

TEACHER EVALUATION POLICY REFORMS INVOLVING STUDENT DATA:
TEACHERS' PERCEPTIONS

BY

EMILY A. WEIDNER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Education in Educational Organization and Leadership
with a concentration in Educational Administration and Leadership
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Assistant Professor Rachel Roegman, Chair and Director of Research
Associate Professor James Shriner
Clinical Assistant Professor Mary Herrmann
Associate Professor Anjale Welton

Abstract

This study explored the perceptions of elementary school teachers in two Illinois public schools who have been evaluated using student test data. One school used a student learning objective approach and one school used an all-in approach. The theoretical framework for this study was relational trust. The research questions were: (a) To what extent do Illinois public school teachers who are evaluated using student learning objective and all-in student growth approaches perceive the processes to be understandable, fair, and valid? (b) To what extent have these student growth approaches influenced their professional practice? (c) How does the principal's approach to teacher evaluation during the student growth component of the evaluation influence teachers' professional development? In order to answer these questions, a qualitative interview study was conducted. Two principals and nine teachers were interviewed using semi-structured, open-ended interviews. Three main findings emerged after the interviews were coded and analyzed. First, teachers at the SLO school had a more positive view of the student growth model at their school than teachers at the all-in school. Teachers at the all-in school reported that the student growth model influenced their professional practice more than teachers at the SLO school. Both principals maintained a high level of trust with their teachers through the process by exhibiting competence and integrity and demonstrating care and respect for their teachers.

Acknowledgements

Thank you to the members of my committee Dr. Mary Herrmann, Dr. James Shriner, Dr. Anjale Welton, and especially to my advisor and committee chair, Dr. Rachel Roegman. I could not have finished this project on this timeline without the quick and constructive feedback that Dr. Roegman consistently provided. I aspire to be as positive and kind as her someday. I would also like to acknowledge the other instructors that helped guide me at the University of Illinois, including Dr. Don Hackmann, Dr. Christopher Dunbar, Dr. Asia Fuller-Hamilton, Dr. Steve Epperson, Dr. Kim Wolowiec-Fisher, and Dr. Kern Alexander.

I am beyond grateful to the teachers and principals who volunteered to participate in this study. They made space in their busy schedules to help a fellow educator, and I am indebted to them for their kindness.

I would also like to thank my superintendent, Dr. Vic Zimmerman, for being an outstanding mentor, encouraging me to pursue this degree, and offering advice along the way. My staff at the best school in the land, White Heath Elementary School, has also supported and encouraged me through this process as have my fellow principals in MCUSD #25.

Finally, I could not have done this without the help and sacrifice of my family. Thank you to my mom, Marty, and my dad, Tom, for their never ending willingness to drop everything and help me with the kids or anything else that I need, and for teaching me the value of hard work through their example. Thank you to my brother, Gabe, for always cheering me on in anything that I do. Thank you to my study buddies, my sister, Sarah, and my best friend Kim, for the companionship during long study sessions at Harvest Market on many nights and weekends. Special thanks to my kids, Jack, Ali, and Grace, who barely remember a time when their mom wasn't busy writing every night at the kitchen table and who are my proudest

accomplishment. I hope that this process has shown you that hard work and perseverance are required to do great things. Last but not least, thank you to my husband, Justin. He has sacrificed the most in order for me to complete this study. In spite of the extra work and responsibility that this brought on him, he never complained, and he consistently stepped up to take care of housework, homework, carpools, cooking, groceries, and a wide variety of last second errands. I could not have finished this without him, and I am grateful to have a husband who is my biggest fan.

*Dedicated to my Mom and Dad, Tom and Marty Bridwell, my husband, Justin Weidner,
my kids Jack, Ali, and Grace Weidner, my sister, Sarah Young, my brother,
Gabe Bridwell, and my best friend, Kim Patton*

Table of Contents

Chapter 1 Introduction	1
Chapter 2 Review of the Literature	10
Chapter 3 Methodology	49
Chapter 4 Results	62
Chapter 5 Discussion and Implications	84
References	100
Appendix A Email Requesting Access From Superintendents	112
Appendix B Email Confirming Study Eligibility and Soliciting Participation From Principals	113
Appendix C Email Requesting Participation From School Teachers	114
Appendix D Teacher Evaluation Policy Reforms Involving Student Data: Teachers' Perceptions	115
Appendix E Interview Protocol for Principals	118
Appendix F Interview Protocol for Teachers	120
Appendix G Coding Chart	122

Chapter 1

Introduction

The fear that students in the United States are not keeping pace with their global peers has had an ongoing influence on educational policy reform and dates back to at least 1957 with the Soviet launch of Sputnik. This fear spiked again after the release of *A Nation at Risk* in 1983 (National Commission on Education, 1983) and more recently with the poor results of U.S. students on the Program for International Student Assessment (PISA), National Assessment of Educational Progress (NAEP), and Third International Mathematics and Science Study (TIMSS; Gurl et al., 2016). After Sputnik, educational policymakers pushed mathematics and science education reforms in order to win the Space Race. In the last 50 years, policymakers have sought to find the reason why students in the U.S. are not achieving the same test scores as their global peers and have looked for solutions that will keep American students globally competitive. Many policymakers developed the belief that schools and educators must be held more accountable in order for the students to gain ground on students in other countries. Perhaps if schools and educators faced real consequences for poor student performance, the schools would be more motivated to ensure that all students succeed.

Based on the idea of global competitiveness as well as concerns about the achievement gap between different demographic groups in the United States, the federal government in the United States placed a great deal of emphasis on student achievement data as an indicator of the success of schools and the competence of teachers (Lavigne & Good, 2014). The No Child Left Behind Act of 2002 marked the U.S. federal government's foray into holding schools accountable for the test scores of their students. Schools whose students did not meet the government's adequate yearly progress goals on standardized reading and math tests faced

restructuring or even closure. District and school scores were made public so that parents could be informed about the quality of schools and have the information needed to make decisions about enrollment. The goals for student test scores increased every year under the No Child Left Behind Plan, with the expectation that 100% of students would meet or exceed standards on the mandated standardized tests by 2014. This goal was never attained, and another problem emerged in the United States that caught the attention of policymakers. They became concerned that traditional teacher evaluations were not successful in identifying and removing incompetent teachers.

As a result of the push for accountability and the desire to make sure that underperforming teachers were not allowed to continue negatively impacting student performance, the federal government pushed the use of student test scores for educator accountability into the realm of teacher evaluation by tying \$4.35 billion in 2009 Race to the Top (RttT) education grants to a variety of reform criteria, including the adoption of teacher evaluation systems that have student growth components (U.S. Department of Education, 2009). States scrambled to implement reforms that would allow them to qualify for RttT grant money with the result that by 2016, 42 states had teacher evaluation systems that included a student growth component.

Illinois was one such state, passing the Performance Evaluation Reform Act (PERA) in 2010. This law required districts to adopt teacher evaluation programs that included both a professional practice rating based on clearly established standards of teaching practice and also a student growth component based on student assessment scores by September of 2016 (Growth Through Learning, 2012). School districts were given some degree of flexibility in adopting teacher evaluation systems that met the guidelines of the new law. Illinois also offered a state

model that districts could adopt if the joint committee of educators could not come to a consensus on a new evaluation model. Many districts implemented a student learning objective approach. Student learning objectives involve the teacher giving students a pre-test, then setting individual learning objectives for student achievement, and finally giving a post-test to see if the students met or exceeded the goal. Every teacher in every subject would be required to give two such assessments with student learning objectives for each one. The specific assessment must be agreed upon with the administrator or evaluator and the student learning objective must also be agreed upon in advance. The student learning objective approach allows for educators to give assessments that are specific to their subject area. For example, a Physical Education teacher would give assessments involving Physical Education. However, there are parameters for the type of assessments that educators can use based on whether tests are national, districtwide, or teacher-created. Another approach that some districts adopted is called an all-in approach. Under this type of plan, the entire school or grade level shares the same assessments and goals. Generally, the goals would be related to Math or English Language Arts, and all teachers in all subject areas would be rated based on the same student assessment scores, not on their own subject areas assessments.

For many school districts, the PERA mandates required a great amount of work and change in a short period of time. Districts created joint committees of administrators and teachers to select new evaluation models, administrators had to be trained, and then teachers had to be trained. Since all districts had to implement new systems that met PERA requirements by 2016, the implementation stage is now complete. Thus, it is a logical time to see what type of effect these reforms are actually having on the teachers and their practice without having to factor in

the probable discomfort of the change process negatively skewing teacher and principal perception of the reform.

Statement of the Problem

In order for teacher evaluation to positively influence the professional practice of teachers, it is important that teachers understand the evaluation process, trust that it is fair, and find value in the feedback received. Prominent educational researchers have pointed out serious concerns about many of the value-added models (VAMs) for assessing student growth that have been implemented in the United States (Amerein-Beardsley, 2014; American Statistical Association, 2014; Baker, et al., 2010; Gabriel & Allington, 2012). Research also indicates that teachers have concerns about the use of test scores in their evaluations (Crystal, 2014; Foutch, 2017; German, 2014; Jiang, Sporte, & Luppescu, 2015; Pressley, Roehrig, & Turner, 2018). However, in Illinois, districts were given some flexibility in student growth model selection and a VAM approach was not applied across the state. There is a gap in the research regarding teacher perceptions of two of the most common models that were implemented in Illinois: SLOs and the all-in approach. Furthermore, the majority of research regarding PERA in Illinois was conducted during the implementation phase of the new models and is likely to reflect perceptions of the initial change process rather than the actual procedures as they operate now. Another problematic issue is the lack of research into the experiences of elementary school teachers, including those whose students are too young for the standardized state tests that many student growth models use. It is important for principals to understand the effect that the student growth model has on their teachers so that they can achieve positive educational outcomes and effectively guide teachers through the process. Relational trust between the principal and the teacher is at stake when new policies are implemented, but this same trust may be able to

positively influence the experience of the teacher if the principal understands the teachers' feelings and perceptions about the process.

Statement of Purpose

The purpose of this study was to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using various student growth models, and to what extent the principals' approach to the process influences the teachers' engagement with these approaches. By gaining an understanding of the experiences of teachers, principals will be better equipped to implement the student growth component in the most effective way. Additionally, teachers' perceptions could demonstrate the need for districts to make revisions or improvements to teacher evaluation systems.

Research Questions

This study examined three main research questions.

1. To what extent do Illinois public school teachers who are evaluated using student learning objective and all-in student growth approaches perceive the processes to be understandable, fair, and valid?
2. To what extent have these student growth approaches influenced their professional practice?
3. How does the principal's approach to teacher evaluation during the student growth component of the evaluation influence teachers' professional development?

Theoretical Framework

The framework for this study is relational trust. Bryk and Schneider (2002) found that when relational trust is high within an organization it is easier to develop and maintain long term improvement efforts. Teachers and principals depend on each other to work in a symbiotic relationship and trust is key to ongoing cooperative efforts toward common goals.

Methods

I conducted a qualitative interview study in order to understand the experiences of elementary school teachers who have experienced teacher evaluations since the implementation of PERA reforms. Two public school districts were selected using purposeful sampling, one that uses the SLO process and one that uses an all-in approach. The study was narrowed to elementary schools in order to shed light on the experiences of teachers at this level, including those whose students are not tested using standardized tests. Teacher interviews, principal interviews, and document reviews were utilized in each district in order to gain an understanding of teacher perceptions, principal approaches, and processes within the schools.

Interview data was coded and analyzed based on the theoretical framework of relational trust between teachers and principals. Member checks were completed with the subjects in order to ensure that the data reported matched what the subjects provided (Miles, Huberman, & Saldana, 2014). Internal validity was ensured by establishing a chain of evidence. Ethical considerations were addressed by receiving IRB approval prior to beginning the research study, providing informed consent to all participants, and using pseudonyms for districts, schools, and individuals.

Limitations

This study had some limitations. Interviews are a valuable source of firsthand information, but it is possible that interview subjects may not share complete or truthful information if they are concerned about workplace repercussions, portraying themselves in a negative light, or criticizing others. Furthermore, those interviewed can speak only for themselves, not all members of the school or district. Interview subjects have their own motives for participating in the study and are not without their own personal biases. While document

review can also offer insight into policies and procedures, documents do not always accurately reflect how policies are being implemented in reality, and documents are not always updated as processes evolve (Creswell, 2014).

Generalizability was another limitation of this study. Research was conducted in only two districts in Illinois. The small sample of districts and educators negatively impacted the generalizability of the results. Since districts in Illinois were given flexibility in PERA implementation, research from individual districts will not be applicable to other districts within the state or in other states in the United States.

Delimitations

This study was delimited in several ways. First, the study focused only on the student growth component of PERA, and not the professional practice portion of teacher evaluation. The search criteria also delimited the research to public schools in Illinois that implemented the two specific student growth processes of SLOs or all-in. The study was also narrowed to elementary school principals and teachers in two separate school districts.

Significance of the Study

This study is significant because the use of student growth measures in teacher evaluation in the United States has a controversial history. During the implementation of PERA reforms, feedback from principals and teachers was likely to be tinged with the discomfort that naturally accompanies change efforts. Now that the implementation of the reforms is complete in Illinois, data regarding the effect that the reforms have had on the teachers and their relationships with their principals can be valuable for assessing the effectiveness of the reforms. While this information is likely to apply most directly to other public schools in Illinois, many districts

across the country use some form of student growth in teacher evaluation and could therefore benefit from the knowledge gained through this research project.

Definition of Terms

All-in approach. An all-in approach refers to a student growth model in which the same learning goals, assessments, and outcomes are applied to all teachers in the school or district regardless of the teachers' subject area.

Illinois State Board of Education. The Illinois State Board of Education (ISBE) develops policies and guidelines for public and private schools in Illinois. The board recommends legislation to the Illinois General Assembly and Governor based on educational data collected from public and private schools across Illinois.

Performance Evaluation Reform Act. The Performance Evaluation Reform Act (PERA) was signed into law in Illinois in 2010. PERA required that principal, assistant principal, and teacher evaluations include student growth as a significant factor and use a four category rating system (Excellent, Proficient, Needs Improvement, and Unsatisfactory). Additionally, evaluators must be complete ISBE evaluation training. (ISBE, 2019).

Race to the Top. Race to the Top was a competitive federal grant program that was signed into law in 2009 as part of the American Recovery and Reinvestment Act. \$4.35 billion was made available for states that were willing to implement reforms in four key areas: adoption of rigorous learning standards and assessments, utilizing data systems that measure student growth, recruitment and retention of high quality teachers and principals, and improving the lowest-achieving schools (U.S. Department of Education, 2009).

Student growth. Student growth is defined as a demonstrable change in a student's or group of students' knowledge or skills, as evidenced by gain and/or attainment on two or more assessments, between two or more points in time”

Student learning objective. As defined in Illinois Administrative Code 50.10, SLOs include a learning goal, assessment, measurement model, and student growth prediction. Growth predictions are made based on the needs of students in specific teachers' classrooms (ISBE, 2015).

Value added model. Value added models use regression techniques to predict student scores on assessments based on previous scores while statistically controlling for student characteristics such as special education status, English language learner status, mobility, and socioeconomic status (Reform Support Network, 2015).

Summary

This chapter included an overview of this study including the problem, purpose, research questions, methods, limitations, delimitations, key definitions, and the significance of the study. Chapter 2 provides a review of the literature regarding teacher evaluation, types of student growth models, and pertinent research regarding teacher perceptions of student growth in teacher evaluation. Chapter 3 describes the methodology selected for the study, which is a qualitative interview study utilizing semi-structured open-ended interviews and document reviews at two public school districts in Illinois. Chapter 4 will provide the findings of the study and Chapter Five will summarize the findings and describe the implications of the findings for educators, policymakers, and researchers.

Chapter 2

Review of the Literature

This chapter provides a review of relevant literature on the topic of teacher evaluation, with a focus on evaluation processes that include measures of student growth. There is an abundance of research on the use of student test data in evaluation, especially as it relates to the use of standardized test scores and value-added statistical models (VAMs). However, there is limited research examining how teachers perceive the use of student learning objectives or all-in approaches and the influence that these models have on their teaching practice and relational trust with colleagues and administrators. Background on teacher evaluation, related reform movements, and research on the use of test scores in evaluations is included in this chapter. This chapter also presents relational trust as a theoretical framework. This framework highlights the connections among establishing relational trust, protecting teacher self-efficacy, and successfully implementing school improvement initiatives.

Danielson and McGreal (2000) and Tucker and Stronge (2005) stated that the first purpose of teacher evaluation is to ensure teacher quality, and the second is to promote teachers' professional growth. Danielson and McGreal (2000) asserted that teacher evaluations often fail to achieve either of these purposes and have instead become merely a series of steps that principals and teachers resign themselves to complete. As a result of these concerns and continued warnings that students in the United States score lower than their global peers on standardized tests (U.S. Department of Education, 2010), teacher evaluation has been the subject of major reform efforts at the federal and state levels over the last two decades. These reform efforts focused on both the teacher quality portion of the evaluation and the professional improvement portion, based on the notions that ineffective teachers were not being identified

under current systems and that teachers would be motivated to improve student performance if they were held accountable (Pianta & Kerr, 2014; Ravitch, 2014). To assess teacher quality, 38 states and the District of Columbia have now implemented the use of student test scores as part of a teacher's evaluation; this is based on the belief that teacher evaluation systems will be more objective and data-based. As a result, advocates argue that these new teacher evaluation plans will be more likely to identify teachers whose students are not achieving (NCTQ, 2017). Now that these restructured systems have been implemented, it is important to determine the positive and negative impacts of the reforms in order to guide future actions.

History of Teacher Evaluation

Teacher evaluation processes and policies in the United States have largely been influenced by state and federal mandates and reforms. Government policies and reforms will be outlined in a later section. According to Marzano, Frontier, and Livingston (2011), teacher evaluation in the United States has evolved from unstructured and mostly supervisory processes completed mostly by local clergy in the 1700s to the complex and federally regulated evaluation systems of the 21st century. As common public schools became more widespread in the 1800s, supervision of teachers shifted from the clergy to lead teachers or principals who had more expertise in teaching skills. However, according to Shinkfield and Stufflebeam (1996), formal evaluation of teachers was nearly non-existent until the 1900's and it was not until the second half of the century that these formal evaluation processes became widespread in the United States.

By the 1960s, clinical supervision began being used in teacher evaluation across the United States (Marzano et al., 2011). It was developed at Harvard University by Morris Cogan and other researchers and was based on how doctors were observed for improvement of their

professional practice in teaching hospitals (Marzano et al.). Goldhammer (1969) furthered Cogan's work by developing five phases of clinical supervision that are still evident in many teacher evaluation formats today. The process includes the pre-observation conference, classroom observation, analysis, supervision conference, and a plan for further professional development of the teacher (Mette, Range, Anderson, Hvidiston, & Nieuwenhuizen, 2015). Goldhammer did not identify specific criteria for effective teaching but sought to use his approach to supervision as a way to spark rich conversations with teachers about improving instructional practice. Despite his intent, the five parts of the clinical supervision model that Goldhammer popularized became widely used as a teacher evaluation model rather than the supervisory tool he attempted to create (Marzano et al.).

In the 1980s, Hunter (1989) influenced the course of teacher supervision and evaluation in the United States. Teachers began planning their lessons based on Hunter's seven-step framework, and supervisors evaluated their effectiveness based on alignment to Hunter's model (Marzano, et al., 2011). Hunter also popularized the use of script taping during clinical supervision of teachers. In script taping, the teacher's words during an observation were recorded and organized into categories in order to determine their impact on student learning (Marzano et al.). Hunter believed that script taping would ensure accuracy of observations so that the observer and teacher could determine if: (a) the teacher was following the main objective of the lesson, (b) the students' responses show that they are achieving the objective, (c) the level of difficulty was adjusted properly, and (d) the principles of learning were applied (Shinkfield & Stufflebeam, 1996).

Evaluation Models

Many models for teacher evaluation and variations of those models have been used in the United States since the formal evaluation of teachers became prominent in the 1950s (Shinkfield & Stufflebeam, 1996). Brandt, Thomas, and Burke (2008) noted that decisions about teacher evaluation processes are often left to school districts and collective bargaining units. However, most states provide policy guidelines that specify some criteria for teacher evaluation. Manatt introduced one of the country's earliest teacher evaluation models, the Teacher Performance Evaluation (TPE), in the early 1980s and expanded the use of clinical supervision. The steps in the TPE cycle were (a) establish rules, (b) orient teachers, (c) analyze a lesson plan, (d) preobservation conference, (e) classroom observation, (f) postobservation conference, (g) analyze data, (h) evaluation report, and (i) set improvement goals (Shinkfield & Stufflebeam). Also during the 1980s, a time when President Reagan sought to increase local control over educational policy-making, McGreal worked with hundreds of districts across the country to implement the eight commonalities that he identified for effective teacher evaluation systems (Shinkfield & Stufflebeam). McGreal (1983) stated that districts should (a) adopt an appropriate attitude toward evaluation, (b) design a model to meet the desired purpose of the district, (c) separate supervisory behavior from administrative behavior, (d) set goals, (e) focus on teaching, (f) improve observation skills, (g) use additional sources of data, and (h) adopt a training system that complements the evaluation system (McGreal, 1983). McGreal suggested that districts should choose policies and procedures that work for them locally, while adhering to these eight principles.

Danielson shared McGreal's viewpoint that districts should be specific in their processes and policies and outline expectations for teachers and administrators throughout the evaluation

process (Danielson & McGreal, 2000). Danielson (1996) developed a framework for effective teaching that was originally created to help identify all of the things that teachers should know and be able to do in order to guide professional growth. It is commonly referred to as the Danielson Model. Danielson first developed the framework to encourage productive conversations about effective teaching, not for the purpose of teacher evaluation (Danielson Group, 2019). The Danielson model is outlined in further detail here because it is the default model in Illinois.

