Examination of Teacher Observation Dynamics: Role of Observer Effort on Teacher Growth

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Michael Shaun Bury

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Kyla Wahlstrom, Adviser

July 2017

Acknowledgments

I wish to extend my endless thanks to my adviser, Dr. Kyla Wahlstrom, who continued to provide guidance, support, encouragement, insight, patience, and grace over the duration of this journey. It was quite the climb, and I can't count the number of times I rounded the switchback, expecting to see the summit, only to see more of the trail. When that switchback had a fallen tree blocking the path, and I wanted to just drop my pack and head back down the mountain, she was there, hand extended, pulling me to the other side. Dr. Wahlstrom is a great scholar, but most importantly, she is an exceptional person.

I also want to thank the university faculty members who have shared their time and wisdom with me. I particularly want to thank those who served on my examination committees for their insight and attention to detail: Dr. Karen Seashore, Dr. Ann Werner, Dr. Jennifer York-Barr, Dr. Peter Demerath, Dr. Katie Pekel, and Dr. Nicola Alexander.

I owe a special thanks to my wife, Kristin, who kept a candle in the window so I could always find my way back home. No one has a better life partner than I do. Brianne, Grace, Keira, and Riley, thanks for being understanding when dad couldn't play and for playing extra hard when I could.

Finally, I owe a great debt to my parents, Kathy and Jerry Bury, for a lifetime of support, love, and encouragement regardless of my shenanigans.

Abstract

This study examined the teacher observation cycle to understand the effect of observer knowledge, observer effort, observer power, and school culture on teachers' perceptions of whether the observation process helped them grow, implement strategies, or increase student learning.

The concepts of power and expertise were defined by blending the definition of expertise of Berliner (2004) with the framework of power developed by Michelson (2001). Surveys and interviews were used to gather data on teacher perceptions and provide additional context and understanding on these perceptions. Linear regression was applied to the survey data to determine the relationship and significance between variables. Interviews were coded originally based on defined variables, but two of these variables had subcomponents that emerged as significant in the final analysis.

The results indicate that the effort and the content and pedagogical knowledge of the observer are more significant factors in perceptions of teacher growth and implementation, as well as in perceived student learning, than the factors of observer power or school culture. Therefore, observers and school systems that want to improve teacher quality through the observation process should try to match teacher and observers in like-content areas, train observers on pedagogy and the evaluation process, and prioritize teacher observations over other work demands.

Table of Contents

List of Tables

List of Figures

Chapter One: Critical Issue

Introduction

In December 2015, President Obama signed the Every Student Succeeds Act (ESSA) which was a reauthorization of the Elementary and Secondary Education Act (ESEA) and replaced the Bush era iteration of ESEA, No Child Left Behind (NCLB). The Council of Chief State School Officers (CCSSO, 2016) made some comparisons between these two reauthorizations that have an impact on selection and development of teachers. First, ESSA drops NCLB's requirement that teachers in Title I schools and core subjects be "highly qualified" as defined by statute. Instead, teachers must meet the state's licensure and certification standards. Second, ESSA adds a requirement that state plans include provisions to ensure that Title I schools are not disproportionately staffed by out-of-license, inexperienced, and/or ineffective teachers and principals. This stipulation is an important element in reducing the achievement gap, as research demonstrates that teacher quality is the most significant school-based factor in student achievement (Louis et al., 2010; Leithwood, Louis, Anderson, & Wahlstrom, 2004; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2000; Rowan, Correnti & Miller, 2002; Wright, Horn, & Sanders, 1997). This stipulation also gives schools an important responsibility for selecting and maintaining an effective teaching force as one element of reducing the achievement gap.

Teacher evaluation systems are the mechanism to meet this responsibility. Evaluations are used by schools in districts in two ways. First, evaluations can be summative and used to measure a teacher's impact on student achievement. This is important to the selection process during a teacher's probationary period. Second,

evaluations can be used formatively to provide teachers targeted feedback and suggestions for development. Formative evaluations are also important during the probationary period, but they are essential for maintaining and increasing teacher effectiveness over time as a means to address educational equity. Schools, therefore, need to understand what evaluation system structures and practices lead to effective summative and formative evaluations. Research can identify these effective practices.

Perspectives on Teacher Evaluation

Given the scope of research on teacher evaluation, it is useful to focus on three specific areas that have guided and established the research base. The first area focuses on overall effectiveness of a program's design to determine if a teacher evaluation program is meeting the purposes for which it was designed. Sample research topics include looking at a system's objectivity (Heneman & Milanowski, 2003), effectiveness (Wise, Darling-Hammond, McLaughlin & Bernstein, 1984), impact on student achievement (Kupermintz, 2003), or underlying philosophy (Taut, Santelices, Araya, & Manzi, 2010). While these researchers have different research perspectives, they all assume that a well-designed system of evaluation leads to a high quality evaluation. They also assume that interaction between system elements is more influential on validity and reliability than are the individual elements.

Expertise plays a dual role in a second body of scholarship. First, this body focuses on the usefulness of specific types of evaluation evidence to evaluate teacher's expertise in content and pedagogy. The specific types of evidence considered include multiple lines of evidence (Bill &Melinda Gates Foundation, 2010; Peterson, 1987), rating criteria (Kane, McCaffrey, Miller & Staiger, 2013; Epstein, 1985), and specific

data collection tools (Kane & Staiger, 2012; Evertson & Burry, 1989). Several of these aforementioned studies were conducted as a part of the Measures of Effective Teaching (MET) Project funded by the Bill and Melinda Gates Foundation. Second, the MET researchers generally believe that high quality evidence results in high quality evaluation, and collecting high quality evidence depends on the evidence collection expertise of the evaluators. Furthermore, they assert that validity and reliability of evidence increases the utility of evaluation evidence. This latter belief is the defining element of this research body.

The final group of scholars focus their research on the stakeholders, both the evaluators and the evaluated, involved in the evaluation process. Considering the summative and formative purposes of evaluation, it is clear that evaluations are a tool through which evaluators attempt to influence the behavior of the evaluated. As such, the interactions between the stakeholders involve uses of power (French and Raven, 1959). However, sources of power vary between stakeholder groups. French and Raven (1959) noted that power is based in both positional and personal sources. Administrators who conduct evaluations inherently have positional power and could also have personal power. Peer evaluators might have some positional power, particularly if they are involved in pay for performance evaluations, but personal power, based on their knowledge and experience, is the primary source of influence for peer evaluators. Researchers have examined various stakeholder roles in evaluation processes including administrators (Ovando & Ramirez, 2007), teachers (Ovando & Harris, 1993), students (Bill & Melinda Gates Foundation, 2010), and the interaction between groups (Johnson & Shields, 2007). Research on stakeholders assumes that human perceptions and context

affect evaluation. Additionally, it assumes that understanding stakeholders' perceptions and contexts lead to a better teacher evaluation system.

Gaps in Research

Because there likely are an infinite number of factors that could impact the construction of a teacher observation, it is necessary to narrow the focus. Three factors that emerged in the research literature that clearly affect evaluations are school climate, observer expertise, and observer power.

The research indicates that there is a need to measure teacher expertise in observations, and there is a need for observers to have expertise in the observation instrument (Kane, McCaffrey, Miller & Staiger, 2013). These two elements of expertise appear in the literature and have been studied to determine their reliability (Ho & Kane, 2013) for measuring a teacher's strengths and weaknesses.

