tools in a large suburban school district. Northeastern University, Boston

   MA.

Schafer, W. D., Lissitz, R. W., Zhu, X., Zhang, Y., Hou, X., & Li, Y. (2012). Evaluating

   teachers and schools using student growth models. *Practical Assessment, Research &*

   Evaluation, *17*(17).

Shinkfield, A. J., & Stufflebeam, D. L. (2012). *Teacher evaluation: Guide to effective*

   practice. Massachusetts: Kluwer Academic Publishers.

Shugart, K. (2017). *The pedagogical influences of a value-added model evaluation system from the perspectives of elementary school teachers in north Georgia: A phenomenological study*. Liberty University, Lynchburg, VA.

Smith, A. (2015). *Negative impacts of teacher evaluation reform: The role of federal policy in student achievement and teacher quality*. University of Nevada, Las Vegas, NV.

Smith, W. C. (2015). Framing the debate over teacher unions. *Mid-Atlantic Education Review*, *1*(1), 17-26.

Smith, A., & Spatariu, A. (2015). Student achievement and teacher evaluation: A review of past and current government policy reform and scholarly research with an emphasis on multicultural educational implications. *Journal of Education Policy, Planning, and Administration*, *5*(9), 1.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of curriculum studies*, *36*(1), 3-34.

Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education review*, *27*(5), 556-563.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks: Sage Publications.

Stake, R. E. (2010). *Qualitative research: Studying how things work*. Guilford Press.

State Department of Education. (2014).

State Department of Education. (2017).

State Department of Education (2019).

Steinberg, M. P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, *38*(2), 293-317.

Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's excellence in teaching project. *Education Finance and Policy*. Washington DC.

Stewart, Alison (2014). Case study. In Jane Mills & Melanie Birks (Eds.), *Qualitative methodology: A practical guide* (pp.145-159). Thousand Oaks, CA: Sage.

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

Strunk, K. O., & Grissom, J. A. (2010). Do strong unions shape district policies? Collective bargaining, teacher contract restrictiveness, and the political power of teachers' unions. *Educational Evaluation and Policy Analysis*, *32*(3), 389-406.

Swearingen, M. (1946). Looking at supervision. *Educational Leadership, 3*(4), 146-151.

Superfine, B. M., & Gottlieb, J. J. (2014). Teacher evaluation and collective bargaining: The new frontier of civil rights. *Michigan State Law Review*, 737.

Taylor, F. W. (1914). *The principles of scientific management*. Harper.

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching. *Education Next*, *12*(4), 10.

Thattai, D. (2001). A history of public education in the United States. *Journal of Literacy and Education in Developing Societies*, *1*(2), 2001-11.

The New Teacher Project (August 2007). *Hiring, assignment, and transfer in Chicago public schools.* Retrieved from https://tntp.org/assets/documents/TNTPAnalysis-Chicago.pdf

Thompson, E. (1952). So begins-so ends the supervisor's day. *Educational Leadership, 10*(2), 80-84.

Thomsen, J. (2014). Teacher performance growing role in employment decisions. Teacher

      tenure: Trends in state laws. *Education Commission of the States*. Washington DC.

Toch, T., & Rothman, R. (2008). Rush to judgment. *Washington: Education Sector*, 479-508.

Tracy, S. J. (1995). How historical concepts of supervision relate to supervisory practices

      today. *The Clearing House, 68*(5), 320-325.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative

      research. *Qualitative inquiry*, *16*(10), 837-851.

Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy

      beliefs of novice and experienced teachers. *Teaching and teacher Education*, *23*(6),

      944-956.

Turner, D. W. (2010). Qualitative interview design: A practical guide for novice

      investigators. *The qualitative report*, *15*(3), 754.

United States Department of Education. (2010). *A blueprint for reform: The Reauthorization

      of the Elementary and Secondary Education Act*. Washington, DC: U.S. Dept. of

      Education, Office of Planning, Evaluation and Policy Development.

U.S. Department of Education. (2015). Every Student Succeeds Act (ESSA). Retrieved from

      https://www.ed.gov/essa.

Vardas, M. (2014). *The influence of federally mandated policy on educators* (Doctoral

      dissertation) Northeastern University, Boston, MA.

Wade, M. (2015). *Labor's last stand? The great recession and public sector collective

      bargaining reform in the American states.* University of Minnesota, Minneapolis,

      MN.

Waldron, N. L., & McLeskey, J. (2010). Establishing a collaborative school culture through comprehensive school reform. *Journal of Educational and Psychological Consultation*, *20*(1), 58-74.

Weingarten, R. (2010). A new path forward: Four approaches to quality teaching and better schools. *American Educator*, *34*(1), 36-39.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *New Teacher Project*. Washington DC.

Wetzel, W. A. (1929). Scientific supervision and curriculum-building. *The School Review, 37*(2), 117-123.

Whitehead, M. (1952). Teachers look at supervision. *Educational Leadership, 10*(2), 1011-1106.

Winkler, A. M., Scull, J., & Zeehandelaar, D. (2012). How strong are US teacher unions? A state-by-state comparison. *Thomas B. Fordham Institute*. Washington DC.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1985). Teacher evaluation: A study of effective practices. *The Elementary School Journal, 86*(1), 61-121.

Wright, L. P. (2015). *Assessing and guiding instructional practice: Administrators' and teachers' perceptions of the framework for teaching evaluation.* Walden University, Minneapolis, MN.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks: Sage Publications.

VITA

Ryan Curtis Most graduated from Gardener-Edgerton High School located in Gardner, KS in 1995. Immediately following, he enrolled at the University of Kansas and completed an undergraduate degree in secondary education and certification for English/language Arts for grades 7-12. Additionally, Mr. Most completed a Master's degree with the University of Kansas in curriculum and instruction during the summer of 2004. After a short break, he enrolled with Kansas State University and completed coursework to add the English as a Second Language certification to his license. Upon completion of the additional licensure, Mr. Most enrolled at the University of Missouri-Kansas City, completing a specialist's degree for education leadership in 2007. Mr. Most began work towards his Ed. D. in PK-12 Education Leadership in 2014.

Following the completion of the undergraduate degree, Mr. Most served as an eighth grade and freshman English teacher in the Atchison County Community School District. Mr. Most returned to the Kansas City area and acquired a position with USD 500-KCKPS as an eighth grade teacher. Upon completing the ESL certification, he became an ESL instructional coach in the same district, and upon completion of the specialist's degree, Mr. Most pursued administrative positions in the same district, first as an assistant principal in a middle school, then as head elementary school principal.