In an effort to capture all aspects of teaching, Danielson (1996) broke down the teaching profession into four domains with multiple sub-domains in each area.

- Domain 1: Planning and Preparation
- Domain 2: The Classroom Environment
- Domain 3: Instruction
- Domain 4: Professional Responsibilities

Danielson provided a rubric with four different ratings for each domain and sub-domain. She proposed that the framework could be used for new teacher preparation, recruitment and hiring of teachers, to guide novice teachers, to help experienced teachers grow, to focus professional development, and to communicate teacher effectiveness with stakeholders (Danielson, 1996). Parts of the Danielson Framework can be observed by evaluators in the classroom, while other components require evaluators to review documents like lesson plans, unit plans, assessments, student work, and evidence of teachers' professional development efforts and contributions to the school (Danielson). Danielson (1996) noted that effective use of the framework must include a pre-observation conference, observations, a post-observation conference, and artifacts provided by the teacher that provide evidence of competence in the domains.

The first revision of the framework came in 2007 and included rubrics for specialist teachers whose job responsibilities are different from those of classroom teachers (Danielson, 2013). In 2011, Danielson revised it again so that it could be used in the Measures of Effective Teaching study sponsored by Bill and Melinda Gates, and the most recent version was released in 2013 in order to be more applicable to teaching under the Common Core State Standards (Danielson, 2013). The Danielson model has been adopted in school districts across the United States and the world.

Marzano also developed a teacher evaluation model that has been used in several states, including Florida and Michigan (Carbaugh, Marzano, & Toth, 2017). Marzano's model includes Focused Model protocols that support evidence of student learning. The plan also includes a five step process for classroom observations: (a) Observe elements of the protocol that the teacher is using and determine if they are being used correctly. Rate the teacher on a sliding scale accordingly. (b) Observe how the teacher monitors whether the teaching had the desired effect upon the student. (c) Determine the percentage of students who reach the desired outcome and rate the teacher from 2-4 based on the percentages. (d) If the teacher does not reach a level 3 or 4 after the outcomes are assessed, the observer monitors whether the teacher makes adaptations to instruction that change the outcomes and rate the teacher accordingly. (e) Review teacher-provided evidence during a post-conference to clarify how many students eventually demonstrated the desired effect (Marzano, 2017).

Some teacher evaluation models use different methods than the typical clinical supervision and rating scale models. The Tennessee Value-Added Assessment System (TVAAS) provides a prime example. The TVAAS was mandated by the state of Tennessee and took effect in 1992. This system tracked student progress on state assessments over time and attributed

growth or lack of growth to teachers, who were evaluated based on the students' growth as compared to the growth that the statistical model proposed they should have achieved (Shinkfield & Stufflebeam, 1996). The trend of using student data in evaluation became more widespread at the turn of the 21st century, and by 2015, 42 states and the District of Columbia mandated that student test or other objective measures of student performance be included in teacher evaluations (NCTQ, 2015). Many districts in the United States now use a combination of traditional models of clinical supervision coupled with student test scores to create a summative teacher evaluation rating.

Teacher evaluations systems still vary greatly across the United States and are influenced by federal, state, and local policies. As of 2016, nine states and the District of Columbia had state-mandated teacher evaluation models with limited flexibility for local districts (Center on Great Teachers & Leaders at American Institutes for Research, 2016). Examples of teacher evaluation programs in these states include the TAP Evaluation System in Tennessee, COMPASS in Louisiana, ADEPT in South Carolina, and DC IMPACT in Washington, D.C. All of these systems included a student growth component as well as observation components (Center on Great Teachers & Leaders at American Institutes of Research). Eighteen states have adopted elective state models. These states have a model teacher evaluation system that districts can choose to adopt or adapt to their local needs. Districts within these states can also create their own evaluation systems that fit the state criteria. Among these states, Illinois, Arkansas, South Dakota, and Wisconsin selected the Danielson Framework to be used for the professional practice component of the teacher evaluation (Center for Great Teachers & Leaders at American Institutes of Research). Additionally, 23 states allowed local districts to create their own evaluation systems without a required or default model to follow, provided that the districts met

the requirements for evaluation systems established by the state. Two of these states, Idaho and New Hampshire, specified that districts must develop plans based on the Danielson Model (Center for Great Teachers & Leaders at American Institutes of Research).

Purpose of Teacher Evaluation

It is widely argued that highly effective teachers have the greatest effect on student learning outcomes (Lavigne & Good, 2014; Sanders & Horn, 1998; Shinkfield & Stufflebeam, 1996; Stronge, Ward, & Grant, 2011). “The transformative power of an effective teacher is something that almost all of us have experienced and understand on a personal level” (Tucker & Stronge, 2005, p. 1). According to Archer, Kerr, and Pianta (2014), “real improvement requires quality measurement” (p. 1). In order to ensure that students have access to quality teachers, it is necessary to identify what constitutes effective teaching and learning practices and how to quantify student learning. These authors posited that it is essential to evaluate teachers against recognized standards of best practice so that those who need further development are identified for improvement and those who excel are retained and promoted. Danielson (2011) and Marzano (2012) asserted that the purposes of teacher evaluation are two-fold: to ensure teacher quality through measurement and to promote professional learning. Peterson (2000) specified that evaluation should “protect children, reassure teachers that they are doing good jobs, assure audiences interested in teacher performance, make personnel decisions, inform teacher educators, and shape further practice” (p. 36). These assertions are related in that they all note that teacher performance must be measured. How that performance should be measured is a complicated question that researchers have varied opinions on, largely stemming from their perspectives on the reasons that teacher evaluation is needed. Those who promote teacher evaluation that focuses on improving teacher performance suggest comparing teacher

performance to standards of best practice and then recommending professional development based on areas in need of improvement related to those standards. Those who promote teacher evaluation that is focused more on teacher accountability and ensuring that ineffective teachers are removed from practice support less subjective measurements like student assessment data. These two approaches underlie the two major components of teacher evaluation: formative and summative evaluation (Shinkfield & Stufflebeam, 1996).

Formative evaluation. The formative component is specifically designed to help teachers improve their practice. It involves setting professional goals, recognizing outstanding work, planning individualized professional development, and helping to improve teaching skills and curricular choices. According to Marzano et al. (2011), “the purpose of supervision should be the enhancement of teachers’ pedagogical skills, with the ultimate goal of enhancing student achievement” (p. 2). The formative portion of teacher involves supervision of the teacher for improvement. Carroll (1997) defined supervision as the identification of what is happening in a teacher’s classroom with the intent of improving teaching practices and student learning. Danielson (1996) noted that teachers must understand how quality teaching is defined so that they can improve their professional practice in specific and meaningful ways.

The formative component of teacher evaluation is generally based on one or more formal observations of the teacher conducted by the supervisor. Typically, the teacher and evaluator will have a pre-observation conference, a formal observation, and a post-observation conference. At the pre-observation conference the teacher explains the learning goals and the lesson that will be taught. The supervisor communicates the teaching standards that are in use in the evaluation model. In Illinois, the state default model uses the Charlotte Danielson Framework for Effective Teaching, but joint committees had the option of adopting other frameworks. During the

observation, the supervisor notes what the teacher is doing effectively and what areas are in need of improvement. At the post-observation conference, the supervisor and teacher talk about how the lesson went. The teacher and supervisor discuss ways to improve the observed lesson and also how the teacher can improve curriculum and instruction in general. Ideally, the evaluator gives the teacher individualized goals and suggestions for professional development. While supervision is the goal of the formative portion, evaluation or the rating of what happens in the teacher's classroom is the purpose of the summative component of evaluation (Carroll).

Summative evaluation. The summative aspect of teacher evaluation is the quality control component. The summative or final rating of a teacher is used to make important decisions about teachers' careers and their continued employment. It is a rating and can be used to compare teachers to each other. Principals use these ratings to determine if a teacher is retained, remediated, or dismissed, to determine merit pay, and to decide who to hire (Danielson & McGreal, 2000; Marzano, 2012; Peterson, 2000). Teachers use these ratings to determine if they are competent in their job duties and to ascertain how their supervisors view their performance. The public also has a vested interest in the summative side of teacher evaluation, because the effectiveness of the teacher workforce directly affects most members of society. Lawmakers use summative evaluation data to assess the quality of the public schools, and parents and board members want reassurance that the teachers in their community are effective (Danielson & McGreal, 2000). In Illinois, the percentage of teachers in a school who are rated "excellent" or "proficient" is posted on the state report card so that any interested party can access that information (Illinois State Board of Education, 2018). Some states have even posted individual teacher evaluation ratings in newspapers and online.

Education Reforms in the Era of Accountability

Over the last 20 years, a focus on accountability and the use of student learning data has dramatically changed the landscape of public education across the United States. A new sense of urgency emerged in the 1990's after U.S. students performed poorly relative to their international peers on the Program for International Student Assessment (PISA), Third International Mathematics and Science Study (TIMSS), and National Assessment of Educational Progress (NAEP; Gurl et al., 2016). Many policymakers subscribed to the belief that using student test data to evaluate teacher performance and determine teacher pay would motivate teachers to improve their professional practice, thereby increasing student learning (U.S. Department of Education, 2009). Federal education reform policies like No Child Left Behind (NCLB) and Race to the Top changed the purpose of standardized testing in the United States. With the enactment of NCLB in 2002, the federal government began using standardized tests less for comparing student achievement to learning standards and more for determining the effectiveness of schools, administrators, and teachers. NCLB ushered in an era where students' test scores determined whether schools and districts were considered to be successful or failing. Those schools that did not make Adequate Yearly Progress toward annually increasing attainment benchmarks faced sanctions, restructuring, or even closure. Adequate Yearly Progress was defined by individual states, but the federal government mandated that states had to be the primary measure of progress, those tests had to include mathematics and reading/language arts, and subgroups of students had to improve test scores, not just the student population as a whole. The target for proficiency for Adequate Yearly Progress was 100% by 2014 (U.S. Department of Education, 2004). The theory behind this approach was that if schools were held accountable for their students' scores, the schools would either improve or students and parents could choose a

school with better test scores (U.S. Department of Education, 2004). This focus on student test scores continued with the re-authorization of the Every Student Succeeds Act in 2015.

In 2009, President Obama bolstered the federal government's dedication to the use of student test data to evaluate teachers by signing the American Recovery and Reinvestment Act, which included \$4.35 billion for Race to the Top education grants, which were awarded to states through a competitive proposal submission process. Grant application criteria included the adoption of common learning standards and common assessments as well as the development of teacher evaluation systems that included a student growth component (U.S. Department of Education, 2009). In addition to Race to the Top, the U.S. Department of Education also had developed the Teacher Incentive Fund (TIF) in 2006 to provide grants to schools that tied student achievement gains to teacher and principal compensation. The state of Illinois received \$42.8 million in the third phase of the Race to the Top competition (U.S. Department of Education).

As a result of these federal grants, states policy officials and school district leaders began looking for ways to hastily implement student growth models that would meet the Race to the Top grant or TIF criteria. Many state officials treated the reforms with urgency because the grant funds could help alleviate the funding hardship of the Great Recession, the economic crisis that occurred in the late 2000s (Callahan & Sadeghi, 2015). By 2010, at least 200 schools in 13 states had implemented one example of this type of teacher evaluation system, called the Teacher Advancement Program (TAP). TAP was designed to provide performance incentives to teachers in order to help with recruitment of quality teachers and increase student achievement (Glazerman & Seifullah, 2010). However, a 2-year study of TAP schools in Chicago found that students in TAP schools did not score higher on state assessments and retention rates for teachers were not impacted by TAP.

In 2009, only 15 states had legislation requiring student growth as a component of teacher evaluation (Education Commission of the States, 2018). By 2014, at least 35 states and the District of Columbia had adopted evaluation systems that included measures of student growth (Walker, 2014), and by 2015, 43 states had teacher evaluation systems that included some form of student growth measurement (Education Commission of the States, 2016). However, since 2017, 10 states have recently passed legislation or resolutions that reduce, eliminate, change, or require further study for the use of student assessments in teacher evaluation due to new flexibilities in The Every Student Succeeds Act (ESSA; Education Commission of the States, 2018).

Problems with Teacher Evaluation

The importance of evaluating teachers is evident, but the best methods for evaluation are still a source of debate. The traditional approaches that were widely used in United States public schools prior to Race to the Top Era reforms were fraught with problems and generally not considered to be valuable by teachers or administrators (Calabrese, Sherwood, Fast, & Womack, 2004; Carroll, 1997; Danielson & McGreal, 2000; Shinkfield & Stufflebeam, 1996). These traditional models of teacher evaluation typically involved a supervisor conducting one or two classroom observations of a teacher, writing up a report, sharing it with the teacher, and then filing it in the teacher's personnel file (Danielson & McGreal). Criteria for teaching effectiveness were often very subjective and unclear to teachers and recommendations for improvement were few.

Marzano (2012) noted that traditional teacher evaluation systems have not been successful at distinguishing between effective and ineffective teachers, and they have not built a highly skilled teacher workforce. According to Marzano (2014), there was insufficient time for

administrators to do a substantial number of observations in order to eliminate sampling error. It is unlikely that the evaluator could get a true picture of a teacher's professional practice through one to two observations in a year. Marzano also suggested that raters tend to score teachers higher than they deserve or than the teachers rate themselves. Danielson (2011) noted that teachers are often rated very high with few suggestions for ways that they can improve their practice. Stiggins (2014) agreed, noting that teachers often receive little to no feedback, and the outcome is almost always that teachers are retained. Pianta and Kerr (2014) stated that most common practices of teacher evaluation result in all teachers achieving good ratings and no differentiation between strong and weak teachers. According to Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012), "practitioners, researchers, and policymakers agree that most current teacher evaluation systems do little to help teachers improve or to support personnel decision making" (p. 8).

In a survey study across four states and 12 districts that included 15,000 teachers and 1,300 administrators, Weisberg, Sexton, Mulhern, Keeling, Schunck, Palcisco, and Morgan (2009) concluded that districts in the United States have consistently failed to acknowledge that teachers vary in their effectiveness, and districts have also consistently failed to appropriately address variations in teacher effectiveness. The evaluation systems in the surveyed districts varied. All the systems called for annual evaluation of non-tenured teachers, but not necessarily for tenured teachers. Most of the systems also involved two or fewer observations of teachers totaling 76 minutes or less (Weisberg et al., 2009). Weisberg et al. called the phenomenon they observed "The Widget Effect," meaning that schools act as if all teachers are equal and interchangeable. Weisberg et al. identified five ways that indifference to teacher effectiveness has been reinforced, including: (a) rating 94-99% of teachers as satisfactory, (b) no recognition

or reward for excellence, (c) lack of appropriate professional development linked to teacher evaluation, (d) no special attention to new teachers, and (e) not dismissing poorly performing teachers.

Calabrese, Sherwood, Fast, and Womack (2004) completed a case study in a midwestern school district that included interviews with 40 teachers and principals, a left- and right-hand case method, and review of district documents. The teachers and principals in the study indicated that they were frustrated with the evaluation process, which included observation, post-classroom observation, and a summative evaluation conference. Final ratings were based on a rating scale with rubrics (Calabrese et al.). Principals noted that they were required to implement an evaluation system that they did not help to create or support. They felt that the instrument was not clear in how it should be completed. They expressed the desire to perform evaluations in a more formative way, but they felt like they were limited by their local process to act in a more summative way. Teachers felt that their ratings were too subjective and that there were rules limiting the number of teachers who were allowed to achieve high ratings due to district policies. The teachers wanted evaluations that were not perfunctory but that would improve their performance.

The problems with teacher evaluation noted above are not new. Ginsberg and Berry (1989) found that teachers in South Carolina were frustrated with the teacher evaluation system that involved a checklist and a portfolio. They felt that some evaluators did not know enough about what they were supposed to be evaluating, and the criteria for the evaluations was unclear. Additionally, teachers felt that it was possible to “fudge” or embellish their portfolios in order to influence their evaluation ratings. In this study, 95% of teachers and principals were rated as satisfactory or better (Ginsberg & Berry).

Student Test Scores in Teacher Evaluation

There is an ongoing dialogue in the United States based on the idea that traditional teacher evaluation models were too subjective and created a system where ineffective teachers were not identified, and consequently, were neither required to improve their performance nor were terminated for poor performance (Weisberg et al., 2009). This idea, coupled with the desire to hold teachers and schools accountable for student test scores, gave rise to the use of value-added models (VAMs). These assessment models were designed “to estimate effects of individual teachers or schools on student achievement while accounting for differences in student background” (American Statistical Association, 2014, p. 1). The concept is that student test scores can be compared longitudinally with their previous scores while statistically controlling for outside factors like gender, race, or socioeconomic status. Ideally, these VAMs could quantify the effect of the school or of individual teachers on students’ academic growth during the year (Pearson Education, 2004). Proponents of these models feel that these scores are highly valuable because they are quantitative and objective measures of student performance (Collins, 2014). They believe that using student test data eliminates the subjectivity from teacher evaluations and shifts the focus to student achievement scores that are quantifiable and are likely more defensible in court. The belief is that value-added data could help to identify poorly performing teachers so that they could be provided professional development to improve their methods or be more easily terminated (Collins).

William Sanders was one of the original advocates of VAMS and the creator of the Tennessee Value-Added Assessment System (TVAAS). Horn and Sanders (2009) asserted, “race, socioeconomic level, class size, and classroom heterogeneity are poor predictors of student academic growth. Rather the effectiveness of the teacher is the major determinant of

student academic progress” (p. 3). They claimed that by using a database of students’ state standardized test scores and comparing them longitudinally, the effectiveness of a teacher could be quantified. The TVAAS system was originally created to be part of a summative evaluation for teachers, but Horn and Sanders also said, “the real power of the process lies in its ability to serve as a data source for formative evaluation and for educational research” (2009, p. 3).

Concerns About Value-Added Models

There is abundant research aimed at assessing the advisability of using value-added models in teacher evaluations. Scholars have pointed out major shortcomings with using VAMs in teacher evaluation including false assumptions about test data, flawed evaluation models, and unintended side effects (Amerein-Beardsley, 2014; American Statistical Association, 2014; Baker et al., 2010). Gabriel and Allington (2012) found that value-added models were often unreliable, invalid, and not understood by those who are supposed to use them to improve teaching.

The first assumption is that standardized tests can effectively measure student learning. Baker et al. (2015) noted that, “Standardized tests now in use are not perfect and do not provide unerring measurements of student achievement” (p. 5). At the national level, these tests only cover core subjects like English Language Arts (ELA), math, and sometimes science, which omits many categories of teachers from value-added models or lumps them into goals about core subjects that are not aligned with the learning standards of their discipline. Baker et al. (2010) questioned the fairness of evaluating only certain teachers with VAMS and of evaluating specialists, like physical education teachers or music teachers, based on students’ Math or Reading achievement scores.

A second assumption is that monetary incentives tied to higher standardized test scores will increase student achievement. Research shows that linking student test scores to teacher pay does not increase student scores. In a study of teacher incentives and student achievement in New York City Public Schools by the U.S. Department of Education, the authors found that the program had a negative impact on middle school achievement in ELA and math, lower graduation rates in high schools, and no statistically significant different impact in elementary schools (U.S. Department of Education, 2013). Additionally, research indicated that the four most prominent merit pay systems in the country, the Iowa Pay-for Performance Program, the Texas Governor’s Educator Excellence Grants, the Chicago Teacher Advancement Program, and the Denver Professional Compensation for Teachers Program did not show evidence of raising student test scores (Education Commission of the States, 2010). Moreover, connecting teacher pay to student test scores has motivated some misguided educators to cheat. In 2015, nine teachers in Georgia were sentenced to prison for racketeering after an investigation revealed widespread cheating on state tests that actually involved more than 180 staff members (Fausset & Blinder, 2015).

Another assumption is that using value-added models will make it easier to terminate ineffective teachers. The argument is that under traditional evaluation systems, supported by powerful teacher unions, teachers could easily challenge termination based on highly subjective models of performance evaluation. However, as evidenced by a wave of lawsuits across the nation, value-added models are just as likely to invite litigation when used to terminate teachers. For example, in 2013 Florida teachers in conjunction with the National Education Association the Florida Education Association challenged the state’s evaluation system in court, citing a violation of the teachers’ constitutional rights. While the court agreed that the VAMs were

flawed, the matter was not considered a violation of constitutional rights (Walker, 2014). Additionally, in 2014 teachers in Houston, Texas also filed suit over the use of value-added measures in their evaluations (Sawchuk, 2014). The Tennessee Education Association filed three lawsuits challenging the use of the TVAAS between 2014 and 2015 (Balakit, 2015). In 2015, the American Federation of Teachers, along with New Mexico teachers and legislators, filed a lawsuit against New Mexico's Public Education Department, claiming the teacher evaluation system was punitive and error-ridden (Brown, 2015).

Flawed models. The lawsuits largely stem from claims regarding significant defects in the value-added models, such as the lack of reliability in summative teacher evaluation ratings. Teachers who are rated effective one year may not be effective in a subsequent year, in spite of using the same teaching strategies and assessments. In a study on the SAS Education Value-Added Assessment System in Houston, 404 out of 874 teachers indicated that their scores were inconsistent from year to year, and 24.4% of those with inconsistent scores cited changes in the categories of students being taught as impacting their scores (Collins, 2014). In the same study, about half of teachers also noted that their scores were not consistent across grade levels, subjects, and groups of students taught.

In addition to problems with reliability, teachers and principals often did not understand the data they were given. Amrein-Beardsley (2014) found in studies of the widely-used EVAAS and TVAAS VAMs that teachers not only did not understand the calculations, but they were also not trained on how to apply and interpret the data to improve teaching and learning. According to Sparks (2011), teachers in Houston who got bonuses associated with good VAM scores “consider it like winning the lottery: I don't know what I did to deserve this, but I'll enjoy spending it” (p. 11).