Another unexplored area comes from the power dynamic inherent in the observation process. The higher scores associated with teacher selected videos in the MET project (Ho & Kane, 2013) indicate that teachers try to influence principals by demonstrating their best lessons when possible. The MET project also demonstrated that impressions tend to linger over time, so this influence is cumulative. This effect might explain why principals rated their own teachers higher (Ho & Kane, 2013). Principals also influence teachers in the evaluation process through the identification of growth areas. This identification could influence teacher development since the principal has positional power. Many educators perceive that peer evaluators do not have the same positional power and must draw on personal power to influence growth. The literature

also does not address how power issues impact teacher perceptions of the validity of the evaluation experience.

Finally, the literature on school climate and teacher experience suggests that factors outside the evaluation process itself can have an impact on the evaluation (Garmston & Wellman, 1999; Steffy, Wolfe, Pasch & Enz, 2000). Teachers working in a culture that facilitates change can try new strategies with the understanding that their overall performance evaluation will not be affected based on their initial implementation. This could increase their willingness to take a risk based on their evaluation. Teachers in the early phases of their career may place more emphasis on the suggestions of their supervisor because they have not yet connected themselves to the larger profession. Conversely, teachers who have been in the profession for many years may value the goals of the profession over perceived limited feedback from their immediate supervisor.

Research Questions

Given the lack of investigation into the links between expertise and power in teacher evaluation research, three questions emerge:

1. What role does the perceived content area expertise of the observer play in teacher observations for the observer and the observed?
2. What role does the perceived expertise of the observer play in teacher observations for the observer and the observed?
3. How does the perceived power relationship between observer and observed shape teacher observations?

Critical Frameworks

Two frameworks defined the central issues of power and expertise for this study. Michelson (2001) outlined a framework of power that includes five factors of positional power and three attributes of personal power. The five factors of personal power are centrality, criticality, flexibility, visibility, and relevance. The three attributes of personal power are knowledge/information, personal attraction, and effort. Teacher expertise is defined as a set of characteristics by Berliner (2004). These characteristics are aligned with and reinforce Michelson's concepts of knowledge/information and effort. These two frameworks and their underlying concepts will be more completely defined and discussed in Chapter 3.

Data Collection and Analysis

To answer the aforementioned research questions, this study used a combination of surveys and interviews. Since the research questions involved gauging participant perceptions, survey questions that assessed perceptions provided insight into the role that perception plays in the observation.

Interviews were conducted following the administration of the survey. Following data collection, surveys and interviews were subjected to analysis. The survey data was explored using regression, and interviews were thematically analyzed and coded based on emerging findings from the survey analysis.

Limitations

This study had some limiting factors. First, surveys were distributed in the two high schools, each at a staff meeting that occurred at the end of the day. As a result, teachers who were not in attendance at that meeting did not have an opportunity to

participate. Also, even though the teachers were assured of anonymity in completing their surveys, it is suspected that there were still a few respondents who were not comfortable providing some answers that may have been perceived as negative or self-revealing. Second, the research sites are part of the same district. Although this could be viewed positively, since each site follows the same process for observations and uses the same evaluation rubric based on the work of Charlotte Danielson (2007), the findings may not be generalizable to other districts in the state and nation.

Key Terms

There are several terms used in this study that can have different meanings in different setting and/or contexts. For clarification, these terms are defined below for the context of this study.

Observation. The observation process for a teacher consists of a pre-observation meeting between the teacher and observer, an in-class observation by the observer for entire class-period, and a post-observation meeting between the teacher and observer.

Observation Cycle. The observation cycle consists of three sets of observations which collect evidence of proficiency in twenty-two component areas. Building administrators are observers for high cycle and probationary cycles. Peer Evaluators are observers for low cycles.

High Cycle. Every third year, staff in the teacher bargaining unit are considered to be in the "High Cycle" of evaluation as required by Minnesota Teacher Development and Evaluation law (MN statute 122A.40).

Probationary Cycle. Teaching staff in their first three years of employment or first year in the district are considered probationary as defined by Minnesota Teacher Development and Evaluation law (MN statute 122A.40).

Low Cycle. Teachers who are not probationary (i.e., tenured) and who are not in their high cycle are considered to be in their low cycle.

Peer Evaluator. Peer evaluators are continuing contract teachers, with a minimum of 7 years' experience, who are hired for a three-year rotation to observe fellow members of the teacher bargaining unit. Peer evaluators are assigned to low cycle teachers based on matching or similar content areas.

Personal Growth Plan. Teachers develop an individual plan for their own professional development that includes specific goals and action steps to meet those goals.

Rubric. The district's rubric is based on the work of Danielson (2007). Danielson identified 22 domains of teaching and four levels of performance. The rubric is a matrix that has rows with the 22 domains of teaching and columns with the levels of performance. Each domain and performance combination has text describing them.

Summative Evaluator. These are building-level licensed administrators who are assigned to high-cycle and probationary teachers. The assignments for the roster of teachers to be observed by a summative evaluator are made at the building level.

Pedagogy. This term refers to the teaching strategies, including materials and instructional language, which a teacher uses to provide instruction.

Q-Comp. Quality Compensation law (Q Comp) was enacted in the Minnesota Legislature in July 2005. It is a voluntary program intended to improve teacher

professional growth that allows districts and teachers to design and collectively bargain a plan that meets the four components of the law: Career Ladder/Advancement Options; Job-embedded Professional Development; Teacher Evaluation; and Performance Pay and Alternative Salary Schedule.

Summary

The 2015 reauthorization of the Elementary and Secondary Education Act made school districts responsible for selecting and maintaining a high quality teaching force in an effort to reduce the achievement gap. Classroom observation is the primary means by which to assess teacher performance. Research has been conducted to explore ways to make observations more accurate, but little research exists to explore what elements of the observation process leads to increased teacher development. This study addressed this gap by focusing on three research questions:

1. What role does the perceived content area expertise of the observer play in teacher observations for the observer and the observed?
2. What role does the perceived expertise of the observer play in teacher observations for the observer and the observed?
3. How does the perceived power relationship between observer and observed shape teacher observations?

Surveys and interviews were used as a data source to answer these questions. Subsequent chapters describe in more detail the previous research, methodology, analysis, and the role that effort has on teacher growth.

Chapter Two: Literature Review

Introduction

As schools look to increase student achievement, they must identify school-based improvements over which they have the most control, and research has identified teacher quality as the most important school-based factor in student achievement (Bill & Melinda Gates Foundation, 2010; Louis et al., 2010; Leithwood, Louis, Anderson, & Wahlstrom, 2004; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2000; Rowan, Correnti & Miller, 2002; Wright, Horn, & Sanders, 1997). Therefore, it is paramount to retain and develop high quality teachers in order to increase achievement. Schools use teacher evaluations to gather data to support this goal. Summative evaluations gather data to make decisions about staff retention and formative evaluations provide feedback to teachers for professional growth.

Furthermore, legislative actions in the United States and Minnesota increased the need for quality teacher evaluations. For example, the Elementary and Secondary Education Act was reauthorized by Congress in late 2001 and titled No Child Left Behind (NCLB). This act accelerated the education accountability movement. In addition to calling for increased student achievement, NCLB also defined "highly-qualified teachers" and called on teachers to make instructional decisions based on researched "best practices" (NCLB, 2002). The Minnesota Alternative Teacher Pay System (also known as Q-Comp) was created by legislative action in 2005, which included a provision requiring participating districts to have an objective teacher evaluation system that used multiple lines of evidence (Minnesota Office of the Legislative Auditor [MOLA], 2009). The federal Race to the Top Act of 2009 included criteria for "Great Teachers and

Leaders" that further codified the need for quality evaluations (U.S. Department of Education, 2009). During the 2011 legislative session, Minnesota became the 15th state to establish yearly teacher evaluations (Laws of Minnesota 2011).