In addition to a lack of understanding of the systems, researchers have identified many concerns about the validity of value-added scores. Collins (2014) noted that in Houston, 9.1% of teachers were assessed based on scores for a grade level that they did not teach, and 9.6% had been evaluated based upon a subject for which they were not a teacher of record. Additionally, 17.5% of teachers had been evaluated using student data for whom they were not the primary teacher, and 57.6% of teachers indicated that their evaluations were inconsistent with their observation scores. After value-added scores were released to the press in New York City, 73% of principals said that the teachers who were labeled as ineffective were not considered as such based on observations (Amrein-Beardsley, 2014).

Finally, researchers questioned the validity of value-added models because teachers, especially at the elementary level, are not assigned to random classes of randomly sorted students. Often students were grouped into classrooms and assigned to teachers based on academic ability, avoidance of personality conflicts, special needs, English Language Learner status, and/or parent preference. This non-random sorting led to some teachers having students who were more likely to demonstrate growth due to innate capability, parental support, access to technology at home, resources for enrichment learning outside of school, and more. Yet, other teachers may have been assigned students who were less likely to show growth due to socio-economic status, absenteeism, disabilities, and mobility (Amrein-Beardsley, 2014). According to Darling-Hammond et al. (2012), “teachers are advantaged or disadvantaged based on the students they teach” (p. 10). A study in Houston, Texas, where the EVAAS model was in use, found that:

- Teachers of grades in which English language learners (ELLs) are transitioned into mainstreamed classrooms are least likely to show “added value.”

- Teachers of large numbers of special education students in mainstreamed classrooms are also found to have lower “value-added” scores, on average.
- Teachers of gifted students show little value-added because their students are already near the top of the test score range.
- Ratings change considerably when teachers change grade levels, often from “ineffective” to “effective” and vice versa. (Darling-Hammond et al., 2012, p. 12)

Unintended side effects. The movement toward using value-added models in the United States has also had many consequences. In many cases, financial pressures motivated states to develop systems that met Race to the Top or TIF criteria. States and districts rushed to write grant applications to meet the requirements and many have since spent time over multiple years planning and implementing new and often more involved evaluation systems (Sparks, 2011). Evaluation statutes, administrative rules, procedures, and models generally had to be developed at the state level and then adapted to the local level. Evaluators had to be trained and then teachers had to be trained on local district processes. Often the local process involved ongoing meetings and discussions between teachers’ unions and administrators. Drew Furedi from the Los Angeles School District said:

It was the Noah’s Ark of task forces: We had two of everyone from teachers and principals to union and central-office staff, parents, students, and community leaders. The group of 35 met for six months, setting up criteria for a new, comprehensive evaluation system. (Sparks, 2011, High visibility, para. 4)

When the new systems were implemented, the time administrators and teachers had to invest in the evaluation process increased due to the student growth component. The Teacher Advancement Program, which was founded in 1999 and includes schools across the United States, provides an example of the significant amount of time involved in some evaluation processes. In TAP, teachers are observed at least four to six times per year and their evaluators have to be trained for 4 full days in order to be approved as evaluators. Teachers also watch and rate videotaped lessons in addition to the traditional post-observation conferences (Darling-

Hammond et al., 2012). The time spent on creating and implementing systems takes away from time teachers could use to plan for students (Foutch, 2017).

Perhaps the most troubling side effect of value-added models is that teachers and schools are encouraged either directly or indirectly to narrow the curriculum. In order to improve scores on value-added measures, teachers are incentivized to shift their focus almost entirely to English Language Arts and mathematics. This shift increases the likelihood that instruction in other vital areas like social studies, science, and the arts is reduced (Baker et al., 2010). Moreover, instruction in ELA and math is likely to emphasize test taking skills and test preparation than to include projects that are applicable to the real world or promote creative thinking. Generally standardized tests are more efficiently graded if they include mostly multiple-choice questions. Short-answer and essay responses are kept to a minimum to reduce the time and money required for scoring. Preparation for such tasks promotes the reduction of instruction in writing, discussion, and real-world problem solving (Baker et al., 2010). Berliner (2011) stated that narrowing the curriculum in order to improve test scores has multiple negative consequences. Creative students who have talent outside of the narrow scope that is tested might have limited opportunities to show their talent. Skills that are important for modern careers may be undervalued in favor of test-taking skills, and activities that are enjoyable for teachers and students may be reduced (Berliner, 2011).

Additionally, use of student test scores and VAMs to evaluate teachers can threaten teacher morale. According to Baker et al. (2010), “pressure to raise student test scores, to the exclusion of other important goals, can demoralize good teachers and, in some cases, provoke them to leave the profession entirely” (p. 15). Darling-Hammond et al. (2012) noted that in a study of teachers in Tennessee, “after three years, 85% thought the VAM evaluation ignored

important aspects of their performance that test scores didn't measure, and two-thirds thought VAM didn't do a good job of distinguishing effective from ineffective teachers" (p. 12). If teachers do not buy into the programs, it can certainly lead to job dissatisfaction. Moreover, teacher morale is especially at risk in parts of the country like Los Angeles, CA, where the value-added scores of teachers along with their names, and in some cases their pictures, were published in the *Los Angeles Times* (Sparks, 2011).

Alternative Student Growth Models

Due to the concerns about VAMS that are outlined in the research and new flexibility for state teacher evaluation policies under ESSA, educators and policymakers have looked for more effective and less controversial ways to implement student assessment data into teacher evaluation (Education Commission of the States, 2018). Arkansas, Florida, Indiana, Kentucky, Michigan, South Carolina, and Utah passed legislation in 2017 that allowed for more local district flexibility in determining student growth measures (Education Commission of the States, 2018). Stiggins (2014), a leading researcher in classroom assessment and a proponent of using student assessment data in teacher evaluations, is opposed to the use of standardized test scores for teacher evaluation. However, Stiggins and Berliner (2018) reported that classroom assessments, used appropriately, can be the answer to including test information into evaluation processes. Stiggins outlined five criteria that must be met in order to effectively use classroom assessment data in teacher evaluations.

1. Specific academic standards must be agreed upon by the evaluator and teacher.
2. Those standards must align with the teacher's normal instructional responsibilities.
3. Each standard should have an assessment plan and high-quality assessments.
4. Assessments must be conducted in a pre-test/post-test format.

5. Teachers should have the opportunity to explain things that impacted the results.

There are a variety of student growth models now in place across the United States that attempt to strike a balance between the purely statistical VAMs and classroom level assessments. The federal government recommended that states and districts allow the use of Student Learning Objectives (SLOs) for teachers in non-tested subjects and grades (Stiggins, 2014). In districts that employ SLOs, teachers give a pre-test, predict the growth that students will make through their classroom learning experiences, give a post-test, and then are rated on the percentage of students who meet their growth targets. Tucker and Stronge (2005) described systems employing SLOs in the Thompson, Colorado School District and in the Alexandria, Virginia School District. In these districts, teachers were allowed to select their assessments from a listing of district-approved assessments and then demonstrate growth through a pre-test to post-test approach. They were not required to predict their students' growth in this model. By 2014, in a shift from the use of VAMS, SLO processes had been outlined for the cities of Austin, Texas Denver, Colorado, and Houston, Texas and the states of Washington, Oregon, Louisiana, Rhode Island, Georgia, Indiana, Ohio, New York, Maryland, Wisconsin, and Illinois (Stiggins). There is less research regarding the use of SLOs in teacher evaluation than there is regarding the use of VAMs.

Some districts have adopted student growth measurement models in which the whole school or entire grade levels adopt the same student growth goal instead of each individual teacher developing goals. This all-in approach ensures that the group is focused on key goals and simplifies the process of selecting assessments and writing goals for both the evaluator and the teacher. However, under an all-in approach, teachers could potentially be evaluated on the performance of students who they do not teach and/or in subjects that they do not teach.

Teacher Evaluation Reform in Illinois

Illinois was one of several states that implemented major reforms in learning standards, student assessment, and teacher evaluation in an effort to garner Race to the Top funds. In 2010, Governor Pat Quinn signed the Performance Evaluation Reform Act (PERA; Growth Through Learning, 2012). PERA required that all Illinois public school districts develop teacher, principal, and assistant principal evaluation systems that assess professional practice and also incorporate measures of student growth. Districts had to create clearly established standards of professional practice in order to “add objectivity to a practice that almost universally was subjective” (p. 1). Under PERA, districts had to form joint committees of teachers and administrators to collaboratively develop new evaluation systems. Student growth had to account for at least 30% of teachers’ final ratings, and the law specified that incorporation of student growth applied to all teachers with the exception of support personnel like counselors, social workers, speech and language pathologists, psychologists, and school nurses (ISBE, 2010). Joint committees that were unable to reach an agreement on a new system would default to the state’s plan, which prescribed Charlotte Danielson’s Framework for Effective Teaching (Danielson, 1996) and mandated that 50% of the teacher’s summative rating would be based on student growth on assessments (ISBE, 2010). PERA required that non-tenured teachers who were rated as excellent or proficient would be evaluated every other year while non-tenured teachers and tenured teachers who were rated as needs improvement or unsatisfactory would be evaluated every year (Growth Through Learning).

Additionally, teachers had to be rated as either excellent, proficient, needs improvement or unsatisfactory, and the percentage of teachers who were rated as proficient or excellent would be published as part of the annual school report card (ISBE, 2010). However, teachers would not

be identified individually as part of the report card. Reduction in force (RIF) procedures were also drastically changed under PERA. Instead of the traditional system of dismissing teachers solely based on seniority, PERA required that teachers be placed into four categories for RIF decisions. The first category included part-time teachers, those who had not be evaluated yet, or those who were covering for a leave of absence. The second category included teachers who had received an unsatisfactory or needs improvement rating on one or both of their last two evaluations. The third category consisted of teachers who had at least a proficient rating on the last two evaluations, and the fourth category was teachers who had excellent ratings on their last two evaluations or excellent ratings on their last two out of three evaluations, with the third being proficient. Teachers would be dismissed starting with group one and then dismissals would proceed to groups two through four (ISBE).

The deadline for implementing the new evaluation systems was different for various districts across the state. All of the Chicago Public Schools had to be compliant by September 1, 2013, and Illinois districts whose students' academic performance fell in the bottom 20% of schools of their type had to be compliant by September 1, 2015 (ISBE, 2010). All remaining schools in the state had to implement the changes by September 1, 2016. Administrators were required to complete formal training prior to evaluating teachers, principals, or assistant principals. (ISBE).

A notable component of PERA is that local districts retained control of the design of the evaluation systems as long as they met the state's broad criteria. The result of this local control is that districts in Illinois have adopted a variety of approaches to teacher evaluation, including SLOs and all-in approaches.

Teacher Perceptions of the Use of Test Scores in Teacher Evaluations

In response to the reform efforts of the early 2000s that encouraged the use of standardized tests in teacher evaluations, there was an outpouring of research pointing out the flaws in rating teachers based on students' standardized test performance (Amerein-Beardsley, 2014; American Statistical Association, 2014; Baker et al., 2010; Gabriel & Allington, 2012). As a result, many districts and states have moved away from using standardized test scores and incorporated different kinds of student growth measures into blended teacher evaluations. Now that these reforms have been implemented, it logically follows that the effects of the new models should be assessed. Some research does exist on the use of standardized test scores in teacher evaluation, but less research is found regarding models that use other student growth approaches. After reviewing the research of teacher perceptions of various types of VAMS and student growth models, key themes emerged, including unfair evaluation criteria, teacher stress, lack of clarity in the evaluation criteria, and influence on teaching practice. It is important to note that much of the research covers teacher evaluation reforms that involve both the professional practice component combined with various forms of VAMS and student growth models.

Unfair evaluation criteria. Fairness in the evaluation process was a consistent theme among teacher groups. Crystal (2014) conducted a qualitative study in Connecticut using questionnaires and interviews that was aimed at determining how teachers perceived a new evaluation system that implemented a VAM that accounted for 45% of their teacher summative ratings. Teachers in this study reported that they were not opposed to evaluation, but they wanted the evaluation to be valid and reliable (Crystal). Jiang, Sporte, and Luppescu (2015) reported similar teacher sentiments in their research in Chicago, Illinois. Their research included 32 interviews and survey data from 12,000 teachers per year from 2012-2013. While the teachers

who were surveyed reported positive feelings about the observation component of their new evaluation system, 65% perceived their evaluations relied too much on student growth data, and half did not feel that the assessment being used to measure student growth was fair (Jiang et al.).

The main theme regarding fairness was the influence that factors other than the teacher's classroom practices can have on test scores. Pressley, Roehrig, and Turner (2018) surveyed and interviewed 13 elementary school teachers across five school districts in Florida in their qualitative case study. Respondents reported that they did not feel that their scores represented their teaching or their students' learning due to the lack of accounting for issues that students face outside of school that are the teacher's control. In a qualitative interview study in Ohio, nine middle school English and mathematics teachers expressed similar concerns. Specific variables affecting student performance that they mentioned were lack of sleep, inadequate nutrition, hormones, poor attendance, social relationships, and relationships with teachers (German, 2014). Jiang et al. (2015) found that teachers in Chicago also felt that student performance was impacted by factors outside of school as much as by teaching practices, so evaluating them based on test scores was not completely fair to the teacher.

Another issue related to unfair evaluation criteria was revealed in schools where teachers were evaluated based on test scores outside of their teaching area. Jiang et al. (2015) noted that Chicago Public Schools were using school-wide literacy value-added scores that teachers felt were unfair and invalid. In addition to concerns about being evaluated using scores outside their areas of expertise, teachers reported that VAMs do not measure important skills that they need to teach in order for students to be successful. For example, character education and how to write a five-paragraph essay are important for students, but are not measured on VAMs (German, 2014).

Teachers in Chicago also noted the potential for cheating among teachers using VAMs when the systems permit teachers to score their own pre-tests and post-tests (Jiang et al., 2015).

Increased teacher stress. Teachers consistently reported increased stress, anxiety, and tension related to student test scores in teacher evaluation across studies (Crystal, 2014; Foutch, 2017; German, 2014; Hertel, 2014; Jiang et al., 2015; Pressley, 2018). However, it is important to note that many studies were conducted during the implementation phase of new evaluation models. It is likely that teacher emotions are higher during the early implementation phase, when the effect of the new evaluation systems is not fully known, and teachers are still learning about the process. The fact that the stress and anxiety linked specifically to learning a new system is now finished, is another reason that an updated study of teacher perceptions post-implementation is needed.

Jiang et al. (2015) found in their 2014 survey that 79% of teachers felt stress or anxiety related to their teacher evaluation process. German (2014) studied the effect of value-added scores on teachers' perceived self-efficacy in Ohio; teachers noted that it sometimes took them days or even a week to emotionally recover from receiving their VAM reports. Even teachers who received the highest rating did not report any feelings of pride or joy. Several teachers in German's study had experienced both high and low VAM ratings and did not feel that they had performed differently as a teacher when their scores had declined. Although the VAM scores fluctuated, they believed that their professional practice was consistently effective.

Jiang et al. (2015) found that 60% of teachers in their study in Chicago felt that the process took more time than the results were worth. Foutch (2017) also commented on the large amount of time that was involved in the new processes as a stress factor. Foutch's interview study of 14 teachers in five districts focused on the Performance Evaluation Reform Act in

Illinois and how teachers felt about both the new teacher observation regulations as well as the student growth component. His findings revealed that teachers felt that the new system took time away from planning for students' needs and focused on the new mandates instead. Foutch also noted that schools have a limited amount of time for professional development each year and need time to work on local goals as well as mandated reform initiatives. This issue is less likely to be a concern in Illinois now that the implementation phase is over; however, ongoing professional development time for teacher evaluation is likely to be necessary as teachers move through "on" and "off" cycles of evaluation. Non-tenured teachers must be evaluated every year, but tenured teachers are only required to be evaluated once every 2 years. Tenured teachers may not remember the nuances of the evaluation process after having a full school year without formal participation in it. Hertel (2014) also studied teacher evaluation in Illinois as it relates to the implementation of new policies related to teacher observation using the Danielson Framework; 84.4% of teachers and 52.8% of administrators agreed that teacher evaluation raises tension and anxiety.

Stress levels have been found to be especially high for teachers when scores are publicly reported. In 2012, the New York City Department of Education released individual value-added reports for 18,000 teachers (German, 2014) In Florida, teachers expressed concerns that parents and the public would not understand the scores and how they were calculated and that they would be unfairly judged in the public eye (Pressley et al., 2018). They worried that parents would not want their children in their class if their scores were low and that their parent relationships would be negatively impacted.

Foutch (2017) reported relational stress and frustration related to the new system in Illinois, with teachers perceiving the growth requirement was harming their relationships with

colleagues. Since Illinois implemented a new policy where reduction in force (RIF) was based on teacher ratings instead of seniority, teachers became more competitive and less willing to collaborate. Since they were in competition with their peers, they were less likely to share and work together. Foutch reported that in one school district in Illinois teachers felt that the new evaluation system was pushing the wrong teachers out of the profession.

Teacher understanding of scores. Teachers across many studies reported that they did not fully understand their value-added scores (German, 2014; Pressley et al., 2018; Jiang et al., 2015). Ohio teachers reported that they had a sufficient understanding of the scores to use them to evaluate their own practice, but they did not understand the statistical part of the VAMs, including the predicted growth calculation. This lack of clarity made them less trusting of the scores, and they stated they would probably trust the scores more if they understood the statistical model (German, 2014). Teachers in Florida did not always understand their VAM scores and did not feel that their evaluators were able to accurately explain their scores to them. Teachers in Chicago were not sure how much student growth even counted toward their evaluations (Jiang et al., 2015).

Impact on curriculum on and instruction. Although it was not a predominant theme in the research, teachers expressed various ways that value-added evaluations have impacted their teaching practices. German's (2014) study in Ohio revealed that value-added scores caused some teachers to re-evaluate how time is used in the classroom, but the motivation to change was to improve the learning for students and not to increase the VAM score. The same study found that teachers increased their use of formative assessment, increased the teaching of test-taking skills, and strived to develop improved relationships with students. Another by-product of VAMS was

that some teachers included less discussion, less debate, less project-based learning, and more constructed-response questions (German, 2004).

A 2-year survey study in New Jersey asked teachers how much the new ACHIEVENJ evaluation, which combined professional practice observations with student test data, had changed their teaching practices in the areas of teaching students with special needs, the emphasis placed on raising student test scores, how teachers handle student discipline, teachers' knowledge of the subject and pedagogy, and classroom management. More than half of the 618 respondents reported that they made no changes in these areas based on their 2012 and 2014 evaluation scores (Callahan & Sadeghi, 2015).

Pressley et al. (2018) found that in Florida teachers reported feeling pressure to narrow the curriculum in order to teach to the test. The teachers said that they were given pacing guides that started in January in order to prepare for standardized tests. The 90-page test-prep books sometimes took the place of more engaging activities. Additionally, teachers reported that Social Studies and Science classes had been eliminated except for when the concepts were involved in Reading and Writing. In the same study, teachers in another district in Florida were not required to make scripted changes to curriculum and instruction. Instead, they were allowed to implement test-taking strategies and review relevant standards with autonomy. These teachers had more positive feelings towards the process than those in the district who were given specific mandates for test preparation (Pressley et al.).

Positive outcomes related to student growth measures. The data on teacher perceptions of student test data in evaluations is mostly negative. However, there were some positive areas. Ohio teachers reported that they used the value-added scores as a way to assess themselves and their professional growth. They felt that being evaluated with these scores has

made them more reflective as teachers (German, 2014). Crystal (2014) noted that teachers consistently mentioned that the necessary increased communication between the teachers and evaluators under the new system in Connecticut was viewed as positive. This positive reaction to more evaluator time spent in the classroom was echoed by teachers in Chicago (Jiang et al., 2015) and Florida (Pressley et al., 2018). However, it is unclear whether the increased time with evaluators was solely part of the professional practice observations or part of the value-added portion of the evaluations.

Theoretical Framework

The relationship between principals and the teachers who they evaluate is likely to influence the effectiveness of the evaluation process, since principals are generally solely responsible for carrying out teacher evaluations. According to Bryk and Schneider (2002), there is a complex web of social interactions within schools that directly impacts daily operations as well as improvement efforts. The people within the school depend on each other to engage in cooperative efforts around local problems. Schools function most efficiently when trusting relationships exist between school leaders, teachers, students, parents, and the community. Each of these groups have obligations related to the school and they rely on the other groups to fulfill their own obligations for the mutual benefit of all groups. When this trust is not present, schools and communities can become mired in conflict and reform efforts become stagnant.

Bryk and Schneider (2002) developed the theory of relational trust based on their work studying reform efforts in the Chicago Public School system. They define it as a three-level theory that involves the (a) intrapersonal process of determining the intentions of others; (b) interpersonal roles within the context of the culture, history, climate, and local norms; and (c) impact of relationships on decision-making, support for change, and improvement (p. 22). People

within the school are constantly determining whether the actions of others are in line with their own interests and assessing the intentions of others. When they have a history of interaction, they base current perceptions on past experiences. In the absence of background information, people are likely to make assumptions based on reputation or social categories (Bryk & Schneider, 2003). Bryk and Schneider identified four areas that usually play a role in how people determine the intentions of others: respect, personal regard for others, competence in core responsibilities, and personal integrity.

Respect. This component of relational trust theory starts with actively listening and attempting to understand the perspectives and opinions of others. It does not require that all parties agree and have the same opinion, rather that all parties feel heard and understood. Furthermore, relational trust is improved if the perspectives of others are considered when making decisions in the future. Bryk and Schneider (2002) gave examples of respect in relationships that include parents trusting that teachers will consider their opinions about the education of their children, teachers trusting that principals will consider their perspectives before making decisions, and principals trusting that teachers will embrace important changes that are needed to improve the school. Respect as a component of relational trust relates to teacher evaluation, especially with regard to student growth scores. As noted in previous sections, teachers have concerns about test scores being used in their evaluations. They are unlikely to improve from the evaluation process if they do not feel that their evaluator respects these concerns and is genuinely interested in helping them improve.

Personal regard for others. Personal regard for others means that people within the school community show that they care about others and are willing to go beyond the normal scope of their responsibilities to help those in the school or community. Relational trust is

stronger as people begin to believe that others care about them as individuals. This regard for others is especially powerful when it involves a person in a position of power. Principals who show genuine concern for teachers are likely to build relational trust that helps teachers develop a sense of identity and community related to the school. This is likely to encourage them to work beyond the scope of their identified responsibilities when they know that it is likely to help the school to improve. A team mentality begins to take the place of an individual mentality.

Personal regard for others relates to the principal's role in teacher evaluation as the principal must ensure that the process is carried out in a way that is fair to the teacher. There are obligations within teacher evaluation for both the evaluator and the teacher. The principal must not only fulfill their own portion of the work for the teacher evaluation but must also ensure that the teacher understands the process and is supported throughout the cycle. It is even more important for the principal to guide the teacher and make sure they understand the system when it involves a new initiative or reform that is likely to cause the teacher anxiety or confusion.

Professional competence. In a survey study involving 610 teachers and 37 schools in Belgium, Tuytens and Devos (2010) found that teachers had a more positive perception of their new teacher evaluation system when the principal provided structure for the program and established clear standards and steps. People within the school community make judgements about the competence of others within the system. Teachers are judged by principals, parents, and students based on their instructional approaches, curricular choices, classroom management, and behavior management. Principals are judged based on their organization of the school and systems within it, test scores, and public perception of the school. Parents are judged on their efforts to support their students and meet their basic needs. Relational trust depends on the perception that others within the system are competent and working toward common goals.

Parents must trust those in the school to teach their children the knowledge and skills that they need. Teachers rely on the principal to organize the school, systems, and schedule so that they can achieve their goals. Principals rely on teachers to carry out their duties effectively. When there is a perception of incompetence it is damaging to the relational trust within the school community. This issue of competence relates to teacher evaluation because principals must ensure that teachers who are not competent are remediated or released. If ineffective teachers are allowed to continue practicing, trust erodes. Colleagues of the ineffective teacher will not want to collaborate with that person and parents will lose trust in the school to educate their children. Additionally, competent teachers must be acknowledged and recognized so that those within the system recognize effective teaching.

Personal integrity. Relational trust depends on the perception that others within the system can be trusted not only to do what they say they are going to do but also that they can be trusted to do the right thing. Trust can quickly erode if people do not keep their word or if there is a perception that they are not working for the best interest of the school, even if that means making difficult decisions or initiating difficult reforms. Principals show integrity when they make decisions based on what is best for student learning, rather than what is easiest or likely to encounter the least resistance. Integrity is important in regard to teacher evaluation. Teachers must believe that the principal will carry out the evaluation in a fair way that will benefit both the teacher and the school. This was evident in the study by Calabrese, Sherwood, Fast, and Womack (2004) in which the teachers questioned if the principals were avoiding high ratings at the direction of upper administration, and principals were afraid to rate teachers as needing improvement because teachers take it so personally.

In cases where teachers establish student growth targets and select their own assessments, principals must trust that teachers will choose fair and rigorous assessments and set appropriately high growth goals for their students. Additionally, parents and community members must believe that the principal will act if there is a teacher who is ineffective or who has done something that is detrimental to students.

Importance of relational trust in schools and teacher self-efficacy. Bryk and Schneider (2003) found through their research in Chicago that a school with a low score on relational trust had only a one-in-seven chance of showing academic improvement. Additionally, half of the schools with high relational trust scores showed improvements on academic measures. Schools with consistently low relational trust scores showed almost no improvement in reading or math scores. This growth is attributed to the collective decision-making and teacher buy-in that results from strong relational trust. Tuytens and Devos (2010) found that the level of trust that teachers had for their principal influenced their perceptions on the practicality of their new teacher evaluation system.

Forsyth, Barnes, and Adams (2005) pointed out the importance of the trust relationship between teachers and the principal for fostering cooperation and collaboration. “The level of teacher trust for the principal is shaped by behaviors of the principal, observed by teachers, that show he/she is respectful, benevolent, competent, reliable and honest” (Forsyth et al., p.125). Relational trust between teachers and their principals has consequences for the teachers’ self-efficacy. Tschannen-Moran, Hoy, and Hoy (1998) explained that teacher self-efficacy means how a teacher perceives their own competence. Bandura and Wessels (1997) suggested that, “People with high assurance in their capabilities approach difficult tasks as challenges to be mastered rather than as threats to be avoided” (p. 1). This self-perception impacts the effort that

they put forth and how well they utilize the skills that they possess. These teachers set lofty goals and persevere to achieve them. They feel that they are in control and can have an impact on their students.

Importance of relational trust in schools and teacher evaluation. A teacher evaluation system that considers a teacher's self-efficacy and maintains or builds relational trust is more likely to benefit the school and the students. Given the current perceptions of using student test scores in teacher evaluation, protecting trust and teacher self-efficacy may be especially important. In this study, I sought to determine how the trust between the principal and the teachers who are being evaluated influences the level of engagement of the teachers with the student growth component of their evaluation.

Conclusion

In summary, teacher evaluation should be a lever for improving teaching and learning. Traditional models of teacher evaluation were very subjective and generally resulted in most teachers achieving a satisfactory rating. Federal education policy began to shift towards using student test scores to hold schools and then teachers accountable for student achievement around the beginning of the 21st century. In order to get federal money, schools across the country implemented teacher evaluation systems that included student test scores. Various approaches to incorporating student test data have been developed, most notably value-added models and student learning objectives. Researchers have raised serious concerns with using student test data to make high stakes decisions about teachers, especially in regard to value-added models. Limited research shows that teachers also have reservations about the use of student test scores in their evaluations. It is important to assess the impact that the student growth component of teacher evaluation has had on teachers now that the implementation phase of the reform has

passed. Assuming that relational trust and teacher self-efficacy are key components of operating a successful school and implementing continuous improvement within a school, a study of how the student growth model in Illinois has influenced the practice of teachers in elementary schools would provide valuable information to principals and policymakers.

Chapter 3

Methodology

Illinois public school districts have implemented PERA-aligned evaluation systems that incorporate student test data, through a phased-in approach, beginning in September 2012 (ISBE, 2010). By September 2016, all Illinois districts were required to have evaluation models that incorporated PERA mandates. Research is needed to determine the effect of these reforms. Because local districts were allowed some discretion with regard to how PERA reforms were implemented, there are various teacher evaluation plans in place in Illinois now. Therefore, research data related to the effectiveness of specific models can help inform future best practices for districts and also guide principals toward implementing a worthwhile evaluation program. The review of the literature found that teachers generally viewed the use of value added models for evaluation purposes negatively, due to concerns over validity, reliability, lack of clarity in the evaluation criteria, and unintended consequences. There is less research regarding perceptions of teachers on alternate student measurement models like SLOs and all-in approaches.

Bryk and Schneider (2002) suggest that social interactions between people within schools have a significant effect on the daily operations of the school and on improvement efforts within the organization. Schools function most efficiently when trust is developed and maintained between teachers, administrators, students, and parents (Bryk & Schneider). This proposed study will examine how elementary teachers and principals in two different school districts perceive the teacher evaluation processes that have been adopted by their districts and will seek to explore the role of relational trust between the teacher and principal in an effective evaluation process.

In this chapter, I present my research questions in order to illustrate the research design that I selected for this study. I then describe the qualitative approach, site and sample selection

process, data collection, and data analysis. I also discuss validity, reliability, and trustworthiness as well as ethical considerations.

Research Questions

This qualitative interview study integrates the framework of relational trust and is based on the following research questions.

1. To what extent do Illinois public school teachers who are evaluated using student learning objective and all-in student growth approaches perceive the processes to be understandable, fair, and valid?
2. To what extent have these student growth approaches influenced their professional practice?
3. How does the principal's approach to teacher evaluation during the student growth component of the evaluation influence teachers' professional development?

Research Design

This study uses a social constructivist qualitative design. Creswell and Poth (2018) defined qualitative research as inquiry that uses a conceptual framework to study the meaning that people get from a problem by observing or studying the people in the setting in which the problem occurs. In the social constructivist framework,

individuals seek understanding of the world in which they live and work. They develop subjective meanings of their experiences—meanings directed toward certain objects or things. These meanings are varied and multiple, leading the researcher to look for the complexity of views rather than narrow the meanings into a few categories or ideas. (Creswell & Poth, p. 24)

Teachers and principals will have complex and varied views and perceptions of teacher evaluation processes. This study strives to understand complex viewpoints with multiple meanings and understandings based on the lived experiences of the teachers and principals.

This qualitative interview study used purposeful sampling to find two elementary schools in separate Illinois public school districts that have implemented PERA-aligned teacher

evaluation systems incorporating student growth data. One district utilizes the student learning objective approach and one uses the all-in approach.

Site and Participant Selection

This study was conducted in the state of Illinois, with two public school districts. Sites were selected using random purposeful sampling of public primary schools (Creswell & Poth, 2018). The purposeful component of the sampling focused specifically on recruiting one primary school that uses an SLO process for the student growth component of teacher evaluation and one school that uses an all-in approach. Only two schools were identified that used the all-in approach. One was eliminated because it had just started implementation of the all-in model this school year. The other was recruited and agreed to participate. Many schools that use SLOs were identified and five that are located within central Illinois were contacted. Two principals initially agreed to participate and then one was unable to recruit teachers. Selecting sites that use different models allowed for comparison between the two approaches. The study was limited to elementary schools in order to illustrate the experiences of teachers of younger students, including those whose students are not tested using standardized tests.

Potential sites were identified by contacting the superintendent of Regional Office of Education #39 (ROE 39) in Illinois to help identify districts that used the desired approaches to teacher evaluation. ROE 39 was selected due to proximity to where I live. I also researched school district websites in order to determine which schools met the criteria of the study. I sought two schools that were similar in size and teacher and student demographics. I excluded Chicago Public Schools (CPS), since studies have already been completed there, and there are no comparable districts to CPS in Illinois. I contacted districts that were recommended by calling superintendents in order to determine if the district met the student growth criteria of using an

SLO approach or using an all-in approach. If the superintendents were willing to allow district participation, I then contacted building principals via phone call. I selected principals who had served in the district prior to implementation of PERA so that they could provide insight into teacher evaluation at the site both before and after the new policies and procedures were implemented. Selection of schools was dependent on the principals' willingness to be interviewed and also to allow teachers in the building to be interviewed.

Candidates for teacher interviews were not limited based on rating or years of experience, but I sought out at least five teachers from each building for individual interviews. I chose not to limit the type of teachers recruited, because differences in perceptions may arise from different levels of experience or rating levels, and such trends in the data could be valuable for principals or evaluators to understand. Therefore, I included veteran and novice teachers and those who have had both positive and negative ratings with student growth. I also conducted interviews with the building principals in order to gain insight into their role in carrying out the teacher evaluation process and how the principal leads the teacher evaluation process might impact teacher perceptions. I interviewed two principals and nine teachers from two different schools. Interviews took 30-40 minutes. I now discuss contextual aspects of each of the two schools and the participants.

Monroe Elementary School. Monroe Elementary School is a small, rural Illinois school that houses 275 students in grades pre-kindergarten through sixth grade (ISBE, 2020). The following statistics were retrieved from Illinois Interactive Report Card from the 2018-2019 school year, but reference to a specific website is not included in order to protect the anonymity of the school and district. Ninety percent of students in Monroe School are white, 2% are black, 3% are Hispanic, and 5% are two or more races. The most recent student test data results show

that 50% of students met or exceeded expectations on the English Language Arts test, 30% met or exceeded expectations on the Math test, and 90% met or exceeded expectations on the Science test. Forty percent of students are considered low income, 10% have individualized education plans, and 1% are homeless. There are 40 teachers employed at Monroe School, and the teacher retention rate was 90% over the last 3 years. One hundred percent of teachers were rated as proficient or excellent during the 2018-2019 school year (ISBE).

The principal and four teachers participated in the research study at Monroe Elementary School (see Table 1).

Table 1

Participant Profile of Monroe Elementary School

Name	Gender	Position	Years in position	Grade
Samantha Douglas	Female	Principal	10	PreK-6
Lexie Jenkins	Female	Teacher	24	3
Stefanie Clark	Female	Teacher	16	K
Kim Tracy	Female	Special Education teacher	6	K-6
Sarah Smith	Female	Teacher	6	PreK

Roosevelt Elementary School. Roosevelt Elementary School is a rural school that serves 325 students in pre-kindergarten through fifth grade (ISBE, 2020). The following statistics are from the Illinois Interactive Report Card for the 2018-2019 school year. The specific link is not referenced in order to protect the anonymity of the district and participants. Ninety percent of students were white, 2% were black, 4% were Hispanic, 1% were Asian, 3% were two or more races, and 1% were Pacific Islander. Sixty percent of students were considered low income, 3% were English Language Learners, 20% had individualized education plans, and 1% were homeless. The most recent test data results show that 30% of students met or exceeded expectations on the English Language Arts test, 40% met or exceeded expectations on the Math

test, and 50% met or exceeded expectations on the Science assessment. Roosevelt Elementary School employs 100 teachers, and 99% of them were rated proficient or excellent during the 2018-2019 school year. The school retained 90% of their teachers over the last 3 years (ISBE).

The principal and five teachers participated in the research study at Roosevelt Elementary School (see Table 2).

Table 2

Participant Profile of Roosevelt Elementary School

Name	Gender	Position	Years in position	Grade
Justin Foley	Male	Principal	12	PreK-5
Cati Mosser	Female	Special Education teacher	5	3-5
Jamie Jones	Female	Math intervention	13	2-5
Jack Taylor	Male	Teacher	2	4
Grace McCarter	Female	Special Education teacher	2	5
Allison Kaiser	Female	Teacher	8	3

Data Sources

The majority of data for this study was collected through semi-structured open-ended interviews with principals and teachers. Individual interviews were conducted in a face-to-face setting. Sitting in the same room with the interviewees helped the interviewer to establish rapport. Avoiding group interviews encouraged more open sharing since teacher evaluation is a personal subject, and participants did not have to worry about potential problems in their workplaces if they criticized processes or people during the interview (Simons, 2012). An interview protocol was developed, and pilot interviews were conducted based on access and convenience in order to help clarify and refine the interview questions (Creswell & Poth, 2018). The interview protocol was refined and reorganized after the pilot interviews (see Appendix F). The pilot participant suggested eliminating the questions about experience at the beginning, because they were redundant after asking the ice breaker questions. The questions about the

participants' participation in the district joint committee to implement PERA reforms were moved to the top of the list, because this allowed the participants to explain their history and experiences with teacher evaluation reforms in chronological order. Interviews were audio recorded and then transcribed electronically using an online application called TEMI.

Records and artifacts were reviewed in order to supplement the interview data (Creswell & Poth, 2018). These documents included district teacher evaluation handbooks, Education Consulting Research Analytics Group (ECRA) information,¹ and documents used in the teacher evaluation processes at each school. The teacher evaluation manual for Roosevelt School was available online and the principal at Monroe School supplied the Monroe evaluation handbook. ECRA information was accessed online, and two teachers at Roosevelt School shared their goal worksheets.

Data Analysis

The steps for effective data analysis include organizing the data, conducting a preliminary reading of the data, coding data around themes, showing the data, and developing an interpretation (Creswell & Poth, 2018). In order to organize the data, digital files containing interview recordings and transcripts were stored in a secure, password-protected, University of Illinois Box.com account. All files were named starting with a pseudonym for the teacher. Documents that were collected from the sites were stored in a locked file cabinet in my office.

As data were collected, I reviewed the responses to become familiar with the data (Tesch, 1990). I read and edited the transcripts for typos and minor errors shortly after each interview. As I read, I wrote memos about big ideas that stood out to me in each interview. These memos were largely related to comparisons between respondents at the two different schools, but also

¹ ECRA is a consulting firm that supports districts with advanced data analysis techniques.

included overarching themes that I noticed, like the importance of time to teachers. This memoing was used throughout the data analysis process in order to help identify patterns and themes (Miles, Huberman, & Saldana, 2014). I also wrote a brief summary of responses for each participant. I then made a spreadsheet that listed interview questions across the top row and had participants in the first column. I wrote summaries of the participants' answers to each question or included a quote from them in the column for each question. This allowed for easy comparison between respondents and schools.

The theoretical framework for this study was relational trust. The coding and analysis of the data was aimed at determining the relationship between teacher perceptions of the student growth model of evaluation and the trust level that teachers share with their principals through the process. I established deductive codes based on concepts from the theoretical framework, including principal trust, teacher trust, and communication, and from the literature on teacher evaluation, including fairness, accuracy, and time. While the initial codes were framed around relational trust and the literature, I also included codes based on what I saw in the data that were not reflected in my initial codes. There were themes that came up in the data that were not anticipated, but were still valuable to the study. These included artifacts and data usage, which ended up being prominent codes across respondents in both schools. I coded the data independently, and my dissertation advisor, Dr. Rachel Roegman, reviewed my work.

Once the data were coded, I sorted the data by code so that I could look at all of the participant responses based on each code. I had a spreadsheet for each code, a spreadsheet for each participant, and a spreadsheet of summarized answers from all participants. Once I had analyzed all of the coded data, I developed pattern codes that allowed me to categorize the data. These pattern codes were developed by sorting the initial codes into groups of related ideas,

including how teachers understand the process, fairness and accuracy, time and value, teacher preference of processes, influence of SLO and all-in on teaching, influence of SLO and all-in on collaboration, relationships before during and after PERA, principal steps in evaluation, and teachers value feedback. I then grouped these pattern codes based on my research questions, and after analyzing the data sorted in that manner, I arrived at three major findings that involved comparisons between the two schools and their approaches (see Appendix G).

Validity, Reliability, and Trustworthiness

Creswell and Poth (2018) stated that qualitative researchers must use a process in order to assess the accuracy of their findings. The authors recommended nine strategies for ensuring validity and suggested that researchers should use at least two of these strategies. The strategies are divided into three categories: (a) researcher's lens, (b) participant's lens, and (c) reader's lens.

I established validation through the researcher's lens by clarifying researcher bias. My positionality and biography are included later in this section. I ensured validity through the participant's lens by using member checking to verify data with the participants. To do this, I shared summaries of the interviews at each site to ensure that the data was in line with what the subjects reported (Miles et al., 2014). The information sent to the respondents for clarification and review included a bulleted list summarizing their views on the major topics for the study. No corrections or clarifications were needed based on the responses of the participants. From the reader's lens, I used the peer review strategy. I asked a qualified researcher, my dissertation advisor, Dr. Rachel Roegman, to review my work throughout the study (Creswell & Poth, 2018). Yin (2009) suggested that researchers must keep careful record of their procedures in order to

validate their research. I established a chain of evidence to allow my readers to trace my steps back and see how patterns were developed and conclusions were drawn.

Booth, Colomb, and Williams (2008) stated that research is reliable when it is “sufficient and representative, reported accurately and precisely, and taken from an authoritative source” (pp. 135-136). They suggested that researchers must take great care to eliminate even small mistakes that the reader might take as indicators of carelessness or a lack of reliability of the research. I checked data accuracy as it was collected and clarified it afterwards with member checks in order to ensure that the research study was reliable. It is also important to use precise language to provide evidence in qualitative research (Booth et al.). In this study, I provided accurate and precise statistics and data when making claims.

Trustworthiness is an important part of establishing the validity of a qualitative study (Creswell, 2014). Researchers can gain the trust of readers not only by putting forth accurate data and claims based on evidence, but by acknowledging differing viewpoints (Booth et al., 2008). In this study, I shared specific quotes from all respondents in order to allow the reader to have access to the raw data. Trustworthiness can also be established by including data that does not fit well into the researcher’s claims or arguments and explaining its relevance. As I analyzed data and drew conclusions for this study, I carefully considered data that did fit the pattern and clarified what was representative of the sample of teachers and what was outside the norm. I shared any outliers in the research as well as the more common themes.

Researcher Positionality

As a researcher, it is also important that I reveal my positionality in relation to this study. I was a public high school teacher in Illinois for 6 years. During that time, I was evaluated under a traditional model that included one observation per semester and a subjective rating system that

resulted in teacher summative ratings of satisfactory or unsatisfactory. I did not find tremendous value in my teacher evaluation, other than that I appreciated the acknowledgment of my efforts and encouragement that I was doing a good job.

I became a dean of students at that high school for 1 year before accepting a principalship at a neighboring school that served 100% K-12 special education students whose behavior led to their change of placement from their public schools. This principalship was my first experience formally supervising and evaluating teachers. I used a subjective checklist that resulted in teachers either being rated as satisfactory or unsatisfactory. It was a traditional approach that called for one observation per semester for non-tenured teachers; there were no tenured teachers in my school. I found it difficult to rate teachers based on the checklist and using limited observational data. After 1 year, I accepted a principal position in a small rural elementary school, where I have been employed for the last 11 years. Initially, the evaluation tool was very similar to what I had previously experienced. It was a checklist with what amounted to “yes” or “no” questions. After PERA was enacted, I was tasked with leading the district’s joint committee to develop a new teacher evaluation system. We adopted the Charlotte Danielson Framework first and then implemented student growth measures the following year. I did extensive research during that time to determine what other districts were using and what the research recommended. Our district implemented an SLO approach that allows teachers and principals to agree on assessments pertaining to the specific subject area that the teachers teach ,and then the teacher and principal set goals jointly.