In many cases, these legislative changes, rather than being revolutionary, merely codified changes that had already been occurring in schools. First, there was a movement to increase the frequency of evaluations, such as using "walk-throughs". Walk-throughs are short, three to ten minute observations in a classroom. Second, a movement to include evidence of effectiveness, other than observations, brought student test results, goal-achievement, professional studies, and other data sources to teacher evaluations. Finally, schools began to experiment with the relationship between observers and observed by including teachers, or other non-administrators, as observers (MOLA, 2009).

Even with all these new types of evidence, purposes, and participants, the fundamental tool of evaluation has remained classroom observations. In this research study, observations were explored by examining how perceptions of power and expertise affected the relationship between the observers and the persons observed, particularly as they relate to efficacy and change. The review of the literature that follows investigates how scholars of teacher evaluation processes have framed their research and explores how these frames address power and expertise.

Perspectives on Teacher Evaluation

Given the scope of research on teacher evaluation, it is useful to focus on three specific areas that have guided and established the research base. These three areas focus on different aspects of the evaluation process, from a broad systems perspective to more

focused examination of the roles of evaluator and evaluated. Specifically, the three areas are:

- Program Effectiveness Scholarship
- Evaluation Evidence Scholarship
- Stakeholder Scholarship

The first area focuses on the overall program effectiveness of a teacher evaluation program to determine how well that program meets its intended purposes. Sample research topics include looking at a system's objectivity (Heneman & Milanowski, 2003), effectiveness (Wise, Darling-Hammond, McLaughlin & Bernstein, 1984), place in a larger system of teaching and learning (Darling-Hammond, 2012) impact on student achievement (Kupermintz, 2003), ability to differentiate performance (Weisberg, Sexton, Mulhern & Keeling, 2009), best weighting for composite scores (Mihaly, McCaffey, Staiger & Lockwood, 2013), or underlying philosophy (Taut, Santelices, Araya, & Manzi, 2010). A second body of scholarship focuses on the usefulness of specific types of evaluation evidence, such as multiple lines of evidence (Kane, McCaffey, Miller & Staiger, 2013; Peterson, 1987), rating criteria (Epstein, 1985), or specific data collection tools and protocols (Evertson & Burry, 1989; Ho & Kane, 2013; Kane & Staiger, 2012). The final focal point examines the stakeholders in a teacher evaluation program. This might be the administrators (Ovando & Ramirez, 2007), the teachers (Ovando & Harris, 1993), students (Kane & Staiger, 2010), or the interaction between groups (Johnson & Shields, 2007).

Program Effectiveness Scholarship

Program effectiveness researchers try to determine if a teacher evaluation program is meeting the purposes for which it was designed. While they come from different traditions, these researchers all assume that a well-designed system of evaluation leads to a high quality evaluation. They also assume that interactions between system elements are more influential on validity and reliability than individual elements. These assumptions can be seen in the following examples.

Examples of program effectiveness research. In the first example, Wise, Darling-Hammond, McLaughlin, and Bernstein (1984) studied "evaluation practices with a view to analyzing how teacher evaluation can be used to improve personnel decisions and staff development (p. 3)" They surveyed 32 school districts to identify practices that were and were not effective. They also conducted interviews in these districts and case studies for four of these districts. They concluded that:

1. To succeed, a teacher evaluation system must suit the educational goals, management style, conception of learning, and community values of the school district.
2. Top-level commitment to and resources for evaluation outweigh checklists and procedures.
3. The School district should decide the main purpose of its teacher evaluation system and then match the process to the purpose.
4. To sustain resource commitments and political support, teacher evaluation must be seen to have utility. Utility depends on the efficient use of resources to achieve reliability, validity, and cost-effectiveness.

5. Teacher involvement and responsibility improve the quality of teacher evaluation (pp. 66-76)

In her more current work, Linda Darling-Hammond (2012) continues to look at teacher evaluation but in a larger context. She states that teacher evaluation be one element in "a teaching and learning system that supports continuous improvement (Darling-Hammond, 2012, pp. 1-2). She states five desired elements of this system: Common statewide standards; performance assessments, based on these standards, guiding state function; local evaluation systems aligned to the same standards; support structures; and aligned professional learning opportunities (Darling-Hammond, 2012)

Taut, Santelices, Araya, and Manzi (2010) sought to explicate the theories underlying Chile's national teacher evaluation system (NTES) as held by four stakeholder groups who were the original designers of the program: the Chilean Education Ministry, Chile's Teacher Union, Association of Local Authorities, and the Measurement Center of the Catholic University of Chile. The work by Taut et al. (2010) was the first phase in evaluating the system and their role was "to help program designers and implementers formulate their underlying program theories regarding the NTES" (Taut, et al., 2010, p. 477). These researchers analyzed policy documents and interviewed fourteen leaders from the stakeholder groups to reconstruct the intent of the program. In their reconstruction, they melded the perspective of the stakeholder groups, as well as their own, demonstrating their assumption about the importance of interaction between elements. This reconstruction was intended to inform further evaluation of the program and illustrates their validity assumption. They found that each group had a

different intent for the program which accounted for some of the difficulty in implementation.

In another example, Kupermintz (2003) examined the validity of the Tennessee Value Added Assessment System (TVAAS) which is used to evaluate teacher effects and teacher effectiveness. He looked at how the TVAAS defined effectiveness, how effectiveness was calculated, and how accurately this calculation explained student gains to evaluate the case for using the instrument as an evaluation tool. This highlights assumptions about the interaction between elements and system validity. Kupermintz used previously published data from the TVAAS to run a validity assessment of the program's strategy for determining students' prior achievement. He also applied theoretical values to the TVAAS computation algorithm to analyze the validity of assigning growth effects to the teacher instead of the student. He concluded that the TVAAS did not contain enough validity evidence to support its use in teacher evaluation.

Heneman and Milanowski (2003) evaluated a standards-based evaluation system in the Cincinnati public schools. Looking at the first two years of a district-wide implementation of the program, they hoped to determine the degree of inter-rater reliability and the teachers' reactions to the new system. To determine the degree of inter-rater reliability, they drew a sample of teachers and compared how they were evaluated by teacher and administrator evaluators. To determine teacher reactions, they used surveys and interviews to collect data. They found that "positive reactions of teachers imply an acceptance of the system and its administrative features, and a willingness to have the evaluation results used for their intended purposes, such as feedback to improve instructional practice" (Heneman & Milanowski, 2003, p. 179). In

other words, they concluded that teachers believed the system is valid and evidence is balanced. They suggested, based on the teacher reaction data they collected, that future standards-based systems start with a teacher competency model, and that leaders must decide on the specific purposes of the system, stress implementation over instrumentation, anticipate different and increased role expectations, prepare teachers and administrators thoroughly, align other human resource management systems with the evaluation system, and evaluate the system (Heneman & Milanowski, 2003).

Weisberg, Sexton, Mulhern, and Keeling (2009), working with The New Teacher Project, examined the current state of evaluation systems in the United States. The worked with 12 districts across four states. These districts supplied staff and student demographic data and data from their teacher evaluation systems. The researchers also conducted surveys and interviews with teachers and leadership from these districts. From this research, they identified the Widget Effect which "describes the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher" (Weisberg et al., 2009, p. 4). They found that the results of this effect were that all teachers were rated good or great, excellence went unrecognized, professional development was inadequate, no special attention was given to novices, and poor performance went unaddressed. They postulated that "reversing the Widget Effect depends on better information about instructional quality that can be used to inform other important decisions that dictate who teaches in our schools" (Weisberg et al., 2009, p.7).

The Measure of Effective Teaching (MET) Project, funded by the Bill & Melinda Gates Foundation, was initiated in 2009 to discover better sources of information. Mihaly, McCaffrey, Staiger, and Lockwood (2013) produced one of the MET project

research reports. They examined how different measures of teaching could be combined into a single composite score. Their goal was to understand how the measures "might be combined to improve inferences about a teacher's impact on student achievement (as measured by tests) and about teaching (as measured by observations and surveys)" (Mihaly et al., 2013, p. 7). They used the data collected by the MET project which included value-added data on state and national tests, student survey responses, and assessments of video recorded lessons. They used this data in a statistical model to make predictions about teacher effectiveness and compared those predictions to actual results. They found that all the measures captured a stable component of teaching; all identified some common dimensions related to teaching, and all captured distinct unique dimensions of teaching. They concluded that composite scores that used equal weighting are more optimal across all dimensions of teaching, while scores that are weighted in favor of a particular dimension are more optimal for identifying teachers who excel in that dimension. They recommended that states first identify what they are trying to measure before establishing composite weights. Additional reports from The Measure of Effective Teaching (MET) Project are discussed in the following two research perspectives.

Strengths of scholarship. The MET research has made important contributions to the study of teacher evaluation. First, it has clearly defined the purposes for teacher evaluations and identified their sometimes dueling nature: retention and professional growth. Second, it has identified what evidence is being used to establish the strengths and weaknesses of teachers. Finally, it has defined who the stakeholders are in the evaluation program. It is worth noting that the last two items provide foundational

information for the other two scholarship areas discussed in this chapter: evaluation evidence and stakeholder scholarship.

Gaps in this scholarship. In contrast to these strengths, research from the program effectiveness perspective has a few gaps. First, the fidelity of implementation has been generally ignored. This means the studies might incorrectly identify the underlying correlations between the system as designed and the results of the research. Secondly, program effectiveness research doesn't consider the human and contextual nature of teacher evaluation. This is particularly worrisome given the importance of observation to nearly all evaluation programs. Lastly, findings from program effectiveness are most useful to system designers at the district administrative level and not administrators or teachers who are ultimately responsible for enacting the program.

Approach to observations. Researchers from the MET group studied classroom observation differently than researchers from the evaluation evidence and stakeholder scholarships groups. They were interested in studying the whole program and not an isolated part. They were interested in the interaction between design elements and not the interaction between participants, so issues of power and expertise would not be a concern. These researchers considered and explored observations as a piece of the program, but they did not closely examine just observations, unlike the second group of scholars.

Evaluation Evidence Scholarship

A second group of research has a focus on specific evidence used in the evaluation process. While the researchers who focus on the quality of evidence seem to come from a positivist or post-positivist research paradigm, the research in evaluation

evidence is defined by other common assumptions. Expertise plays a dual role in this second body of scholarship. First, this body focuses on the usefulness of specific types of evaluation evidence to evaluate teacher's expertise in content and pedagogy. Second, these researchers generally believe that high quality evidence results in high quality evaluation, and collecting high quality evidence depends on the evidence collection expertise of the evaluators. Furthermore, they assert that validity and reliability of evidence increases the utility of that type of evidence. This latter belief is the defining element of evaluation evidence research.

Examples of evaluation evidence in early research. In a research study conducted nearly 30 years ago, Peterson (1987) examined an evaluation system that used teachers' dossiers built on multiple and variable lines of evidence. Peterson identified problems in evaluation evidence used in traditional principal-based evaluation systems and analyzed the dossier program to determine its impact on these issues. A sample of dossiers for the Nebo School district in Utah using a "lines of evidence" evaluation system was analyzed to determine if the lines were a better means of evaluating teachers. The program used "eight lines of evidence from which teachers could select: student report, parent survey, student achievement, teacher tests, peer review, administrator report, documentation of professionalism, and 'other'" (Peterson, 1987, p. 313). Peterson concluded that a multiple line evaluation provided a higher quality evaluation because it allowed triangulation of evidence and overcame the limitations of a single bit of evidence. In other words, multiple lines of evidence can be combined to increase the utility of evidence as is assumed in this body of scholarship.

The second example also found increased utility in combined evidence. Epstein (1985) analyzed data from a Maryland school district in which parents and principals provided teacher evaluations. She matched the parent evaluations to the principal evaluations to compare the ratings of individual teachers and the factors that contributed to those ratings. She concluded that parents and principals rate on different factors, and, as a result, a combined rating is more accurate than these individual ratings. Evaluating ratings to determine their reliability is an example of one of the defining assumptions of this current research study.

In the last example from this perspective, Evertson and Burry (1989) were concerned that "valuable information regarding the context of the classroom observation is lost and is not retrievable" (p. 297). As a result, they chronicled the use of the Classroom Activity Record (CAR), previously developed by Evertson. The CAR provides a structure for observations by using codes to describe typical classroom activities. Additionally, descriptive notes are simultaneously recorded. Finally, "The CAR may be implemented with a variety of observation systems including those requiring specimen descriptions, anecdotal records, critical incident recording, and on-line checklist" (Evertson & Burry, 1989, p. 298). They used the CAR in two settings. The first setting, an evaluation on the effects of class size, demonstrated that investigators using CAR were able to better understand why variations occurred in the data. The second setting, which compared administrator and senior teacher evaluations of intern teachers, demonstrated that use of the CAR reduced variability between evaluator groups. The examination of ways to increase the reliability of observations illustrates yet another defining assumption of the current research study.

Evaluation evidence in more recent research. The reliability of observations was also the focus of the Measures of Effective Teaching (MET) project. Kane and Staiger (2012) tested five different approaches to classroom observations: Framework for Teaching (FFT), Classroom Assessment Scoring System (CLASS), Protocol for Language Arts Teaching Observations (PLATO), Mathematical Quality of Instruction (MQI), and UTeach Teacher Observation Protocol (UTOP). All of these instruments are rubric-based and require training and judgment to use. They used the videotaped lessons collected by the MET project. Raters were trained, certified, and monitored in the use of one of the five instruments. Their ratings of the videos were used to determine the reliability of each instrument and the association between the instrument and a range of student outcomes: state tests, alternative tests, and student survey results. Kane and Staiger (2012) found that all the instruments were positively associated with student achievement gains and that reliably characterizing a teacher's practice requires averaging scores over multiple observations. Additionally, they found that combining observation scores, student achievement gains, and student feedback improved reliability and predictive power. Further, this combined measure is a better predictor of student achievement than teachers' educational degrees and experience. They concluded that observations would require several quality assurances, evaluation systems should include multiple measures, and the true promise of observations is the potential to improve practice. One limitation of this study was the inability to use experimental design. Instead, differences in student background were addressed using statistical methods. In a subsequent MET report by Kane, McCaffrey, Miller and Staiger (2013), this limitation was addressed.