As a principal, I am evaluated using a professional practice rubric and through an SLO process for student growth. I have not found much value in the student growth component of the evaluation, but it also does not cause me stress or worry. I review student data regularly as we

plan instruction, but after 18 years in education I am still not able to reliably predict what scores my students are likely to achieve on their post-tests. Due to my personal familiarity with teacher evaluation models and evaluations involving student growth measures, it will be important to have a colleague who has not had personal experience with teacher evaluation review the data and codes so that my potential bias could be identified and eliminated. I will also need to carefully review my analysis and search for signs of my own bias so that it can be removed.

Ethical Considerations

Prior to beginning this study, I received approval from the Institutional Review Board at the University of Illinois. Teachers and principals may fear that information that they share during interviews will be made public or will be shared with colleagues in their districts. They may fear that information that they share could damage their reputation in the workplace or their relationships with colleagues or supervisors. I provided all participants with informed consent and their rights as human subjects. Interview protocols were followed, and no identifying information was shared or released. Pseudonyms were given to sites and subjects to ensure that confidential information was kept private. Information provided by participants was not used with other participants within the study.

Summary

This chapter outlined the research questions, research design, data collection procedures, site and sample selection, data analysis, validity, and ethical considerations for this study. I conducted a qualitative interview study involving teachers and principals in two separate Illinois school districts to attempt to determine how teachers perceive two different and common types of student growth models and how their perceptions connect to relational trust with their principals. This research fills a gap in the research related to teacher perceptions of SLOs and all-

in approaches, and it will help inform the practice of principals who are tasked with completing teacher evaluations. On a broader scale, I hope that this study will inform school districts on which student growth models are viewed as most effective by teachers and what actions principals take during teacher evaluations that lead to improved professional development for their teachers.

Chapter 4

Results

This chapter draws on the information attained from principals and teachers through this qualitative interview study to address how teachers perceive the use of student test scores in their teacher evaluations and how various approaches to the student growth model influence teachers' practice and professional development. This study examined three main research questions:

1. To what extent do Illinois public school teachers who are evaluated using student learning objective and all-in student growth approaches perceive the processes to be understandable, fair, and valid?
2. To what extent have these student growth approaches influenced their professional practice?
3. How does the principal's approach to teacher evaluation during the student growth component of the evaluation influence teachers' professional development?

Background information about each school in the study is provided first. Results are then organized based on the three major findings of this study and divided between the two schools for the sake of comparison.

Contextual Background at Each School

Teachers' perceptions of the evaluation process are influenced by the contextual background at each school. Therefore, it is important to consider factors like demographics, achievement data, school climate, and leadership approaches.

Monroe School. Monroe is a relatively high-performing school in a predominately white town. Monroe School was rated commendable by ISBE during the 2018-2019 school year (the second-highest school rating). The percentage of students who met or exceeded standards was above the state average on both portions of the Illinois Assessment of Readiness in 2019. Disparities in scores were present between non-low income and low-income students, females

and males, students who did not have IEPs and students with IEPs. The student population is 90% white, 12% of students have IEPs, and 37% of the students are low income.

The teaching staff is relatively stable and experienced. 40% of teachers at Monroe School have a master's degree or higher, the teacher retention rate is 95%, and 100% of teachers in the school were rated either proficient or excellent on their evaluation during the last 3 years. The teachers in this study had between 6 and 24 years of experience. The principal, Mrs. Douglas, has served at Monroe for 10 years and has more than 20 years of experience as an educator.

The school participated in the 5Essentials Survey to assess school climate in 2015 and 2017-2019. The survey covers five categories: effective leaders, collaborative teachers, supportive environment, ambitious instruction, and involved families. The survey is taken by teachers, parents, and students who are in fifth grade or above. The survey questions regarding "effective leaders" are mostly related to stakeholders' perceptions of the principal of the school, and the "supportive environment" questions are related to stakeholders' perceptions of staff and faculty. During the 2015 school year, the category of "effective leaders" received the second-highest rating based on all stakeholder responses. In 2017-2019 the "effective leaders" category was rated as the middle rating on the survey. The category of "supportive environment" was rated as second-highest in 2019 and middle rating in 2017-2019. There was no data for the "supportive environment" category in other years.

Roosevelt School. Roosevelt is also a relatively high-performing school with similar student demographics. Roosevelt School achieved the second-highest designation from ISBE during the 2018-2019 school year. On average, students scored above the state average in Math and below the state average in ELA on the Illinois Assessment of Readiness. Disparities in scores were present between low income and non-low income students, female and male

students, Hispanic and white students, and non-IEP students and students who had IEPs. The student population is 90% white, 50% are low income, and 18% have IEPs.

The teaching staff is relatively stable and experienced, though not as experienced as at Monroe. Fifty percent of teachers had a master's degree or higher, and the teacher retention rate was 90% in 2018-2019. Ninety-nine percent of teachers were rated as proficient or excellent over the last 3 years. The teachers in this study had between 2 and 13 years of experience. Mr. Foley, the principal, has been at Roosevelt for 12 years and has been in education for 25 years.

Roosevelt implemented the 5Essentials survey every other year, beginning in 2015. In all years that the survey was implemented, the category of "effective leaders" was rated as average. Data regarding "supportive environment" was only available for 2019 and received the second-highest rating.

Teacher Evaluation Processes at Each School

Each school had a different process that led to its current evaluation system.

Monroe School. Monroe School uses Charlotte Danielson's Framework for Effective Teaching for the professional practice portion of their teacher evaluation process. Professional practice evaluation involves formal observations, pre-observation and post-observation conferences and the submission of artifacts by the teacher. The number of formal observations for non-tenured teachers is a minimum of two, while tenured teachers only have to be observed once. The artifacts are intended to show evidence of performance in the four domains of the Danielson Framework. Professional practice accounts for 70% of the summative rating for teachers

The Monroe district joint committee adopted the student learning objective approach for student growth after PERA, and the committee has not made any changes to that process since

the original implementation. Teachers are required to submit two student learning objectives to the principal by September 15th of the school year when they are being evaluated, and the principal must approve or deny all SLOs by October 1st. Teachers are allowed to choose the assessment that they use from a menu of assessments, and most of them use a halfway to one hundred chart to set goals for individual students. This means that the teachers take the student's pre-test score and subtract it from 100 and divide it in half. For example, if a student scored a 50 on the pre-test, they would be expected to improve their score by 25 points on the post-test. Monroe uses an online teacher evaluation system called Evaluwise to complete all steps of the evaluation process. The student growth component counts for 30% of the summative evaluation rating, and the two SLOs are averaged in order to calculate that score.

For staff at Monroe, adopting the Danielson model aligned to previous work they had done. Mrs. Douglas was the principal at Monroe prior to the implementation of the PERA mandated changes. She was on the district's joint committee that was in charge of that process. The district began asking teachers to collect evidence for all four domains of the Danielson Model in 2018. Mrs. Douglas explained,

I've been here ten years and the evaluation tool that we were using when I came here ten years ago was actually developed by Charlotte Danielson and Tom McGreal. . . So it had elements of what we now know with the Charlotte Danielson domains one through four. . . . So it was like six different sections and I think there were 54 attributes in the evaluation tool . . . without a rubric.

Prior to PERA, the district did not have a rubric for professional practice and did not ask teachers to provide artifacts or evidence. However, even with the rubric, Mrs. Douglas reported that she felt that when the district changed to the Danielson Framework, it was smooth because the teachers were already somewhat familiar with it, and the new process was less subjective. Mrs. Douglas recalled that the district provided various trainings related to the Danielson Framework,

use of artifacts, and the SLO process when the mandates were implemented. She noted that teachers needed professional development to understand that the SLOs should be based on long-term assessments, not short units of instruction and assessment. Mrs. Jenkins, a 20-year veteran third grade teacher, had a similar perspective to Mrs. Douglass. She recalled that the adoption of SLOs was not controversial in the district because the teachers felt that it was just what they were required to do. She said that the Danielson transition was mostly smooth because they were already familiar with it. However, she did note that the rigor of the professional practice evaluation increased with the adoption of the rubric. Mrs. Clark, who has taught kindergarten in the district for 16 years, remembered that there was some resistance among teachers when the SLOs were first implemented. She said that teachers “dug their heels in” until they got comfortable with the changes.

Roosevelt School. The professional practice portion of teacher evaluation at Roosevelt Elementary School is similar to Monroe Elementary School. At Roosevelt School, the professional practice component includes observations, pre-observation conferences, and post-observation conferences. The number of observations varies based on whether or not the teacher is tenured. Professional practice ratings are based on the Danielson Framework for Effective Teaching, with evidence coming from both observation and artifacts submitted by the teacher. Professional practice accounts for 70% of the teacher’s final rating, just like Monroe School. Roosevelt uses an online program called Frontline for the professional practice portion of the evaluation process.

The Roosevelt joint committee for PERA initially adopted the student learning objective approach to teacher evaluation but later switched to the all-in approach, due in part to the time required to implement SLOs and also because they believed that a company with more statistical

expertise was better suited to predict student growth. The district began working with Education Consulting Research Analytics Group (ECRA) during the 2018-2019 school year. Now ECRA establishes predicted student growth scores based on Illinois Assessment of Readiness (IAR) and Measures of Academic Progress (MAP) scores. ECRA provides a growth target for the district, and all teachers are rated the same way based on whether the entire district meets the goal. For the last 2 years, all teachers received a proficient rating on the student growth component of their evaluation. The district has the student growth component of the evaluation set for 30% of the teachers' final summative rating.

Mr. Foley was at Roosevelt School prior to the passage of PERA, and he served on the district's PERA joint committee. He said that when they started with the SLO process, the teachers were allowed to choose two assessments. Many teachers chose MAP tests in Reading and Math, but specialists had the option of choosing content area specific tests. Mr. Foley noted that the high number of different assessments was one of the issues that made the SLO process complicated. The process of making predictions for students' test scores was also troublesome for teachers under the SLO process. Mrs. Kaiser, a third grade teacher, indicated that teachers felt they were almost picking random scores and assessments for their students without really having much to go off of.

In the midst of the struggle with the implementation of SLOs, Mr. Foley recalled that he talked to a superintendent in a different district whose school used an all-in approach and was pleased with the results. This led his superintendent and curriculum director to research the idea. They found that ECRA would compile student data and make predictions for student achievement. The district reconvened the PERA committee, which included administrators and teachers, to study the idea of using ECRA and all-in goals instead of SLOs. Mr. Foley stated that

teachers supported moving to all-in because it eliminated having to find their own assessments and trying to predict their students' growth. He said that moving to all-in also reduced the number of meetings needed for the student growth component by three for each teacher, thereby saving time for both the administrator and the teachers. Mrs. Kaiser commented on ECRA compiling all of the data for the district. "I think it was the best part because it was just one less thing to add to the teachers basically." Mrs. Jones, a math interventionist, recalled that the teachers were somewhat split when they voted to move to the all-in approach. She stated that she voted in favor of going to all-in, despite her worries about being evaluated based on students who were not on her caseload, because she trusted other teachers in the district to do a good job. Mrs. Kaiser added that teachers were okay with it after the administration showed them a chart of how the student growth score would influence their overall rating. She said that once they realized it would only have a negative impact on the final score if a teacher had a professional practice rating of "needs improvement" or "unsatisfactory," that relieved some of the teachers' anxiety about it.

Teacher Perceptions of Student Growth Approaches

Teachers at Monroe School had a stronger understanding and a more positive view of the student growth model at their school than teachers at Roosevelt School.

Monroe School. One of the issues identified in the literature review was that teachers often do not understand their student growth models (Amerein-Beardsley, 2014). This was not the case at Monroe. Mrs. Douglas, the principal, felt that Monroe teachers understood the SLO process because it was explained to them prior to implementation and that the teachers have access to the documents containing information that outlines the process, scores, and ratings. All of the teachers at Monroe School indicated that they understood the SLO process used in their

school and understood where their scores came from. They were not all confident of the actual percentages: Mrs. Smith could not remember the exact percentage of her score that came from her SLOs, and Mrs. Jenkins believed it was worth 30% but was not certain. Mrs. Clark noted that she did not understand the SLO process very well at the time of the initial training and implementation, but she understands it now.

The literature review also illustrated that teachers often have concerns about the fairness and validity of student growth models. These areas did not emerge as major concerns for the teachers at Monroe. Mrs. Jenkins said that she thought the process was fair because she liked the data behind it as a math teacher. However, she proceeded to talk about artifacts related to the Danielson Framework instead of the SLO process. Mrs. Tracy said that she felt like student data should be a part of teacher evaluation, but she felt that she could show her special education students' growth without going through the SLO process. She understood the need for accountability but did not think the SLO approach was necessarily the right way to achieve that. In terms of validity, Mrs. Douglas said that with the SLO process some teachers write rigorous goals, but others write goals that will positively influence their evaluation scores. "The excellent/distinguished teachers are writing goals that not 100% of their kids are going to be making 100% growth . . . and I do feel like our lower performing teachers sometimes write easier attainable goals."

Respondents from Monroe cited a variety of positive outcomes related to the use of the SLO process. Mrs. Douglas said that teachers are invested in helping their students excel on the assessments because they get to choose what assessments they use. Mrs. Clark, a kindergarten teacher, agreed that selecting their own assessments made her more comfortable being evaluated on their scores. Mrs. Jenkins, in third grade, said that she loves having student growth data

included in her evaluation. Mrs. Tracy, a special education teacher, and Mrs. Smith, in pre-kindergarten, reported that they like looking at pre-test and post-test data in order to see the growth that students are making.

While respondents from Monroe were mostly positive, they also expressed some concerns about the process. Mrs. Douglas noted that some teachers just wanted to get the process done and made goals based on short-term assessments. She reported that this was an ongoing professional development goal for the school, because she wanted their goals to be long term learning objectives. Mrs. Smith said that her pre-kindergarten students are at widely varying stages of development, and some of them need very different learning goals than others. She was not opposed to collecting student data but found it difficult to assess them for the SLO process. Mrs. Tracy felt that some kids just do not perform well regardless of how the teacher instructs or prepares them. Mrs. Tracy only had two students in one of her SLO groups and three in the other. Therefore, if one of her students had a bad day, it could negatively impact her score.

Teachers at Monroe School had suggestions for principals to improve the teacher evaluation process that were unrelated to the student growth component. The principal and teachers agreed that more professional development at the time of the implementation of the Danielson Framework would have been helpful. However, teachers reported significant concerns about the collection of artifacts related to the Danielson Framework. Mrs. Jenkins said that the artifacts have been much more of a concern than the SLO process for the teachers. She said, “I’m hearing a lot of pushback from teachers who have done it last year in that it was almost a second job and it was very time-consuming.” Mrs. Tracy said,

We’re constantly having to prove what we’re doing right by pictures or you know. . . . If you know you are doing a good job and they know we’re doing our job, it seems redundant. So, the collection of artifacts is a lot.

Mrs. Douglas, the principal, was aware of the teachers' concerns about artifacts. She said,

Teachers I think freaked out a little bit about all these artifacts. Like how many? They wanted to know like how many do you need for domain one. How many do you need for domain four? And I was like, I can't tell you. I'm not like 23 is the magic number.

Mrs. Douglas encouraged her teachers to look at the rubric and determine how many artifacts they needed to show that they were excellent or proficient. The teachers wanted direction on a specific number of artifacts for each domain.

Roosevelt School. The teachers at Roosevelt School reported significantly less understanding of the all-in approach used at their school than their counterparts at Monroe. Mr. Foley, the principal, stated that when Roosevelt used the SLO process, he did not feel that he had enough understanding of statistics and predicting students' scores, even after training, to evaluate whether teachers' SLO goals were appropriate. In terms of his teachers' understanding of the all-in approach, he said that they know that all of the students have data points that ECRA uses to calculate a score, but that neither he nor the teachers understand how those scores are calculated. Foley stated, "That's left to the statistician." Mrs. Mosser knew that there was a company that took student data and compiled it in order to generate goals, but she was not sure of the time period that the goals covered, the subjects included in the goals, or how the goals were calculated. She did not know how much of her final rating was based on her student growth score, but she did know that her score had been proficient. Mrs. Jones understood that there was a goal calculated by an outside company that accounted for 30% of her evaluation score. She reported that it was based on students' ELA and Math scores on standardized tests, but she was not sure what the actual goal was or if it changed from year to year.

Other teachers agreed. Mr. Taylor reported that he felt that the process was easy to understand, but when he elaborated he was actually talking about ratings on the professional

practice component that includes written commentary from the evaluator, not the student growth component. Mrs. McCarter said,

with the evaluation piece as far as like the data piece of it, we got an email at the end of the year, and it just said everyone got proficient. So, I don't know how, I don't know. I don't even honestly know what the goal was.

She reported that she's seen ECRA data for her students' scores and knew that grade levels wrote SMART goals for student achievement but was not sure how the score of proficient was calculated or what it was based on. Mrs. Kaiser said that they get ECRA data at the beginning of the school year that shows how many students should meet or exceed on the IAR, but she was not sure how those grade level predicted scores fit into the districtwide goal. She could not recall how much of her overall rating was based on her professional practice or student growth. Mrs. Jones expressed concerns that she did not understand the student growth goals, so she was not sure what she could do specifically to increase the rating for the teachers in the district.

At Roosevelt School, the teachers all reported that the process was not fair for different reasons. Mrs. Mosser said that the all-in approach is not fair because there are so many students who affect the score who are not in the teacher's class. Mrs. Jones also said it was unfair and noted that some students are less likely to show growth than others. She specifically mentioned interventionists and special education teachers as having student populations who might be less likely to show growth. Mr. Taylor noted that classes are sorted so the top students are grouped together, making some classes more or less likely to meet growth targets. He did not like being rated based on how other teachers performed. Mrs. McCarter felt that student growth was not fair for special education teachers, because they would be less likely to meet growth targets. Mrs. Kaiser also felt that the student growth process was not fair, because "some kids aren't always having the best day. You just can't get them to, no matter what you do, perform like what's

expected necessarily.” However, the principal, Mr. Foley explained that, “I think it’s fair just because we’re kind of all in this together K through 12.” He used an example that if students did well in third grade, it would be a result of the third grader teachers’ work, but the prior grades also would have contributed to that success.

The teachers at Roosevelt School were also skeptical of the validity of the student growth process. Mrs. Mosser stated that some students have “off” days and might not test well in spite of the teacher’s efforts. Mrs. Jones said that teachers do more than just teach the curriculum and could be successfully helping the kids become better people, but that might not be reflected in their student growth score. Mr. Taylor noted that many of his kids were predicted to have a low probability to meet the expected target compared to his same grade colleague’s students, because he was assigned a cluster of lower students. Mrs. McCarter expressed doubt about the validity of the model but said that if all of a teacher’s students were failing, that would be a sign that something was wrong. Mrs. Kaiser said it was not valid because some students just test poorly, and it does not always reflect what they were taught or what they know. Mr. Foley, the principal, was undecided on the accuracy due to the short time that his district had been using the all-in approach. He said,

From my experience, I think there is some student growth pieces where whether it’s related directly to an evaluation or if you look at cohorts of kids and certain grade levels, even certain classes, that certain teachers were more effective with them for whatever reason.

Roosevelt teachers did note some positives about the all-in approach used by their school. All of the teachers reported that the all-in approach is easier and less time-consuming than the SLO process. Mr. Taylor, a fourth grade teacher with 2 years of experience, was not ever evaluated using the SLO process. He thought that one of the benefits of having all-in goals was that the teachers’ success depends on each other. Therefore, it promotes teachers working

together. Mrs. Kaiser said that the student data provided by ECRA was one of the main benefits of using the all-in approach. Four of the participants at Roosevelt School had used the SLO and the all-in approach. They all preferred the all-in approach. Mr. Foley, the principal, said that all-in saves time, eliminates extra meetings, and reduces paperwork. Mrs. Mosser, a special education teacher, said that she liked the all-in because it is easier, but she did think that the SLO was more effective for evaluating teachers. Mrs. Jones agreed, stating that she prefers all-in due to the smaller amount of time involved, but she also thought the SLO was a better measure. Mrs. Kaiser preferred the all-in because the student growth data is already there and ECRA manages it for the district.

Teachers at Roosevelt had suggestions for improving the teacher evaluation process that included professional development to help teachers understand the all-in model, more observation, and fewer artifacts. Two teachers said that the principal could improve the value of the teacher evaluation process by explaining more about the student growth goals and helping teachers get a better understanding of how they can improve the district student growth score. Mr. Taylor suggested that more evaluator time in the classroom would be beneficial, but he also noted that he thinks his administrator does a nice job on the evaluation process. Mrs. Kaiser valued informal observations and feedback related to her everyday teaching the most. Mrs. Mosser also said that she values feedback on her professional practice. Teachers at Roosevelt also expressed concerns about the time involved in collecting artifacts but had fewer complaints about the process and said that it has gotten easier over time. Mrs. Jones said,

We used to do binders and that took forever. I mean it would take weeks for me to prepare my portfolio and so I would put together just a sheet of every single thing from each page. And so, that was difficult for me just because I didn't really see the purpose of wasting all of the paper on a picture of my expectations that are in the room. Because even for domains two and three, we still had to show the evidence. And so now we really provide mainly one and four, so it saves a ton of time. . . . So each year it gets a little

better and I really can do it. I mean I prepare for a while, but it really doesn't take very long.

Mr. Taylor estimated that the collection of artifacts is a five to six hour process for him, and Mrs. McCarter indicated that the artifact component has gotten easier to understand, but it is definitely the most time-consuming part for teachers.