Kane, McCaffrey, Miller, and Staiger (2013) addressed this limitation by randomly assigning teachers to classes in the 2010-11 school year. They began with the composite score from their previous study (Kane & Staiger, 2010) which weighted each measure equally. Next, principals built the master schedule without assigning teachers. Teachers were randomly assigned to those courses. A predictive model was built using student achievement gains and teacher effectiveness calculations from the 2009-10 school year to predict scores for the 2010-11 school year. Finally, actual end of year results from 2010-11 were compared to these predicted outcomes. They found that the measures of effectiveness from the previous year did identify teachers who had higher than average student achievement following random assignment. Also, the magnitude of this achievement was as expected. One caveat in this study is the difficulties in the randomization plan. Difficulties were caused by numerous factors including students transferring classes or schools, teachers getting new assignments, or principals who did not follow the randomization scheme. District compliance in all aspects of the data collection plan ranged from a high of 66% to a low of 27%. Kane et al. (2013) noted that "no information is perfect, but better information should lead to better personnel decision and better feedback to teachers" (p. 39). This philosophy summarizes not only their research, but all research from the body of evaluation evidence scholarship.

In the final example from this body of literature, Ho and Kane (2013) also conducted their research under the auspices of the MET project. They used the videotaped lessons from one Florida district, the district's observation protocol, and the district's standard training on that protocol to compare how administrators and peers scored the same lesson. Additionally, they allowed teachers to choose the lessons the

administrators watched and compared scores on those videos to non-chosen videos scored by peers. Scores were also compared between administrators in and out of the teacher's building. Finally, half of the observations were scored both after 15 minutes into the lesson, as well as at the end of the lesson.

Seven key findings resulted from Ho & Kane's analysis:

1. Observers rarely used the top or bottom categories on the four-point observation rubric.
2. Compared to peer raters, administrators differentiated more among teachers with a 50% larger standard deviation in teacher scores.
3. Administrators rated their own teachers .1 point higher than administrators from other schools and .2 higher than peers.
4. Although administrators scored their own teachers higher, their rankings were similar to the rankings produced by others outside their school.
5. Allowing teacher to choose their own videos generated higher average scores. However, the relative ranking of teachers was preserved whether videos were chosen or not.
6. When an observer formed a positive (or negative) impression of a teacher in the first several videos that impression tended to linger across all videos for that teacher.
7. There are a number of different ways to ensure reliability of .65 or above. Having more than one observer really does matter.

(Ho & Kane, 2013, p. 4)

Strengths of scholarship on evaluation evidence. The defining assumption of the validity and reliability of evidence increases the utility of that type of evidence in teacher evaluation systems. Furthermore, the research has evaluated individual pieces of evidence and clarified the range of reliability and validity found in various teacher evaluation programs. As such, it informs program effectiveness scholarship. Finally, it illuminates and evaluates different ways of gathering evidence by stakeholders and informs stakeholder scholarship.

Gaps in scholarship on evaluation evidence. This body of scholarship also has some gaps. First, it deemphasizes the context of the teacher evaluation programs. Students, teachers, classrooms, schools, and district vary considerably from location to location and evidence that is useful in one location might not be useful in another. For example, an affluent district might be able to define student achievement by passing rates on standardized tests, whereas, a more distressed district would find student growth a better measure. Even evidence scholarship that tries to address context, such as in the work of Evertson and Burry (1989) and the use of CAR, still relies on a moment in time to define the context. A second weakness is that interpretation of the evidence depends on human perception and this appears to have not been considered in the analysis. For example, Ho and Kane (2013) found that principals scored their own teachers higher, but did not explore the reason. The two situations Evertson and Burry (1989) examined spent considerable time training staff to use the CAR. The program Epstein (1985) examined did not train parent or principal raters at all. This calibration of the CAR explains why it produced consistent results, while the findings of the parents and principals in Epstein's work were inconsistent. Finally, like program effectiveness

scholarship, this body of scholarship is most useful during the design phase of a teacher evaluation program.

Approach to observations. Unlike researchers from the program effectiveness perspective, researchers from this group of scholars would examine observation closely. Their focus would be on examining the accuracy of the observation to see if what is observed is reported reliably, particularly between raters, as demonstrated in Ho and Kane (2013) and Epstein (1985). These researchers would focus on expertise. They would be concerned about raters having expertise using the rating instruments. Researchers from evaluation evidence scholarship would, and do, investigate the reliability of ratings from administrators and peers, but would not be concerned about the power relationship in observations. This is the approach taken by Evertson and Burry (1989) and Ho and Kane (2013). Researchers in this group would take a strictly objective approach, unlike scholars in the final body of scholarship reported below.

Stakeholder Scholarship

The third area of scholarship focuses on the people involved in the evaluation process. This group of scholars focus their research on the stakeholders, both evaluators and evaluated, involved in the evaluation process. Considering the summative and formative purposes of evaluation, it is clear that evaluations are a tool through which evaluators attempt to influence the behavior of the evaluated. As such, the interactions between the stakeholders involve uses of power (French and Raven, 1959). However, sources of power vary between stakeholder groups. French and Raven (1959) noted that power is based in both positional and personal sources. Administrators who conduct evaluations inherently have positional power and could also have personal power. Peer

evaluators might have some positional power, particularly if they are involved in pay for performance evaluations, but personal power, based on their knowledge and experience, is the primary source of influence for peer evaluators. These researchers assume that human perceptions and context affect evaluation. Additionally, they assume that understanding these perceptions and contexts lead to a better teacher evaluation system. These assumptions can be seen in the following examples beginning with Ovando and Ramirez (2007).

Examples of research on stakeholders. Ovando and Ramirez (2007) conducted a study to "identify principals' perceptions regarding their instructional leadership actions within the context of the performance appraisal system for teachers in successful schools" (p. 93). They were concerned by the lack of research in teacher evaluation that reflected a principal's voice and felt adding this voice would enhance the discussion. Through their principal interviews they found three common instructional leadership actions: setting clear expectations to clarify process and activities, monitoring instruction through walk-through observations, and connecting teacher's performance evaluation data to professional development. Finally, they concluded that "school leader preparation programs should aim at the development of instructional leadership competencies and dispositions" (p. 108). The research design and conclusions of the work of Ovando & Ramirez illustrates the belief that principal perceptions affect teacher evaluation and that understanding these perceptions is important as is typical of this research focus. Other research adds the teacher voice.

Ovando and Harris (1993) attempted to clarify teachers' perceptions regarding the post-observation conference. They believed that evaluations that were more

collaborative in nature had a better chance of improving teaching and learning. They established ten characteristics of a collaborative process: mutual respect, tolerance, acceptance, commitment, courage, sharing, adhering, respecting, differentiation, and teaming (Ovando & Harris, 1993, p. 302). They believed that understanding how teachers perceived the post-observation conference could lead to a more collaborative process. Surveys were mailed to a sample of teachers in mideast Texas. Based on survey responses, Ovando and Harris found that teachers thought the conference should be a tool to discuss teaching and learning, follow an orderly sequence, occur in their classroom or other familiar environment, and be completed soon after the observation. The design of this research demonstrates their belief in the value of understanding perceptions to create a better evaluation system.