How the teachers were rated using the student growth models likely influenced their perceptions of the models. Their reports on how the student growth models influenced their ratings is included here for context. At Monroe School, teachers offered mixed reports on how their teacher evaluation ratings were influenced by the SLO process. Mrs. Douglas noted that some teachers improved their rating by not writing rigorous goals. Mrs. Mosser and Mrs. Baker said that their scores were positive and helped their overall evaluation ratings, but Mrs. Baker noted that at a previous school her SLO hurt her score after she did not meet one of her goals. Mrs. Smith reported that her student growth score was neutral to her teacher evaluation. She said that she uses the same goal or a similar goal every year.

At Roosevelt School, the participants unanimously agreed that the student growth portion of their evaluation was neutral to their overall score. Mrs. Jones stated that teachers could set themselves up for success under the SLO process, but her rating with the all-in has not influenced her score. Since the implementation of all-in, the rating for all teachers in the district has always been proficient.

Influence of Student Growth Models on Professional Practice

Teachers at Roosevelt School reported that the student growth model influenced their professional practice more than teachers at Monroe School.

Monroe School. Teachers at Monroe reported that the SLO process has not caused them to make significant changes in their curriculum. They denied any narrowing of the curriculum in

order to increase test scores and teacher ratings. Mrs. Douglas, the principal, said that most teachers use teacher-created assessments, and they focus on the specific topics involved in those areas more closely than they would have before. Mrs. Clark and Mrs. Tracy reported that they have always used the kinds of assessments that they are now using for the SLOs, so it has not really changed their teaching materials or pedagogy. Mrs. Smith said that the only thing that has really changed her curriculum and instruction is the state preschool grant requirements. Mrs. Jenkins felt that the Danielson Framework had influenced her to move toward more student-led instruction but did not report changing her teaching due to the SLOs.

Educators at Monroe did not report as much of an increase in student data usage with the SLO approach. The principal said that she thinks they talk about student data now more than before PERA, but she said the increase is also due in part to their Response to Intervention process and the new school performance designations from the Illinois State Board of Education. Mrs. Smith said that she still uses data like she did prior to PERA, but she added that she probably looks at it more closely now. Mrs. Clark said that she increased her use of student assessment data due to the Danielson Framework assessment domain. She keeps bar graphs of student progress that they help fill out and monitor each month.

Teachers at Monroe School did not report an increase in teacher collaboration due to the SLO process. They reported that they did collaborate to create their SLOs the first time that they had to create them. Mrs. Douglas said that grade levels usually do not work together to create common goals for the SLOs because teachers are rarely on the same evaluation cycle. Mrs. Smith said that new teachers often collaborate with veteran teachers the first time that they have to write an SLO. She also said that she consulted the kindergarten teachers prior to writing her

SLOs to be sure that she was targeting areas that would help prepare her students for Kindergarten.

Roosevelt School. The all-in approach at Roosevelt School and the utilization of ECRA data has changed the professional practice in the school more significantly than at Monroe School, according to these interviews. Teachers did not report making changes to curriculum, but they did change their planning and instruction. Mrs. McCarter noted that she uses the ECRA data to plan student groups for her special education students. Mr. Taylor and Mrs. Kaiser also said that they base grouping of students in learning groups in their classroom on ECRA data. Mrs. Kaiser noted that she and her partner teacher reduced the number of lengthy reading tests that they give in class, because they did not want to burn students out on long testing when they also had to take the Illinois Assessment of Readiness (IAR) and Measures of Academic Progress (MAP). Mrs. Mosser said that she has always focused on her students' IEP goals over any goals for her teacher evaluation, both under the SLO and with the all-in process. Mrs. Jones and Mrs. McCarter said that they had changed their instruction more from the Danielson Framework than the student growth model.

The teachers at Roosevelt School reported a significant increase in the use of data since their school began working with ECRA. The principal, Mr. Foley, said that ECRA gives a cut score for students who are likely to meet expectations on IAR. They focus on those students who are close in order to ensure that the greatest number achieve that level. Mrs. Jones, a math interventionist, reported that she uses the student data provided by ECRA to see under which standards her students need additional instruction. Mr. Taylor said that the ECRA data helps him identify gifted or talented students, as well as students who might need special education and those who need extra attention from him. He uses the data to determine his instructional groups

in Reading and Math. He also said that his team looks at the specific areas of Math in which students are not performing well, and they target those areas in instruction. Mrs. McCarter said that the teachers look at the student data to target students who are predicted to be “on the bubble” of meeting expectations on the IAR. She also noted that her special education students often have very low percentage predictions for the IAR targets, so the data is less valuable to her specific caseload. Mrs. Kaiser also mentioned that the school focuses on students who are very near the target and tries to intervene to boost them up to meeting expectations. She said that she uses the ECRA data for planning instruction for groups, but she admitted that she should look at the data more closely.

In terms of teacher collaboration, Roosevelt School teachers did not report an increase related directly to the all-in approach. Mrs. McCarter and Mrs. Kaiser said that they usually meet as grade level teams when the data comes out but that they had only met once so far in the school year as of October. Mrs. Mosser said that teachers collaborate to create materials and prep for state testing, but she did not report increased collaboration resulting from the all-in process. Mrs. Jones did not feel that there was increased collaboration. The principal, Mr. Foley, noted that the district provides a great deal of professional development in English Language Arts and Math, and the teachers collaborate based on those trainings. He said that the whole process “has probably increased our collaboration as a district.”

While one of the main purposes of teacher evaluation is professional development, teachers at both schools reported very little regarding professional development related to the student growth processes in their schools. The professional development that teachers reported was related to training them for how to understand the Danielson Framework, how to use the online evaluation tools, and how to write their SLOs. Mr. Foley, the principal at Roosevelt, said

that they do a lot of professional development with their teachers and instructional coaches, but not centered around the all-in process.

Influence of Principal Approach on Professional Development

The principals at Monroe School and Roosevelt School had a similar approach to teacher evaluation that seemed to lead to the maintenance of high levels of trust, in spite of teacher concerns about the processes.

Mr. Foley and Mrs. Douglas follow many of the same steps during the teacher evaluation processes at their schools. Both principals send observation dates to the teachers at the beginning of the school year and provide a timeline of what teachers need to complete. They have the same set of steps that include a pre-observation conference, observation, a post-observation conference, review of artifacts and student growth scores, and then a summative conference to give the teachers their final scores and rating. At Roosevelt School, experienced teachers learn about the teacher evaluation process through their principal and the curriculum director. New teachers learn through new teacher orientation, their mentors, the principal, and the curriculum director. The teachers reported that the curriculum director comes and meets with them once a year to explain the student growth score. At Monroe, the teachers get information about teacher evaluation from the principal each year.

Both principals held meetings with teachers as part of the professional practice portion of the teacher evaluation, but neither of them held meetings with teachers regarding the formulation of student growth goals. Teachers submitted goals via email or in the school mailbox at Monroe, and the principal was required to respond within a specific number of days. ECRA provided the student growth goals for the teachers at Roosevelt. Since there was no need to negotiate goals or discuss the rigor of the goals, trust between the principal and teacher would not have been as

significant of a factor as in other districts where the principal and teachers collaborate or negotiate to establish student growth goals.

The principals at Monroe and Roosevelt also reported similar approaches to building trust with their teachers, and the teachers at both schools reported that they had trusting relationships with their principals. Mr. Foley at Roosevelt said that he's built trust with his teachers over time. He explained

We've never looked at evaluation as being a gotcha mechanism. We've always looked at it as our jobs are to help you be the best that you can be in teaching. We're not here to get you with that, and I don't want you to feel like that. So, it was almost like my thought was I'm going to help you with your professional practice regardless of what the goals are.

Mrs. Douglas, the principal at Monroe, expressed very similar ideas related to building and maintaining trust while evaluating teachers. She said, "I'm another pair of eyes. It's not a gotcha. I'm here to celebrate what you're doing great and maybe give you some ideas on how you can get better." She said that she builds trust over time with her staff by letting them see her make mistakes and not having all of the answers. She noted the importance of respecting their professional opinions and not acting like she is an expert in all areas. She felt that the SLO process has had a positive influence on her relationship with teachers because it increases the opportunity for conversations. Mrs. Douglas said, "I think it has over time just given us more of a chance to dialogue." She emphasized that the last thing she wants is more busy work for the teachers, and so she tries to ensure that they are studying meaningful data and using that to reflect on their practice.

Teachers at Roosevelt School said that trust with their principal was positive prior to the all-in approach and has not changed as a result of it. Mrs. Jones said, "Our principal always wanted us to succeed. I think our principal has a lot of trust in us, which is great, but I think he

would be okay questioning if he thought it wasn't a big enough improvement." Mrs. McCarter said that she trusts that Mr. Foley would not require them to do something if he did not think it was important. She said it is nice to build a relationship with the principal through post-observation conferences and other meetings, but she talks to him every day anyway and already has a positive relationship with him outside of the evaluation process. The other teachers interviewed at Roosevelt stated that they did not think the all-in approach had influenced the relationship between the teachers and the principal either way.

At Monroe School the teachers did not feel that there was a change in the relationship with the principal related to the SLO process. Mrs. Tracy, a special education teacher, said that Mrs. Douglas trusts the teachers to set their goals, and Mrs. Smith, a pre-k teacher, said that it has helped open more conversations with the principal. Mrs. Clark said that there was no impact on relationships based on the SLO, but that there were some teachers who were upset when the school transitioned to the Danielson Framework. She reported that Mrs. Douglas addressed the staff concerns by providing professional development for the staff related to the differences between proficient and excellent on the Danielson Framework, and that helped the teachers attitudes toward it. No teachers reported that they had ever had an SLO not approved, that they had been encouraged to increase the rigor of their SLO, or that they had not met one of their SLOs while at Monroe.

Neither school reported a significant change in the trust between teachers related to the use of their student growth approaches. At Roosevelt, Mrs. Jones said that she had a high level of trust in the staff at her school when they decided to go to the all-in approach. That was one of the reasons that she felt that the all-in approach would probably go well. However, she and Mr.

Taylor did mention that some teachers might have concerns about the student growth scores of other grade levels. Mrs. Jones said,

I don't think it's really affected us in our relationships between each other. I think some teachers might want to put blame like we got proficient not excellent because of this group, but you can't pinpoint it. I mean, you can see the scores if you want, but I don't think anyone's really going to put that much effort into finding where the issue is and point fingers.

Mr. Taylor said that he did not think that they really had access to the scores of other grade levels, but he had heard a rumor from another teacher that one grade level at a different school tested very poorly and negatively impacted the score for the district.

Teachers at Monroe School did not report any influence on teacher relationships related to the SLO process. Their scores were independent of other teachers, and most teachers did not collaborate on their SLOs. They were not aware exactly what SLOs their colleagues had written and how the SLO approval process went for others.

Summary of Findings

This study was designed to examine the perceptions of elementary school teachers who have been evaluated using the SLO model and all-in model for student growth in teacher evaluations and to what extent the principals' approach to the process influenced the teachers' engagement with the model. The study found that teachers using the SLO model had a better understanding of the model and viewed it as fair. Teachers who used the all-in approach did not fully understand it and did not think it was fair or accurate. However, those teachers who had been evaluated using both the SLO and the all-in approach preferred the all-in approach because it was less time-consuming, and the data from the all-in approach had value to them in their instruction. Teachers who were evaluated using the all-in approach also reported that their student growth model had a greater impact on their professional practice, because they looked at

data more often and used it in their planning for instructional groups and skills to teach. The principal's approach was similar in each of the schools that participated in spite of the different models in use. The principals built trust by reassuring teachers that they were looking to improve their professional practice, not to deduct points off of a rubric. Both principals emphasized the formative nature of the evaluation and were concerned with the time involved in the process for their teachers. Both principals maintained trusting relationships with their staff throughout the implementation of PERA reforms.

Chapter 5 will discuss these findings in light of the literature. Implications for principals, policymakers, and teacher and principal preparation programs will be followed by recommendations for future research.

Chapter 5

Discussion and Implications

This chapter summarizes this qualitative interview study on teachers' perceptions of student growth models and their incorporation into teacher evaluations. First, the findings will be summarized. Then the implications for current principals, policymakers, and teacher and principal preparation programs will be discussed. Recommendations for further research will be included prior to an overall conclusion of the study.

The purpose of this study was to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using various student growth models, and to what extent the principals' approach to the process influences the teachers' engagement with these approaches. Three major findings were identified. First, teachers at Monroe School had a more positive view of the student growth model at their school than teachers at Roosevelt School. Despite this, teachers at Roosevelt School reported that the student growth model influenced their professional practice more than teachers at Monroe School. Overall, the principals at Monroe School and Roosevelt School had a similar approach to teacher evaluation that resulted in the maintenance of high levels of trust in spite of teacher concerns about the processes.

Discussion

The findings in this study are mostly consistent with the research cited in the literature review. Fairness was a significant concern for teachers in many of the studies involving teachers' perceptions of the use of student test data in their evaluations (German, 2014; Jiang et al., 2015; Pressley et al., 2018). The teachers in this study expressed similar concerns. They felt that student test performance might be poor due to factors outside of the teachers' control. They did

not think it was fair to be rated based on the scores of students who they did not teach, and they said that the tests that were used did not show all that teachers actually teach students.

In the research, teachers had concerns about their lack of understanding of their student growth models (German, 2014; Jiang et al., 2015; Pressley et al., 2018). Jiang et al. (2015) found that teachers in Chicago were not sure how much student growth counted for their evaluation. German (2014) reported that teachers did not understand the calculation used for their VAM or how the growth was predicted. Pressley et al. (2018) stated that teachers in Florida did not feel that their evaluators could accurately explain their scores to them. This lack of understanding was a prominent theme in this study for the teachers at Roosevelt School who were evaluated using the all-in model. They did not understand how ECRA developed their scores or what their goals were. Some teachers also did not know how much of their evaluation was based on the student growth scores. The teachers did not know what they could do to improve the scores. The principal also did not understand how the scores were predicted. However, lack of understanding was not a concern reported by the teachers at Monroe School with the SLO approach.

The findings in this study did differ slightly from the research in the area of curriculum and instruction. German (2014) reported that teachers in Ohio changed their instruction in response to their VAM scores. Teachers reduced discussion and debate in favor of more constructed response practice in order to improve test scores. Pressley et al. (2018) found that teachers felt pressure to narrow the curriculum and focus on test preparation as a result of their student data scores. Narrowing of the curriculum and pressure to improve test scores was not reported at either school in this study. However, teachers at both schools reported that they look at student data more, and teachers at Roosevelt School reported that the use of student data has

changed their planning and instruction. They indicated that they look at student scores to create instructional groups and look for skills that are weak across the class or grade level.

This study was guided by the theoretical framework of relational trust. Bryk and Schneider (2002) developed the theory of relational trust based on their work studying reform efforts in the Chicago Public School system. They define it as a three-level theory that involves the (a) intrapersonal process of determining the intentions of others; (b) interpersonal roles within the context of the culture, history, climate, and local norms; and (c) impact of relationships on decision-making, support for change, and improvement (p. 22). Bryk and Schneider identified four areas that usually play a role in how people determine the intentions of others: respect, personal regard for others, competence in core responsibilities, and personal integrity. In approaching this study, I anticipated that a relationship of trust between the principal and the teacher would be a key theme because of the negotiation of student growth goals. I anticipated that with the SLOs, the teachers would create their SLOs individually, and then the principal would provide guidance on the rigor and the predicted growth targets. I thought that the all-in approach would involve a goal or goals established by a team of teachers and administrators.

However, the actual processes that are in place for the student growth models at both schools did not appear to involve much interaction between the principals and teachers. At Monroe, the teachers submitted the goals to the principal. The goals were then approved by the principal by email or via the Evaluwise system. Since the implementation phase is over, many teachers could now use the same or similar SLOs and same assessments from year to year. This has probably eliminated some of the guidance and negotiation of goals that was originally required. At Roosevelt, ECRA created the student growth goals, thereby eliminating any

conversations around the creation of goals between the teacher and principal. Because principals and teachers had minimal interactions in the process of goal creation, it was not possible to see how trust was part of the process.

Relational trust was an important part of how principals approached their work with teachers, and thus may be underlying the teacher evaluation processes, even though it did not emerge as a strong theme in the data. Bryk and Schneider (2002) defined respect in relational trust as actively listening to others and seeking to understand their perspectives. In this study, Mr. Foley at Roosevelt talked about letting his teachers do most of the talking during his meetings with them, so that they could reflect on their own practice, and he could learn how they feel about their own teaching. He also said that he tries to talk to them about their professional practice outside of the evaluation, so that they know that he is interested in helping them develop as teachers. Mrs. Douglas at Monroe talked about a teacher who came to her with a concern about her SLO. She allowed the teacher to change it because she wanted the teacher to be focusing on something beneficial. She emphasized that she did not want to add busy work to already busy teachers. Both principals emphasized that they wanted to focus on the formative nature of teacher evaluation with their staff and not the summative. Tschannen-Moran (2014) noted that the principal acting as a coach to assist teachers working towards their goals is an effective way to foster a culture of trust.

Bryk and Schneider (2002) said that personal regard for others involves showing genuine concern for people as individuals. In terms of teacher evaluation, the principal must fulfill their obligations for the steps of the teacher evaluation in a timely manner and ensure that the teacher knows what is expected of them throughout the process. In this study, both principals were reported by their teachers to have completed the steps and communicated expectations for the

teacher evaluation. However, the teachers at Monroe still expressed concerns about the collection of artifacts and desired more specific information about what they should turn in. It was evident that this was an area that could erode the teacher's trust for the principal if it was not rectified. At Roosevelt, the teachers had concerns about not understanding the student growth scores and calculations. This lack of understanding has not eroded trust at this point. However, if the student growth scores dropped below proficient, the teachers would probably become more concerned about their lack of understanding. They are certainly placing a great deal of trust in their principal and the system to support a teacher evaluation system that they do not fully understand. This falls in line with Tschannen-Moran's (2014) description of the importance of benevolence in a trusting relationship between teachers and principals. According to Tschannen-Moran, "In a situation in which one is dependent upon and consequently vulnerable to another, faith in the caring intentions or altruism of the other is particularly important" (p.19).

Professional competence was another part of the relational trust framework outlined by Bryk and Schneider (2002). In teacher evaluation, the teacher must believe that the principal is a competent evaluator. The competence of the evaluators in this study was reported by their teachers and was evident from the responses provided by Mr. Foley and Mrs. Douglas. Both said that they outline their schedules for teacher evaluations prior to the school year, and they communicate the process and timeline at the beginning of the school year. Mr. Foley said that he makes it a point to return observation notes to teachers quickly. A teacher at Monroe was appreciative of the fact that Mrs. Douglas reads all of their artifacts. Teachers did not note any concerns about the competence of the principals to carry out the evaluations. This falls in line with Tschannen-Moran's (2014) research in which competence was defined as, "the ability to

perform a task as expected, according to appropriate standards” (p. 30). Tschannen-Moran said that in schools with high trust, principals set a high standard for themselves and others.

Personal integrity was the final component of the Relational Trust framework (Bryk & Schneider, 2002). Principals must complete teacher evaluations in a fair and impartial way in order to show integrity. Although there were concerns about the fairness of the all-in process and the use of student growth scores in general, there were no reports from teachers at either school about a lack of fairness based on any actions of the principals. Teachers reported positive relationships before and after PERA. Mrs. Douglas reported that she implemented additional professional development when her teachers had questions about the Danielson Framework. She heard their concerns and provided additional information. Mrs. Douglas also said that she builds trust by allowing teachers to see her make mistakes. Tschannen-Moran (2014) noted the importance of honesty to build trust between teachers and principals. Tschannen-Moran noted that principals who accept responsibility for both the good things that happen and for the mistakes will more effectively build trust.

Administrators are often required to implement policies and procedures according to new federal, state, or district mandates. With each new election, new laws related to education usually follow. Seasoned principals, like those in this study, have been through the implementation of various mandates such as Common Core State Standards, Response to Intervention, and PERA. Principals have to navigate implementing these new policies with staff members who may be nervous about change, overwhelmed by the number of new initiatives, or resistant to the new mandate (Spillane, Reiser, & Reimer, 2002). Moreover, the principals may be opposed to the new policy. In order for a principal to establish and maintain high levels of trust with staff over time, they must implement the policy in a way that best suits the local needs, acknowledges the

concerns of the teachers, and is designed to improve student outcomes (Tschannen-Moran & Hoy, 2000).

Mr. Foley did not support the addition of test scores into teacher evaluation. He did not feel that he and his staff had the expertise or time to create meaningful SLOs and did not see that they had value in improving teacher practice. The teachers also did not think that the student growth scores belonged in their teacher evaluation. Mr. Foley reassured his teachers that their ratings would not be negatively impacted if they wrote an SLO that did not turn out well, and then he looked for another solution to fill PERA requirements. The all-in approach was his answer to meeting the requirements of PERA but also saving time for teachers and administrators and eliminating part of the process that he did not find value in. Although the teachers did not really understand the all-in process, they trusted that Mr. Foley was telling the truth that the all-in would save them time, and their ratings would not be negatively influenced. It appears that they trusted Mr. Foley due to the amount of time that he has been at Roosevelt and their prior relationships with him. They have established relationships with him over time, and he found a solution that was agreeable to both administrators and teachers. This outcome may have been less likely if he was a brand new principal coming in and suggesting that they trust him that their ratings would not be negatively impacted.

Mrs. Douglas was less skeptical of PERA. She approached it as something that was required, so she would help her staff find the benefit. She talked about encouraging her teachers to use long-term educational goals and that getting them to move to writing those types of goals was a work in progress. Her teachers knew that she was a data-driven person and that she believed their goals could positively influence their work with students. In her case, she also had the long-term relationship with teachers in her building, and the data suggest that they trusted her

not only because of that, but also because she had integrity. They did not necessarily agree with the policy, but they trusted that she was trying to improve teaching and learning in order to get value out of something that they were forced to implement.