While the first two examples examined teachers and principals separately, the work of Johnson and Shields (2007) looked at the interaction between administrators and teachers. In their study, they examined the Teacher Efficiency Agreement (TEA) between the New South Wales Department of Education and Training (DET) and the New South Wales Teachers' Federation (NSWTF). The TEA is the annual performance appraisal system for New South Wales. They were interested in this particular agreement because it "represented a small but significant departure from the adversarialism that had previously characterized employment relations" (p. 1214) and they wanted to understand why. Interviews were conducted to explore this issue. Interviewees were drawn from random, convenience, and purposeful samples. They concluded that a salary dispute and staffing crisis that had preceded the TEA agreement actually established a condition in which trust had been built up at the building level between building administrators and

teachers as these groups worked together to overcome these difficulties. Additionally, the TEA agreement allowed the union to establish a new purpose and the department to claim progress towards improved teaching and learning. Their conclusions illustrate the study of context as a way to investigate teacher evaluation systems.

The final example from the stakeholder body of scholarship comes from the Measures of Effective Teaching (MET) project. The initial research report discussed the five measures used by the MET project: Student achievement gains on differentiated assessments, classroom observations and teacher reflections, teacher's pedagogical content knowledge, student perceptions of the classroom instructional environment, and teacher's perception of working conditions and instructional support at their schools (Bill & Melinda Gates Foundation, 2010). Two of these measures, student achievement gains and classroom observations, were analyzed in detail in separate reports as noted above in the evaluation evidence body of scholarship. However, two other of these measures, teacher's pedagogical content knowledge and teacher's perceptions of working conditions were not subjected to detailed analysis in the MET project. One of these measures, student perceptions of the classroom instructional environment, was analyzed in the initial report of the Bill & Melinda Gates Foundation (2010). They used the survey of the Tripod Project for School Improvement. The Tripod surveys were designed for use with specific age ranges and have observational rather than judgmental items for students to answer. The Tripod survey uses multiple survey items to gauge seven constructs: Care, Control, Clarify, Challenge, Captivate, Confer, and Consolidate. Results from this survey indicated that "student perceptions of a given teacher's strength and weaknesses are consistent across the different groups of students they teach" and

"student perceptions in one class are related to the achievement gains in other classes taught by the same teacher" (Kane & Staiger, 2010, p. 9). Since student input is seldom considered in primary and secondary school teacher evaluations, these findings could offer a new measurement for many evaluation systems. One notable example is the use of longitudinal student engagement data in Minnesota (Education Code, 2016). The Minnesota Department of Education (2013) created a model program for this statute which includes the use of a student survey to meet this requirement.

Gaps in scholarship on stakeholders. The reliance on the unique aspects of each school's cultural context is one of the main gaps in stakeholder scholarship. The conclusions that result from this research have limited generalizability. Recognition of this is evident in three studies: Ovando and Ramirez (2007) noted "it is relevant to acknowledge that this study was limited to three purposefully selected schools" (p. 108). Ovando and Harris (1993) noted "the results indicated, for at least one school district" (p. 309), and Johnson and Shields (2007) noted "this development can only be understood against the backdrop. . ." (p. 1225).

A second weakness is that the research studies accept or are not investigating the structure of the evaluation program and do not consider how that structure impacts that which they are studying. For example, Ovando and Harris (1993) look at the post-observation conference between a principal and teacher, but they do not examine post-observation conferences between a teacher and another teacher who is a peer evaluator. As such, they are not accounting for the positional power differential between observers and observed.

Strengths of stakeholder scholarship. In contrast to these gaps, these studies also have several strengths. First, this body of scholarship recognizes the impact that human perception has on evaluation. All evaluation evidence is filtered through human perception when it is collected and/or analyzed. For example, not acknowledging this is like reporting an "average" and not specifying if it is the mean, median, or mode. The recognition of this fact in this body of scholarship adds a certain perspective that is missing in the other bodies. Second, stakeholder research recognizes the interaction and relationship between stakeholders. In the Johnson and Shields (2007) article, the importance of relationships is clear as they concluded that the relationships between building administration and staff was instrumental in transforming the evaluation system. Ovando and Harris (1993) demonstrated the important role that collaborative relationships played in a successful post-observation conference. These two examples also highlight the final strength of this body of scholarship: it has significant utility for building principals and teachers who actually implement the evaluation program. This is because the principals and teachers can modify their individual practices related to observations and evaluations without the need to redesign the system across the district. The evaluation program designs and evaluation evidence are important, but their effectiveness is dependent on the actual fidelity of implementation. The literature on the role of stakeholders in the evaluation process is limited and appears to be the only area that addresses evaluation at the implementation level.

Researchers on stakeholders look at classroom observations as more as relational events. Observations, to them, seem to be shared experiences that need to be understood from multiple perspectives and in context. They approach the study of observation by

asking participants how they experienced the observation process. Several of the studies (Ovando & Harris, 1993; Ovando & Ramirez, 2007) already consider perception. It would have been a logical extension to focus on specific perceptions of power and expertise, but that was not done in their research.

Additional Factors in Evaluation

Three areas of scholarship—program effectiveness, evaluation evidence, and stakeholder research—form the backbone of teacher evaluation research. However, it is necessary to consider other factors when thinking about the impact of the teacher evaluation process on student achievement: school culture, teacher experience, expertise, and power.

Research into school change shows that some schools have a culture of change that makes it more likely teachers will change practice. These cultures have several labels, such as an "adaptive school" (Garmston & Wellman, 1999) or a "reflective school" (York-Barr, Sommers, Ghere, & Montie, 2006). An adaptive school has a clear identity and is not tied to a particular form. The school asks: 1) Who are we? 2) Why are we doing this? and, 3) Why are we doing this, this way? (Garmston & Wellman, 1999). Reflective schools engage in a theory of action for reflective practice: pause, openness, inquiry, thinking, learning, action, and enhanced student learning (York-Barr et al., 2006). Schools with a professional community have shared values, focus on student learning, collaboration, deprivatized practice, and reflective dialogue (Kruse, Louis, & Bryk, 1994). Teachers in these cultures are supported in their improvement and professional growth because the school culture is focused on continuous improvement.

Teacher experience also contributes to a teacher's willingness to change practice. Steffy, Wolfe, Pasch, and Enz (2000) noted that teachers move through six phases in their career: novice, apprentice, professional, expert, distinguished, and emeritus (pp. 6-10). Teachers in the professional phase "most frequently seek help and assistance from other teachers. They actively participate in collegial network for support and guidance. They begin to look beyond the classroom, seeing themselves and their colleagues as part of a broader profession" (Steffy, et.al, 2000, p. 8). Therefore, teachers in this phase or beyond are more receptive to outside ideas than teachers in the first two phases. Novice teachers gain confidence in the field through their practicums and apprentice teachers take responsibility for planning and instruction. Targeted feedback to help them develop in these areas helps them make the transition to professional teachers (Steffy, et. al., 2000, pp. 6-8). Teachers at the expert phase change themselves as they are "typically self-motivated to improve their teaching" and "pursue reflection in a collaborative manner" (Steffy, et. al., 200, pp. 79-80). Therefore, teachers at various phases of their careers view evaluations and feedback differently as they move from wanting feedback on specific strategies to observations to fuel their own self-reflections.

Berliner (2004) identified a long list of qualities of expert teachers:

> Expert teachers often develop automaticity and routinization for the repetitive operations that are needed to accomplish their goals; expert teachers are more sensitive to the task demands and social situation when solving pedagogical problems; expert teachers are more opportunistic and flexible in their teaching than are novices; expert teachers represent problems in qualitatively different ways than do novices; expert teachers have fast and accurate pattern-recognition

> capabilities, whereas novices cannot always make sense of what they experience; expert teachers perceive meaningful patterns in the domain in which they are experienced; and although expert teachers may begin to solve problems slower, they bring richer and more personal sources of information to bear on the problem they are trying to solve (p. 201).

Michelson (2001) identified three attributes associated with personal power: knowledge/information, personal attraction, and effort. The first attribute, knowledge/information, was described as "expertise acquired by possession of special knowledge or information" (p. 195). Additionally, he noted that "a leader's high level of effort can be parlayed into increased expertise" (Michelson, 2001, p. 195).

In addition to identifying characteristics of personal power, Michelson (2001) also identified five factors that contribute to positional power: centrality, criticality, flexibility, visibility, and relevance. Centrality and criticality are described as being located near the work flow and having a good communication network. Flexibility as it relates to power is having the ability to make adjustments in routines and processes. Visibility is how a person in power makes his or her presence noticeable in the organization. Finally, Relevance is how a leader works in connecting various pieces of an organization. An example of relevance is when leaders connect people to the larger organizational goals or by developing the skills needed by the organization.

Summary

Teacher evaluation and high quality feedback are critical elements in improving instructional behavior. High quality instruction is known to lead to improved student learning (Bill & Melinda Gates Foundation, 2010; Louis et al., 2010; Leithwood, Louis,

Anderson, & Wahlstrom, 2004; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2000; Rowan, Correnti & Miller, 2002; Wright, Horn, & Sanders, 1997). Furthermore, teacher evaluation processes need continued research because with local and legislative changes that are occurring, we need to better understand what works well in an effective evaluation system.

In this review of literature, scholarship in teacher evaluation has been organized from three perspectives: overall program effectiveness, evaluation evidence, and stakeholder perspective. Each of these bodies of scholarship has its own strengths, gaps, and approaches to studying observations.

Gaps in Research

Because there likely are an infinite number of factors that could impact the construction of a teacher observation system, it is necessary to narrow the focus. Three factors that emerged in the research literature presented here that clearly affect evaluations are school climate, observer expertise, and observer power.

The research indicates that there is a need to measure teacher expertise during an observation, and there is a need for observers to have expertise themselves in using the observation instrument. These two elements of expertise appear in the literature and have been studied to determine their reliability for measuring a teacher's strengths and weaknesses.

An unexplored area in the teacher evaluation experience comes from the power dynamic inherent in the observation process. The higher scores associated with teacher selected videos in the MET project (Ho & Kane, 2013) indicate that teachers try to influence principals by demonstrating their best lessons when possible. The MET project

also demonstrated that impressions tend to linger over time, so this influence is cumulative. This effect might explain why principals rated their own teachers higher (Ho & Kane, 2013). Principals also influence teachers in the evaluation process through the identification of growth areas. The identification of areas for further professional growth could influence teacher development, since the principal has positional power. Peer evaluators do not have the same positional power and must draw on personal power to influence growth. The literature does not appear to address how power issues impact teacher perceptions of the validity of the evaluation experience.

Finally, the literature on school climate and teacher experience suggests that factors outside the evaluation process itself can have an impact on the evaluation. Teachers in a culture that facilitates change can try new strategies with the understanding that their overall performance evaluation will not be affected based on their initial implementation attempts. Working in such a culture could increase their willingness to take a risk, based on their evaluation. In addition, teachers in the early phases of their career may place more emphasis on the suggestions of their supervisor because they have not yet connected themselves to the larger profession. Conversely, teachers in the advanced phases of their career may value the improvements that are possible for them and the larger profession over their immediate supervisor.

Research Questions

Given the lack of investigation into the links between perceived expertise of the observer and the observed, plus not having a full understanding of the role of power relative to expertise in teacher evaluations, three research questions emerged:

1. What role does the perceived content area expertise of the observer play in teacher observations for the observer and the observed?
2. What role does the perceived expertise of the observer play in teacher observations for the observer and the observed?
3. How does the perceived power relationship between observer and observed shape teacher observations?

Chapter Three: Methodology

Instrumentation

This study used a combination of surveys and interviews to answer the following research questions:

1. What role does the perceived content area expertise of the observer play in teacher observations for the observer and the observed?
2. What role does the perceived expertise of the observer play in teacher observations for the observer and the observed?
3. How does the perceived power relationship between observer and observed shape teacher observations?

Answering these research questions involved a clearer understanding of the role that participant perceptions play in data collection. Nelson (2008), in his text that explored perception in asking questions, noted:

> Perception is the subjective process of acquiring, interpreting, and organizing sensory information. Survey questions that assess perception, as opposed to those assessing factual knowledge, are aimed at identifying the processes that (a) underlie how individuals acquire, interpret, organize, and, generally make sense of (i.e. form beliefs about) the environment in which they live; and (b) help measure the extent to which such perceptions affect individual behaviors and attitudes as a function of an individual's past experiences, biological makeup, expectations, goals, and/or culture (p. 580).

Therefore, survey questions that assess perception give insight into the role that perception plays in the observation process, in term of both sense-making and behaviors.

Nelson's insights were also used to provide the basis for how the interview questions were formulated. Leonard (2003) noted that interviews "may be used as follow-up to a questionnaire. This allows the researcher to explore in more depth interesting issues that may have emerged from the standard questionnaire" (p. 3). Interviews were used in this manner to give more context to the survey findings.

Survey design

Research also indicated how the constructs of expertise and power could be framed in the survey and interview. Berliner (2004) identified a long list of qualities of expert teachers as listed in Chapter 2. However, these qualities as described by Berliner are cognitive processes and are difficult for the teacher to notice throughout the observation process, thus it was useful to have a more easily observable set of characteristics for the construct of expertise. The set of observable characteristics used to develop the survey for this study were created by blending Berliner's definition and concept of expertise (Berliner, 2004) with Michelson's definition of personal power (Michelson, 2001) as described in Chapter 2. Two attributes of personal power as described by Michelson (2001), knowledge and effort, encapsulate expertise as described Berliner (2004) and were more likely to be observed by teachers during the observation cycle. Therefore, knowledge and effort were used as the basis for teacher perceptions of expertise in the evaluation relationship.

The second observer dynamic under examination in this study was observer power. In addition to identifying characteristics of personal power, Michelson (2001) also identified five factors that contribute to positional power: centrality, criticality,

flexibility, visibility, and relevance. These five criteria were used in the surveys and interviews developed for this study to measure teacher perceptions of observer power.

In general, evaluators can be perceived to have either high or low power. They can also have high or low expertise. Using these dimensions, there are four possible combinations of power and expertise that are possible using this model, as summarized in Figure 1. Observers were classified into these categories based on survey results.