In both schools, the principals had longstanding relationships with teachers that have an enormous influence on trust. At Roosevelt, the teachers trusted that Mr. Foley would protect their time and their ratings. At Monroe, the teachers trusted the principal was trying to get the best educational outcome out of a mandate that they had to implement. Now that the reforms are fully implemented and part of the routine, both schools seem to be going through the motions of the student data portion of the evaluation. They found more value in the professional practice component, and both found a way to emphasize that.

In addition, both schools were relatively high-performing. The teachers and principals did not report any pressure to increase test scores from the administration, the school board, or the community. The fact that test scores are not a major concern and teacher retention is high suggests that teacher evaluation is not under much scrutiny in either district. It appears that both districts are able to recruit and retain teachers. It is likely that proficient and excellent ratings for teachers are considered the norm, and therefore, there is not a significant amount of concern on the part of the teachers or the principal regarding summative ratings. In low-performing schools this would not likely be the case. Such schools would likely face pressure to improve test scores, and the performance of students in individual teachers' classrooms might be scrutinized more heavily in order to try to identify ineffective teachers. Maintaining a trusting relationship around teacher evaluation between the principal and teachers would likely be much more difficult in a school that is low-performing, due to the potential for casting blame and the likelihood that district may struggle with recruitment and retention of high quality teachers.

Implications of the Study

Findings revealed two different schools with different evaluation processes but many commonalities. I now consider implications of these findings for principals, policymakers, preparation, and research.

Implications for principals. Teachers raised several concerns about the evaluation process that are important for principals to understand. Time was an important issue for teachers throughout this study. Roosevelt moved away from the SLO process partly because of how time-consuming the goal writing and meetings were. The teachers at both schools expressed concern over the amount of time that was required for collecting artifacts. Even though the teachers at Roosevelt had serious concerns about the fairness and validity of the all-in approach, they supported it largely because of the time that it saved. This is important for principals to understand, because teachers have a significant amount of responsibilities on a daily and weekly basis. Teachers must plan lessons, create assessments, teach, grade papers, communicate with families, attend meetings, and more. Time is teachers' most valuable and scarce resource, so principals must ensure that the value of any additional tasks assigned to teachers outweigh the cost of lost time.

Professional development at the time of reform was also important to teachers. Several teachers at both schools reported that teachers were anxious about changes at various stages of implementation. Mrs. Douglas reported that she would have provided additional professional development during the early phases of implementation if she could change anything about her approach. Teachers at Monroe reported more frustration with artifacts than teachers at Roosevelt, who had already been collecting them for several years. Teachers at Roosevelt reported that the collection of artifacts had gotten easier over time. Principals are constantly tasked with

implementing reforms, including both those that they choose and those that are forced upon them. The results of this study show the importance of communicating early and often. Principals must provide information in advance for any major change. Principals can smooth the transitions involved in these reforms by making sure that teachers know what is expected of them and that they know how to do what they are expected to do. Principals should not assume that training teachers once for a new initiative is sufficient. Teachers need information about the reform in advance, but they also need ongoing professional development as they go through the new process.

Implications for policymakers. The findings of this study are in line with the research outlined in the literature review. Teachers have concerns about the use of student assessment scores in their evaluations, because the student growth models are often difficult to understand, students are not randomly assigned to teachers, and student performance may perform poorly in spite of effective instruction by the teacher (Crystal, 2014; Foutch, 2017; German, 2014; Jian, Spote, & Luppescu, 2015; Pressley, Roehrig, & Turner, 2018). This study also shows that it is possible for schools to implement processes that allow for the inclusion of student test data in teacher evaluation without connecting individual teachers to those outcomes. One of the intentions of including student tests scores in PERA was to identify effective and ineffective teachers based on student performance. Since the all-in approach connects all of the teachers in the district to one goal, it is not likely that the district can use this process to individually identify effective or ineffective teachers. In essence, the all-in approach has neutralized the impact of the student test data on the teacher evaluation. With an all-in approach, teacher ratings are again differentiated based on subjective professional practice ratings.

The SLO process addresses some of the concerns that teachers have about student growth models. Teachers use assessments in their own subject areas that relate directly to the students in their own classrooms. However, as reported at Roosevelt, teachers often lack the expertise to predict how much a student should grow on a given assessment, and principals may also lack the expertise to guide teachers toward rigorous but attainable SLOs. The result of teachers writing their own goals is that teachers are being held to different standards based on the level of rigor in their goals. This practice is also not likely to help districts identify individual teachers as effective or ineffective.

Policymakers must consider the processes that are actually in use in schools now that PERA has been implemented and how these processes line up with the original intent of the reform. For example, how are the SLOs being written? What percentage of students must show growth, and how much growth must they show in order to achieve each score? These decisions have been left largely to individual districts, so there is likely some variability across the state. Since the intent of the student test data reform is not being met, it is worth considering that the time of the educators involved in these processes might be used more productively for the benefit of students. This study showed that teachers value reviewing student data despite their concerns about its use in their evaluations. Perhaps policymakers could promote reforms that assist schools and districts in analyzing their student data for the improvement of curriculum and instruction instead of for teacher evaluation.

Policymakers also would benefit from understanding that not all principals support the current use of student growth objectives in relation to teacher evaluation. If principals do not find value in a mandated reform, it is possible that they will not implement the reform in the manner that it was intended. This is evident from the all-in approach used in Roosevelt School, where the

district found an approach that met the requirements of the reform but did not meet the intent of PERA. It is likely that most principals will implement required policies, whether they like them or not, but the scale of the reforms influence is likely to be diminished if many principals do not support the change.

Implications for principal and teacher training programs. The experience of the principal and teachers at Roosevelt School illustrates the difficulty that some educators face when using student test data to make decisions. The principal and teachers did not feel like they had the expertise to predict student test scores or estimate what academic levels students should attain. They were writing goals for students that had the potential to change their careers without having a full understanding of how to do so.

In order to use assessment data to make decisions about teacher evaluation, curriculum, and instruction, it is essential for principals to be knowledgeable in selecting appropriate assessments, interpreting student test data, and explaining the data to teachers. It is also important for teachers to have an in-depth understanding of how to create valid assessments and use student test data appropriately when creating instruction, grouping students, and providing intervention. Teacher and principal preparation programs should include training on how to develop and use formative assessment data effectively. Additionally, if some teachers and principals will continue to be required to predict student growth, preparation programs should consider teaching statistical approaches to developing these predictions, so that teachers and principals have the expertise required to make accurate predictions.

Implications for research. In this study, the student growth component of the teacher evaluation was effectively neutralized. It only impacted the scores of teachers who were performing poorly in their professional practice, which was rarely the case, according to the

principal. Teachers who were excellent in professional practice and proficient from the all-in score could still achieve an overall rating of excellent. It would be worth studying other districts to determine if there are different systems in use in which the all-in score has greater potential to alter teacher's overall scores.

In this study, it became evident that SLOs were approved without much, if any, discussion. The principal and teachers said that teachers sometimes set them up for success by writing easily attainable goals. SLOs are under local control, as are measurement models. This means that there is likely wide variability in how these systems are set up, as well as how they are implemented by school leaders. A study that compares SLOs and measurement models across multiple schools in the state would reveal if teachers in Illinois are being evaluated using comparable systems. It would also illustrate if some models are set up in favor of teachers or to the detriment of teachers.

In response to concerns about the use of student test scores in teacher evaluation, many states have now enacted legislation to alter the use of student scores or eliminate their use altogether. According to the Education Commission of the States (2020) in 2017, Arkansas, Kentucky, and Michigan eliminated the requirement that student test data be included in teacher evaluations. That same year Florida, Indiana, South Carolina, and Utah made the requirements for the use of student test scores in evaluation more flexible. Nevada and Tennessee reduced the use of the data, and Louisiana legislated that the use of student data required additional study. States continue to propose legislation to change teacher evaluation. In 2019, New York removed the student data requirement, Arizona made it more flexible, and Maine initiated a revision to their statewide assessment program linked to teacher evaluation. Clearly there are mixed reviews on the effectiveness of using student data in teacher evaluations. A study of states that moved

away from these evaluation procedures or altered them could illuminate problems with the process and show how states are now attempting to address issues with their processes.

Finally, there are implications for the concept of relational trust in schools. In much of the research, trust is a thing to build between principals and teachers. However, in this case, trust appeared to lead to teachers not understanding a key policy mandate. The teachers at Roosevelt School had enough trust in their principal that they were willing to accept the implementation of a teacher evaluation system that they did not understand. This trust was likely based on their history of past interactions with him that resulted in favorable outcomes. While relational trust is generally considered a positive and necessary component of successful organizations, blindly trusting an authority figure could potentially have very negative consequences. It is important that teachers do not assume that the principal is taking the best course of action simply because of past success. A healthy school climate involves input from all stakeholders, active dialogue, and collaborative decision-making. Further study into the potential pitfalls of relational trust could illuminate how high levels of trust could actually be detrimental to an organization.

Conclusion

The use of student test data in teacher evaluations has become widespread in the United States over the last two decades. After the passage of PERA in Illinois, school districts were tasked with creating and implementing new teacher evaluation systems incorporating student growth data as a significant portion of the summative rating. This qualitative interview study examined the perceptions of elementary school teachers who were evaluated using the SLO model or the all-in model, and to what extent the principals' approach to the process influenced teachers' engagement with these processes. The theoretical framework of relational trust guided this study (Bryk & Schneider, 2002).

There were three major findings from this study. First, teachers who were evaluated under both the all-in and SLO approach had concerns related to the use of student test scores in teacher evaluation, but teachers at the school using the SLO model had a more positive perception of their process than teachers who were using the all-in model. Second, teachers at the school using the all-in approach reported a greater influence on their professional practice, specifically related to an increased use of data to make educational decisions. Finally, the principals at each school approached the teacher evaluation process with similar attitudes and utilized similar steps. Both principals have maintained trusting relationships with their teachers through the PERA reforms and throughout the implementation of the new evaluation processes.

The findings of this study indicate the need for further study of the student growth models that are used in schools across Illinois and the United States. Given that many teachers and researchers have major concerns about the use of test scores in teacher evaluation and that there is wide variability in the processes and their application, it is necessary to determine if the intent of the teacher evaluation reforms have been met. If we consider that the two purposes of teacher evaluation are ensuring the quality of the teacher workforce and improving the professional practice of teachers, current approaches may not be meeting those goals. It is unlikely that individual ineffective teachers are being identified, if all teachers in a school or district are being measured on a single growth goal that does not necessarily apply to their specific students or their subject area. It is also unlikely that ineffective teachers will be identified if teachers are allowed to create their own goals that are easily attainable. The silver lining of the use of student test data seems to be that teachers who are evaluated using student growth models like the SLO or all-in approach are looking at student data for instructional purposes more frequently. It is an admirable and important goal to ensure that all students have

high quality teachers and that ineffective teachers are removed or remediated, but the models in use are fraught with potential problems, and the benefits of using these models do not seem to outweigh the costs.

References

- American Statistical Association. (2014). *ASA statement on using value-added models or educational assessment*. Retrieved from American Statistical Association website: https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf
- Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. New York, NY: Routledge.
- Amrein-Beardsley, A., Pivovarov, M., & Geiger, T. J. (2016). Value-added models: What the experts say. *Phi Delta Kappan*, 98, 35-40.
- Archer, J., Kerr, K. A., & Pianta, R. C. (2014). Why measure effective teaching? In K. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 1-5). San Francisco, CA: Jossey-Bass.
- Baker, E., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., . . . Shepard, L. (2010). *Problems with the use of student test scores to evaluate teachers* (Briefing Paper #278). Washington, DC: The Economic Policy Institute.
- Balakit, M. (2015, February 5). Teachers union sues over use of scores in evaluation. *Tennessean*. Retrieved from <https://www.tennessean.com/story/education/2015/02/05/teachers-union-sues-use-scores-evaluations/22946247/>
- Bandura, A., & Wessels, S. (1997). *Self-efficacy*. New York, NY: W.H. Freeman.
- Bass, B. (1999). Two decades of research and development in transformational leadership. *European Journal of Work & Organizational Psychology*, 8, 9-32.

- Beck, M. (2016). *Teacher perceptions of evaluation: Purpose and efficacy of value-added and performance-based evaluation approaches* (Doctoral dissertation, Fordham University). Available from ProQuest. (UMI No. 10145754)
- Berliner, D. (2011). Rational responses to high stakes testing; The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, 41(3), 287-302.
- Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record*, 116, 1-31.
- Berliner, D. (2018, August 11). *Response: The role of student test scores in teacher evaluations* [Web log post]. Retrieved from https://blogs.edweek.org/teachers/classroom_qa_with_larry_ferlazzo/2018/08/response_the_role_of_student_test_scores_in_teacher_evaluations.html
- Booth, W. C., Colomb, G. G., & Williams, J. M. (2008). *The craft of research*. Chicago, IL: The University of Chicago Press.
- Brandt, C., Thomas, J., & Burke, M. (2008). *State policies on teacher evaluation practices in the Midwest Region* (Report No. 004). Washington, DC: U.S. Department of Education.
- Brown, E. (2015, March 1). A closer look at four New Mexico teachers' evaluations. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/local/wp/2015/03/01/a-closer-look-at-four-new-mexico-teachers-evaluations/?noredirect=on&utm_term=.2ddfa54c720b
- Bryk, A. S. (2010). Organizing schools for improvement. *The Phi Delta Kappan*, 91(7), 23-30.
- Bryk, A., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York, NY: Russell Sage Foundation.

- Bryk, A. S., & Schneider, B. (2003). Trust in schools: A core resource for school reform. *Educational Leadership*, 60(6), 40-45. Retrieved from <http://www.ascd.org/publications/educational-leadership/mar03/vol60/num06/Trust-in-Schools@-A-Core-Resource-for-School-Reform.aspx>
- Burns, J. M. (1978). *Model of transactional and transformational leaders*. New York, NY: Harper & Row.
- Calabrese, R. L., Sherwood, K., Fast, J., & Womack, C. (2004). Teachers' and principals' perceptions of the summative evaluation conference: An examination of Model I theories-in-use. *The International Journal of Educational Management*, 18(2/3), 109-117.
- Callahan, K., & Sadeghi, L. (2015). Teacher perceptions of the value of teacher evaluations: New Jersey's ACHIEVE NJ. *NCPEA International Journal of Educational Leadership Preparation*, 10, 46-59.
- Carroll, D. H. (1997, January). *A comparison of clinical supervision and evaluation*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX.
- Carbaugh, B., Marzano, R., & Toth, M. (2017). *The Marzano focused teacher evaluation model: A focused, scientific-behavioral evaluation model for standards-based classrooms*. West Palm Beach, FL: Marzano Center.
- Center on Great Teachers & Leaders at American Institutes for Research. (2016). *Databases on state teacher and principal evaluation policies*. Retrieved from resource.tqsource.org/stateevaldb/StateRoles.aspx

- Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS education value-added assessment system. *Education Policy Analysis Archives*, 22(98), 1-39. doi: 10.14507
- Creswell, J. W. (2014). *Research Design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches*. Thousand Oaks, CA: Sage.
- Crystal, A. D. (2014). *Teachers' perceptions of the implementation of a value-added evaluation system* (Doctoral dissertation, American International College). Available from ProQuest. (UMI No. 3583641)
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: ASCD.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: ASCD.
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.
- Danielson, C. (2013). The framework for teaching evaluation instrument. Retrieved from <http://www.loccsd.ca/~div15/wp-content/uploads/2015/09/2013-framework-for-teaching-evaluation-instrument.pdf>
- Danielson Group. (2019). *Evolution of the framework*. Retrieved from <https://danielsongroup.org/our-story>
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93, 8-15.

- Education Commission of the States. (2010). Teacher merit pay. *The Progress of Education Reform*, 11(3), 1-4.
- Education Commission of the States. (2016). *National snapshot*. Retrieved from <http://www.ecs.org/ec-content/uploads/Use-of-Student-Test-Scores-In-Teacher-Evaluations.pdf>
- Education Commission of the States. (2018). *Policy snapshot: Teacher evaluations*. Retrieved from https://www.ecs.org/wp-content/uploads/Teacher_Evaluations.pdf
- Education Commission of the States. (2020). *State education policy tracking*. Retrieved from <https://www.ecs.org/state-education-policy-tracking/>
- Fausset, R., & Blinder, A. (2015, April 14). Atlanta school workers sentenced in test score cheating case. *The New York Times*. Retrieved from <https://www.nytimes.com/2015/04/15/us/atlanta-school-workers-sentenced-in-test-score-cheating-case.html>
- Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, 43(2), 100-107.
- Flynn, S. I. (2013). Transformational and transactional leadership. *Research Starters: Sociology*, 1(1), 1-6.
- Forsyth, P. B., Barnes, L. B., & Adams, C. M. (2005). Trust-effectiveness patterns in schools. *Journal of Educational Administration*, 44, 122-141. doi:10.1108/09578230610652024
- Foutch, D. R. (2017). *Teacher perceptions of PERA: Influences on teacher morale, satisfaction, and motivation* (Doctoral dissertation, Southern Illinois University at Edward). Available from ProQuest. (UMI No. 10268172)

- Gabriel, R., & Allington, R. (2012). The MET Project: The wrong 45 million dollar question. *Educational Leadership*, 70(3), 44-49.
- German, J. M. (2014). *Teachers' perceptions of self-efficacy: The impact of teacher value-added* (Doctoral Dissertation, Ashland University). Available from ProQuest. (UMI No. 3671351)
- Gershenson, S., & Hayes, M. S. (2018). The implications of summer learning loss for value-added estimates of teacher effectiveness. *Educational Policy*, 32, 55-85.
- Ginsberg, R., & Berry, B. (1989). Influencing superiors' perceptions: The fudge factor in teacher and principal evaluation. *The Urban Review*, 21(1), 15-34.
- Glazerman, S., & Seifullah, A. (2010). *An evaluation of the teacher advancement program (TAP) in Chicago: A year two impact report*. Washington DC: Mathematica.
- Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York, NY: Holt, Rinehart, and Winston.
- Growth Through Learning. (2012). *Illinois performance evaluation frequently asked questions*. Retrieved from www.growththroughlearningillinois.org
- Guarnina, C., Reckase, M., Stacy, B., & Wooldridge, J. (2015). A comparison of student growth percentile and value-added models of teacher performance. *American Statistical Association*, 2, 66-76.
- Gurl, T. J., Caraballo, L., Grey, L., Gunn, J. H., Gerwin, D., & Bembunty, H. (2016). *Policy, professionalization, privatization, and performance assessment: Affordances and constraints for teacher education programs*. Switzerland: Springer.

- Hawley, L. R., Bovaird, J. A., & Wu, C. (2017). Stability of teacher value-added rankings across measurement model and scaling conditions. *Applied Measurement in Education, 30*, 196-212. doi: 10.1080/08957347.2017.1316273
- 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*, 1-28.
- Hertel, P. E. (2014). *Teacher and administrator perceptions of the evaluation process in Illinois* (Doctoral dissertation, Western Illinois University). Available from ProQuest. (UMI No. 3671484)
- Horn, S., & Sanders, W. (2009). *Research findings from the Tennessee Value-Added - SAS*. Retrieved from https://www.sas.com/govedu/edu/ed_eval.pdf
- Hunter, M. C. (1982). *Mastery teaching*. Thousand Oaks, CA: Corwin Press.
- Illinois State Board of Education. (2015). *ISBE non-regulatory guidance on PERA and SB7*. Retrieved from https://www.isbe.net/documents/pera_guidance.pdf
- Illinois State Board of Education. (2018). *Illinois report card*. Retrieved from https://www.illinoisreportcard.com/state.aspx?stateid=IL&source=teachers&source2=teacher_evaluation
- Illinois State Board of Education. (2019). *Educator evaluations-PEAC: Laws and administrative rules*. Retrieved from <https://www.isbe.net/pages/educator-evaluations.aspx>
- Illinois State Board of Education. (2020). *Illinois report card*. Retrieved from <https://www.illinoisreportcard.com>
- Jiang, J. Y., Sparte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher, 44*(2), 105-116.

- Jodlowski, E. (2016). *Value-added measures of student growth: Where we are in Illinois post-PERA* (Doctoral Research Project, Southern Illinois University at Edwardsville). Available from ProQuest. (UMI No. 10132960)
- Johnson, S. M. (2015). Will VAMS reinforce the walls of the egg-crate school? *Educational Researcher*, 44(2), 117-126. doi:10.3102/0013189X15573351
- Kane, M. T. (2017). *Measurement error and bias in value-added models*. Princeton, NJ: Educational Testing Service. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1168716.pdf>
- Konstantopoulos, S. (2014). Teacher effects, value-added models, and accountability. *Teachers College Record*, 116, 1-21.
- Lavigne, A. L., & Good, T. L. (2014) *Teacher and student evaluation: Moving beyond the failure of school reform*. New York City, NY: Routledge.
- Louis, K. S. (2007). Trust and improvement in schools. *Educational Policy and Administration*, 8(1), 1-24. doi:10.1007/s10833-006-9015-5
- Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership*, 70(3), 14-19.
- Marzano, R. J. (2014). Interview by Angela E. Quinn. Looking at the bigger picture with Dr. Robert Marzano: Teacher evaluation and development for improved student learning. *The Delta Kappa Gamma Bulletin*, 81(1), 12-18.
- Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: ASCD.
- McGreal, T. L. (1983). *Successful teacher evaluation*. Alexandria, VA: ASCD.

- 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. *NCPEA Education Leadership Review*, 16(1), 16-30.
- Miles, M. B., Huberman A. M., Saldana, J. (2014). *Qualitative data analysis*. Newbury Park, CA: Sage.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- National Council on Teacher Quality. (2015). *State of the states 2015: Evaluating teacher, leading, and learning*. Retrieved from <https://www.nctq.org/publications/State-of-the-States-2015:-Evaluating-Teaching,-Leading-and-Learning>
- National Council on Teacher Quality. (2017). *State teacher policy database*. Retrieved from <https://www.nctq.org/yearbook/customReport>
- Pearson Education, Inc. (2004). *Policy report: Value-added assessment systems*. Retrieved from https://images.pearsonassessments.com/images/tmrs/tmrs_rg/ValueAdded.pdf?WT.mc_id=TMRS_Value_Added
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Thousand Oaks, CA: Corwin.
- Pianta, R., & Kerr, K. A. (2014). Measuring effective teaching: The future starts now. In K. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 583-590). San Francisco: CA: Jossey-Bass.
- Popham, W. J. (2013). Can classroom assessments of student growth be credibly used to evaluate teachers. *English Journal*, 103(1), 34-39.

- Pressley, T., Roehring, A. D., & Turner, J. E. (2018). Elementary teachers' perceptions of a reformed teacher-evaluation system. *The Teacher Educator, 53*, 21-43.
doi:10.1080/08878730.2017.1391362
- Ravitch, D. (2014). Foreward. *Teachers College Record, 116*, 1-4.
- Reform Support Network. (2015). *Emerging approaches to measuring student growth*. Retrieved from <https://www2.ed.gov/about/inits/ed/implementation-support-unit/tech-assist/emergapprotomeasurstudgrowth.pdf>
- Ruzek, E.A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *Journal of Early Adolescence, 35*, 852-882.
- Sawchuck, S. (2014). *Houston union sues district over 'value added' evaluations*. Retrieved from http://blogs.edweek.org/edweek/teacherbeat/2014/05/houston_union_sues.html
- Schafer, W. D., Lissitz, R. W., Zhu, Xiaoshu, Zhang, Y., Hou, X., & Li, Y. (2012). Evaluating teachers and schools using student growth models. *Practical Assessment, Research, and Evaluation, 17*, 1-21.
- Shinkfield, A. J., & Stufflebeam, D. L. (1996). *Teacher evaluation: Guide to effective practice*. Boston: Kluwer Academic Publishers.
- Simons, H. (2012). *Evolution and concept of case study research*. London, UK: Sage.
- Slaten Frasier, A. M. (2017). *Examining teacher perceptions of the relationships between evaluation policy and teacher practice in a North Carolina school system* (Doctoral dissertation, Michigan State University). Available from ProQuest. (UMI No. 10684225)
- Sparks, S. (2011, November 16). Evaluation reforms put partnerships to the test. *Education Week, 31*(12), S8-S11.

- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002) Policy implementation and cognition: reframing and refocusing implementation research. *Review of Educational Research*, 72(3) 387-431.
- Stiggins, R. (2014). *Defensible teacher evaluation: Student growth through classroom assessment*. Thousand Oaks, CA: Corwin.
- 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*, 64(4), 339-355.
- Tesch, R. (1990). *Qualitative research: Analysis types and software tools*. Bristol, PA: Routledge.
- Tschannen-Moran, M. (2014). *Trust matters: Leadership for successful schools*. San Francisco, CA: Jossey-Bass.
- Tschannen-Moran, M., & Hoy, W. K. (2000). A multidisciplinary analysis of the nature, meaning, and measurement of trust, *Review of Educational Research*, 70(4), 547-593.
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998), Teacher efficacy: Its meaning and measure, *Review of Educational Research*, 68, 202-48.
- Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: ASCD.
- Tuytens, M., & Devos, G. (2010). The influence of school leadership on teachers' perception of teacher evaluation policy. *Educational Studies*, 36(5), 521-536.
- U.S. Department of Education. (2004). *NCLB executive summary*. Retrieved from <http://www2.ed.gov/nclb/overview/intro/execsumm.html>

- U.S. Department of Education. (2009). *Race to the Top program executive summary*. Retrieved from <http://www2.ed.gov/programs/racetothetop/>
- U.S. Department of Education. (2010). *Secretary of Education Arne Duncan's remarks at OECD's release of the program for international student assessment (PISA) 2009 results*. Retrieved from <https://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-oecd-release-program-international-student-assessment->
- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013). WWC review of the report: *Teacher incentives and student achievement: Evidence from New York City Public Schools*. Retrieved from <http://whatworks.ed.gov>
- Viar, M. A. (2016). *An analysis of the relationship between the perceptions of value-added measurement and teacher job satisfaction* (Doctoral dissertation, University of Memphis). Retrieved from ProQuest. (UMI No. 10587627)
- Walker, T. (2014). *New study strikes latest blow against "value added" teacher evaluation*. *NEA Today*. Retrieved from <http://neatoday.org/2014/05/30/new-study-strikes-latest-blow-against-value-added-teacher-evaluation-2/>
- Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunk, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. (2nd ed.). Retrieved from The New Teacher Project website: https://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage.

Appendix A

Email Requesting Access From Superintendents

Dear [Insert Name of Superintendent],

My name is Emily Weidner. I am the principal of White Heath Elementary School in Monticello CUSD #25, and I am also a doctoral candidate at the University of Illinois at Urbana-Champaign in the Department of Education Policy, Organization, and Leadership. Under the supervision of my advisor, Dr. Rachel Roegman, I am conducting a research study focused on the student growth component of teacher evaluation in Illinois. The purpose of this study is to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using the Student Learning Objective process as compared to the all-in approach and to what extent relational trust between the principal and the teacher promotes teachers' engagement with student growth approaches.

If you allow me to complete research in your school district, I would like to collect data from the following sources:

- Individual Interviews
 - Principal
 - 5-6 teachers

- Documents
 - Forms and documents related to teacher evaluation and the student growth model in the school and/or district

Prior to engaging in any research, all participants will be notified and asked to sign letters of informed consent. Local and district research policies will be followed.

If you are interested in participating in this study or have additional questions, please reply via email (bridwell@illinois.edu) or by phone 217-202-5766. Upon your acceptance, all consent forms and letters of invitation to potential participants will be sent for your review.

Sincerely,

Emily Weidner

Appendix B

Email Confirming Study Eligibility and Soliciting Participation From Principals

Dear [Insert Name of Principal]

My name is Emily Weidner. I am the principal of White Heath Elementary School in Monticello CUSD #25, and I am completing my Doctor of Education degree in Education Policy, Organization, and Leadership at the University of Illinois at Urbana-Champaign.

Under the supervision of my advisor, Dr. Rachel Roegman, I am conducting a research study focused on the student growth component of teacher evaluation in Illinois. The purpose of this study is to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using the Student Learning Objective process as compared to the “All In” approach and to what extent relational trust between the principal and the teacher promotes teachers’ engagement with student growth approaches. I am emailing you because your superintendent suggested you as a potential participant in this study.

If you choose to participate in this study, I would collect information from the following sources:

- Individual Interviews (one 45 minute interview with a possible 15 minute follow up interview at a later date)
 - Principal
 - 5-6 teachers
- Documents
 - Forms and documents related to teacher evaluation in the school and/or district

In order to participate in the study, you must meet the following requirements:

- Must be a currently practicing principal in an Illinois public school who practiced before and after the implementation of PERA/SB7 as an evaluator of certified staff.
- School must use either:
 - An All-In approach to student growth in teacher evaluation, in which all teachers are scored based on the same student growth goal for their summative teacher evaluation.
 - A Student Learning Objective approach in which teachers write their own student growth goals to be used in their teacher evaluations.

Prior to engaging in any research, all participants will be notified and asked to sign letters of informed consent. Local and district research policies will be followed. If you are interested in participating in this study or have additional questions, please reply via email (bridwell@illinois.edu) or by phone 217-202-5766. Upon your acceptance, all consent forms and letters of invitation to potential participants will be sent for your review.

Sincerely,

Emily Weidner

Appendix C

Email Requesting Participation From School Teachers

Dear [Insert Name of Teacher],

My name is Emily Weidner. I'm the principal of White Heath Elementary School in Monticello CUSD #25, and I am completing my Doctor of Education degree in Education Policy, Organization, and Leadership at the University of Illinois at Urbana-Champaign. After speaking with [Insert Name of Principal or Superintendent], I am contacting you because you are a teacher who has been evaluated using the [Insert SLO or All-In] approach to the student growth model for teacher evaluation in Illinois.

Under the supervision of my advisor, Dr. Rachel Roegman, I am conducting a research study focused on the student growth component of teacher evaluation in Illinois. The purpose of this study is to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using the Student Learning Objective process as compared to the "All In" approach and to what extent relational trust between the principal and the teacher promotes teachers' engagement with student growth approaches.

If you agree to participate in my study, I would like to collect information from the following sources:

- Individual Interview
 - One approximately 45 minute interview with a possible 15 minute follow up interview.
- Documents
 - Blank forms and documents related to teacher evaluation in the school and/or district.

Prior to engaging in any research, all participants will be notified and asked to sign letters of informed consent. Local and district research policies will be followed.

If you are interested in participating in this study or have additional questions, please reply via email (bridwell@illinois.edu) or by phone 217-202-5766.

Sincerely,

Emily Weidner

Appendix D

Teacher Evaluation Policy Reforms Involving Student Data: Teachers' Perceptions

You are being asked to participate in a voluntary research study. The purpose of this study is to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using the Student Learning Objective process as compared to the "All In" approach and to what extent relational trust between the principal and the teacher promotes teachers' engagement with student growth approaches. Participating in this study will involve one to two in-person interviews. The first interview will be less than one hour and if needed, the second follow-up interview would also be less than one hour. Risks related to this research include loss of confidentiality; benefits related to this research include informing policymakers of the influence of current policies and practices and providing insight for those who evaluate teachers.

Researcher Name and Title: Emily Weidner Doctoral Student at the University of Illinois
Contact Information: 217-202-5766 bridwell@illinois.edu
Principal Investigator Name and Title: Dr. Rachel Roegman
Department and Institution: Department of Educational Policy, Organization, and Leadership
Contact Information: 217.333.0230 roegman@illinois.edu

Why am I being asked?

You are being asked to be a participant in a research study about teachers' perceptions of teacher evaluation policy reforms involving student data.

The purpose of this research is to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using the Student Learning Objective process as compared to the "All In" approach and to what extent relational trust between the principal and the teacher promotes teachers' engagement with student growth approaches. You have been asked to participate in this research because you are an elementary school teacher in Illinois who has been evaluated using student assessment data. Approximately twelve participants will be involved in this research at the University of Illinois at Urbana-Champaign.

Your participation in this research is voluntary. Your decision whether or not to participate will not affect your current or future dealings with the University of Illinois at Urbana-Champaign. If you decide to participate, you are free to withdraw at any time without affecting that relationship.

What procedures are involved?

The study procedures are collection of blank artifacts related to the teacher evaluation process and the student growth component. Artifacts that will be reviewed may include, but are not limited to, district teacher evaluation handbooks or guidelines, school board policies related to teacher evaluation, documents used in the teacher evaluation processes, and actual student learning objectives and all-in documents.

Data will also be collected through semi-structured open-ended interviews.

This research will be performed at your school or a location that is most convenient for you. You will need to meet with the interviewer 1-2 times between September and December of 2019. Each of those visits will last less than one hour. A second interview will be requested only if clarification or further information is needed.

What are the potential risks and discomforts?

Participation in this study carries minimal risk. You may want to ensure that the reputation of your district or school will not be harmed by sharing information about processes and policies. There is a risk of loss of confidentiality.

Are there benefits to participating in the research?

There are no direct benefits to you, but the research could have indirect benefits because you will be reflecting on the evaluation process and your role in it. Information shared by teachers could instruct principals on how to improve the process and make it more valuable for those who are intended to improve from it. The research will inform principals regarding how much impact their relationships with teachers have on the perceived effect of the student growth model being used. The research could also inform policymakers about what effect the reform has had on teachers, principals, and schools. Finally, the research should inform policymakers about the effectiveness of each approach to assessing student growth as reported by the principals and teachers.

What other options are there?

You have the option to not participate in this study.

Will my study-related information be kept confidential?

We will use all reasonable efforts to keep your personal information confidential, but we cannot guarantee absolute confidentiality. When this research is discussed or published, no one will know that you were in the study. But, when required by law or university policy, identifying information (including your signed consent form) may be seen or copied by: a) The Institutional Review Board that approves research studies; b) The Office for Protection of Research Subjects and other university departments that oversee human subjects research; c) University and state auditors responsible for oversight of research.

Will I be reimbursed for any expenses or paid for my participation in this research?

You will not be offered payment for being in this study.

Can I withdraw or be removed from the study?

If you decide to participate, you are free to withdraw your consent and discontinue participation at any time. The researchers also have the right to stop your participation in this study without your consent if they believe it is in your best interests or if you were to object to any future changes that may be made in the study plan.

Will data collected from me be used for any other research?

Your information will not be used or distributed for future use, even if identifiers are removed.

Who should I contact if I have questions?

Contact the researchers, Emily Weidner at 217-202-5766 or bridwell@illinois.edu or Dr. Rachel Roegman at 217.333.0230 or roegman@illinois.edu if you have any questions about this study or your part in it, or if you have concerns or complaints about the research.

What are my rights as a research subject?

If you have any questions about your rights as a participant in this study, please contact the University of Illinois at Urbana-Champaign Office for the Protection of Research Subjects at 217-333-2670 or irb@illinois.edu.

I have read the above information. I have been given an opportunity to ask questions and my questions have been answered to my satisfaction. I agree to participate in this research. I will be given a copy of this signed and dated form.

Signature

Date

Printed Name

Signature of Person Obtaining Consent

Date (must be same as subject's)

Printed Name of Person Obtaining Consent

Appendix E

Interview Protocol for Principals

Script: Thank you for meeting with me. My name is Emily Weidner. I'm a former high school teacher and current elementary school principal. I'm also a doctoral student at the University of Illinois. The purpose of my study is to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using the Student Learning Objective process and the all-in approach. When I say Student Learning Objective, I'm referring to a learning goal, assessment, measurement model, and student growth prediction. Growth predictions are made based on the needs of students in specific teachers' classrooms. When I say, all-in approach, I mean a student growth model in which the same learning goals, assessments, and outcome are applied to all teachers in the school or district regardless of the teachers' subject area.

I plan to interview teachers and principals in one school that uses an all-in approach and one school that uses the Student Learning Objective approach and look for themes in their perceptions of student growth and how their principal's approach influences their engagement in the process.

Ice breaker:

1. Tell me a little be about yourself and your professional history.
 - a. How long have you been a principal?
 - b. How long have you worked in this district?

Interview Questions:

1. To what extent did you help with the planning and implementation of PERA reforms in this school or district?
 - a. Please describe how the plan was developed.
 - i. Has this approach been used since the beginning of the reform or did the district change approaches?
 - b. How did the teachers react to the new policy at the time of introduction and implementation?
 - c. How did you feel about the new policy and procedures at the time of implementation?
 - d. What challenges did your district or school face at the time of implementation?
 - i. How did that influence the perception of the teacher evaluation process with your staff?
2. Walk me through your current teacher evaluation process. Please describe your approach to teacher evaluation including your approach to meetings, conversations that you have with teachers, observations and other data that you collect, and how you give summative ratings?
3. How do you work with teachers so that the teacher evaluation process is meaningful?

4. How do you establish trust with your teachers?
5. In your opinion, how has the new teacher evaluation model using student growth affected your relationship with teachers?
6. How does trust between the principal and teachers influence the process of setting student growth goals?
7. In your opinion, how has the new teacher evaluation model using student growth affected relationships between teachers?
 - a. How if it all has it affected trust?
 - b. Are teachers more or less likely to collaborate?
8. To what extent do you think that the student growth model has improved teaching and learning in your school, if at all?
 - a. To what extent do teachers implement specific changes to curriculum, instruction, assessment or professional development based on the student growth process?
 - b. To what extent do teachers narrow the curriculum to focus on student growth goals?
 - c. Do you think it influences teachers to use student data more regularly?
9. How do you think teachers perceive the student growth component of the teacher evaluation process now?
10. How do you think teacher perception of the student growth component has changed between the implementation and now?
11. To what extent do you feel that your district's student growth model is easy for teachers to understand?
 - a. What steps have you taken to ensure that teachers understand?
12. To what extent do you feel that your district's student growth model is a fair way to assess teacher performance?
 - a. If you answered that it is a) not easy to understand and/or b) not fair, how do you feel that this impacts your relationships and trust with your teachers?
13. How has the student growth model influenced your teachers' overall summative ratings?
 - a. Are summative ratings generally higher or lower since the inclusion of student growth?
 - i. Has the inclusion of student growth ever resulted in a teacher being rated higher or lower than you thought they should be?
 - ii. Have you altered professional practice ratings due to student growth scores?
14. What changes do you believe your district should make to the student growth model?
15. If you evaluated teachers prior to the passage of the Perform Evaluation Reform Act (PERA), please tell me about that process.
 - a. Which plan do you prefer?
 - b. Why?
16. Is there anything else you would like to add?

Appendix F

Interview Protocol for Teachers

Script: Thank you for meeting with me. My name is Emily Weidner. I'm a former high school teacher and current elementary school principal. I'm also a doctoral student at the University of Illinois. The purpose of my study is to develop an understanding of the perceptions of elementary school teachers in Illinois who have been evaluated using the Student Learning Objective process and the all-in approach. When I say Student Learning Objective, I'm referring to a learning goal, assessment, measurement model, and student growth prediction. Growth predictions are made based on the needs of students in specific teachers' classrooms. When I say, all-in approach, I mean a student growth model in which the same learning goals, assessments, and outcome are applied to all teachers in the school or district regardless of the teachers' subject area.

I plan to interview teachers and principals in one school that uses an all-in approach and one school that uses the Student Learning Objective approach and look for themes in their perceptions of student growth and how their principal's approach influences their engagement in the process.

Ice breaker:

1. Tell me a little be about yourself and your professional history.
 - a. How long have you been teaching?
 - b. How long have you been teaching in this district?

Interview Questions:

1. To what extent did you help with the planning and implementation of PERA reforms in this school or district?
 - a. Please describe how the plan was developed.
 - b. What type of student growth model did your district choose? (SLO or All-In)
 - c. Has this approach been used since the beginning of the reform or did the district change approaches?
 - d. Describe how you felt about the new policy at the time of introduction and implementation.

Do you think that your colleagues generally felt the same way?

2. Please walk me through the teacher evaluation process that you are evaluated under currently.
 - a. What steps are you required to complete during the process?
 - b. Describe the amount of time that is required to complete these steps.
 - c. How are the expectations for the steps that you must complete communicated to you?
 - d. What steps is your evaluator required to complete during the process?
 - e. How do you establish student growth goals?
 - f. Are your student growth goals in your content area?
3. Describe the pros and cons of your current evaluation process.

4. One of the policy goals of including student growth in teacher evaluations was to ensure that ineffective teachers were identified. Do you feel that the inclusion of growth has helped to identify poorly performing teachers?
 - a. Do you feel that it has helped to better identify the most effective teachers?
5. To what extent do you believe that the student growth model used in your school and your rating from the model is easy to understand?
 - a. How does your evaluator aid in your understanding of the student growth model and process?
6. To what extent do you believe that the student growth model used in your school is a fair and accurate way to assess your performance as a teacher?
 - a. Does your student growth rating usually match your professional practice rating?
 - b. Does your student growth rating usually impact your overall rating positively or negatively?
 - c. Is your student growth rating consistent from year to year?
7. To what extent has the student growth component influenced your teaching?
 - a. What changes have you made to planning, collaboration, curriculum, instruction, and/or assessment?
 - i. Have you narrowed the curriculum in order to place more focus on student growth goals?
 - ii. If so, has this been a positive or negative change?
8. How has your principal's or evaluator's approach to teacher evaluation influenced your perception of the process and the value that you find in the process?
 - a. Are you encouraged to use student data more regularly or in new ways?
 - b. Do you focus on different learning goals?
 - c. How does it influence your development as a professional?
9. How do you think your principal perceives the student growth component of teacher evaluation?
 - a. How does the principal's perception influence your perception of the process?
10. How has the student growth component of the teacher evaluation process affected the relationship between teachers and the principal?
11. What role does trust between the principal and the teacher play in the formation of student growth goals?
12. What are the key ways that an evaluator can improve the teacher evaluation process for teachers?
13. How has the student growth component of the teacher evaluation process affected relationships between teachers in your building?
 - a. Are teachers more or less likely collaborate?
14. What changes do you believe your district should make to the student growth model?
15. Were you evaluated as a teacher prior to PERA?
 - a. Please describe that process.
 - b. What were the pros and cons of the prior process?
16. Which teacher evaluation model do you prefer and why?
 - a. How does the inclusion of student growth influence your preference between the previous and current system?
17. Is there anything else you would like to add?

Appendix G

Coding Chart

Open Codes	Grouping open codes	Key Findings
Understanding Teacher perception Background Process Easier Time saver Meetings Fair Accurate Artifacts Nervous Consistent	<ul style="list-style-type: none"> • How teachers understand the process • Fairness & Accuracy • Time & value • Teacher preference of processes 	Teachers at Monroe School had a stronger understanding and a more positive view of the student growth model at their school than teachers at Roosevelt School.
Data usage Curriculum Instruction Professional development Teacher collaboration	<ul style="list-style-type: none"> • Influence of SLO & All-In on teaching • Influence of SLO & All-in on collaboration 	Teachers at Roosevelt School reported that the student growth model influenced their professional practice more than teachers at Monroe School.
Principal trust Teacher trust Communication Feedback Conversations Mentor Observation	<ul style="list-style-type: none"> • Relationships before, during, and after PERA • Principal steps in evaluation • Teachers value feedback 	The principals at Monroe School and Roosevelt School had a similar approach to teacher evaluation that seemed to lead to the maintenance of high levels of trust in spite of teacher concerns about the processes.