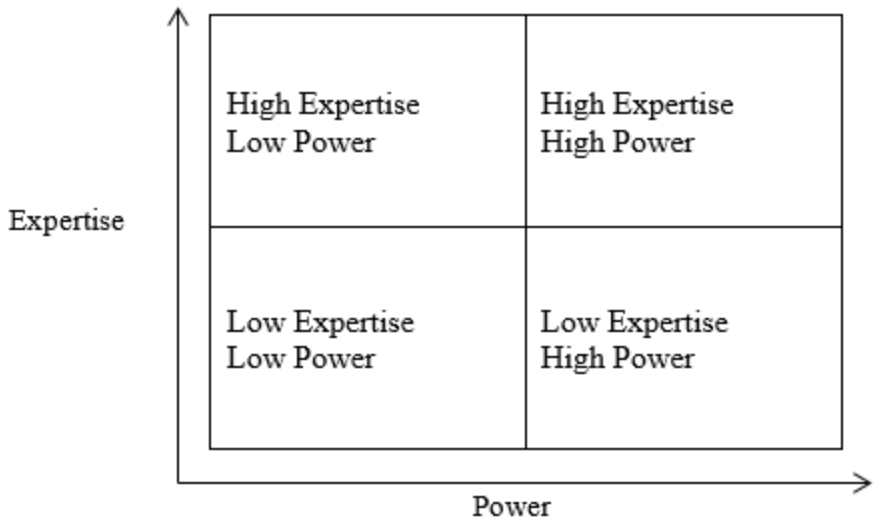


Figure 1- Power and Expertise Matrix

Survey questions

Survey questions were organized around four central ideas: organizational culture, observer power, observer expertise, and evaluation outcomes. The first three of these ideas are familiar concepts, but it is necessary to clarify what evaluation outcomes were in this context. Evaluation outcomes were the self-reported actions of the teacher based on the observation cycle. Specifically, these questions were: Did the teacher change his

or her behavior based on feedback from the observations? Did the teacher experience personal growth? and, Did student learning increase? Questions on the survey had a four point response scale of Strongly Agree, Agree, Disagree, Strongly Disagree, except where noted. Questions are presented below based on the aforementioned central ideas, but the central ideas as thematic headings noted below were masked in the final survey.

Organizational Culture:

- Staff understands school goals
- I have an opportunity to develop my own goals
- Staff challenge existing beliefs and practices
- Staff have a common vision
- My team shares successes and failures
- My team plans together
- Staff considers the impacts of each change

Power:

My last observer:

Centrality

- Has time for me
- Is located in a convenient location

Criticality

- Has influence in building staffing decisions
- Has influence in building scheduling decisions
- Has influence in building capital decisions
- Has influence in building goals

Flexibility

- Has a range of responsibilities
- Leads building initiatives
- Is allowed to adjust plans as necessary

Visibility

- Serves on several committees
- Interacts with staff (use frequency scale: 4-5 times/week, 2-3 times/week, 1 time/week, less than 1/week)

- Supervises common areas (use frequency scale: 4-5 times/week, 2-3 times/week, 1 time/week, less than 1/week)

Relevance

- Has influence in district decisions
- Monitors progress on building goals
- Fosters professional growth

Expertise:

My last observer:

Knowledge/information

- Was knowledgeable in my content area
- Was knowledgeable about general pedagogy
- Was knowledgeable about classroom management
- Was knowledgeable about the evaluation rubric
- Provided new resources

Effort

- Understood my lesson objectives
- Wrote detailed feedback
- Connected feedback to details from observed lesson
- Connected feedback to my personal goals
- Provided adequate meeting time to discuss feedback
- Was available outside of scheduled observations and conferences

Outcomes:

- My last evaluation helped me grow
- I implemented suggestions from my last evaluation
- Strategies I implemented were useful
- My last evaluation improved student learning

Demographics

- Total years teaching including this year
- Years teaching in district including this year
- Content areas taught during last observation cycle
- Grade level(s) taught during last observation cycle
- Gender

- Gender of observer during last observation cycle

(Note: a copy of the survey instrument can be found in Appendix A.)

Interview design

Interviews were conducted using the aforementioned categories of organizational culture, power, expertise, and outcomes. Interviews were semi-structured to allow exploration of existing categories and exploration of new themes that emerged from the survey and the interview itself. Additionally, interviews used open-ended questions so themes could emerge from participants. Thirty minutes were scheduled for each interview.

Interview questions

While some interview questions emerged as a result of the preliminary analysis of survey results, the final list of questions below formed the basis of the structure for the interview:

- Describe the students in your school.
- Describe how staff interacts in your school.
- What does your school value?
- Describe the process of your last observation cycle.
- What is your relationship with your observer?
- Describe your observer's knowledge on the evaluation rubric.
- Describe your observer's knowledge on the evaluation classroom management.
- Describe your observer's knowledge on the evaluation pedagogy.
- Describe your observer's knowledge on the evaluation (your content).
- How did this observation cycle impact your teaching?

(Note: a copy of the interview protocol can be found in Appendix A.)

Participants

Participants were teachers in two high schools in a large suburban school district. This district was chosen based on convenience and several criteria. Travel distance was a

convenience factor in selecting which districts were approached to participate. Another criterion for selection was whether or not a district participated in Minnesota's Alternative Teacher Pay System (Q-comp). Districts that participate in the Q-Comp program use peer evaluators for teachers that are not on undergoing their formal summative evaluation, or high cycle, as defined by Minnesota Teacher Development and Evaluation law (MN statute 122A.40). Therefore, teachers in these districts are more likely to have evaluators with the same content background. In particular, the district in this study is large enough such that only a few teachers do not have a peer evaluator in the same licensed area. Having this content alignment increased the likelihood of observers having perceived content expertise. Additionally, districts without Q-Comp use an administrative evaluation model, so teachers in these districts are more likely to experience high power differentials. The final criterion was that the district had to have at least two secondary schools at the same secondary level, middle or high.

A focus on secondary schools was selected because it narrowed and focused the survey. Elementary schools have fewer licensure areas than secondary schools. Secondary schools, therefore, are more likely to have observers from outside a teacher's licensure area. As a result, secondary observations will have a wider range of observer expertise in the high and low expertise categories. The preference for approaching partner districts was if these two schools also had demographic or programming differences from each other.

The superintendent and two principals from two high schools in the district selected for this study agreed to allow their teachers to be part of this research project. These schools have a number of significant differences that made them distinct. First, the

schools are significantly different in the relative size of their student populations, with one school having approximately 2,000 students and the other having roughly 50% more students and a proportionately larger staff. Second, the schools have different demographic populations, with one school having four times the percentage of students of color. Finally, each school has a different specialized curriculum focus.

Sampling

The number of surveys administered was a convenience sample of each building's staff. The sample was comprised of teachers who were present at an after school staff meeting at each school on a given day. The survey was administered following the meeting and participation was voluntary. Staff members that were not present at the staff meeting did not have the opportunity to take the survey at another time. Overall, 65% of the two buildings' combined teaching staffs completed the survey.

Interview participants were selected from a random sampling of the teaching staff regardless of participation in the survey. A staff list was scrambled and number and a random number generator was used to identify participants. Four participants in each building were identified and asked to participate in an interview. Additional participants were selected as needed until four interviews were conducted in each building, creating a total of eight interviews.

Data Analysis

Regression

The survey had four central ideas: organizational culture, observer power, observer expertise, and evaluation outcomes. For each of the four component areas, responses were assigned a value. Strongly disagree was assigned the value of 1, and

disagree, agree, and strongly agree were assigned 2, 3, and 4 respectively. Again, in this context, evaluation outcomes were determined by whether or not the teacher perceived that he or she changed his or her practice, experienced professional growth, or noticed an increase in student learning based on the observation cycle. Teachers were specifically asked these three questions on the survey and these three components were dependent variables in the final analysis. As noted in the previous section on Survey Questions, each question related to a specific element of Michelson's power and expertise framework (2010), or organizational culture, or observation outcomes. Variables for power, expertise, and culture were created by summing the individual survey items to create the independent variables.

Linear regression using SPSS was used to analyze the relationship between each independent variable and each dependent variable. Model summaries and ANOVA tables as generated by SPSS for each linear regression are included in Chapter 4 for each of these pairings. The strength of the relationship was based on the Adjusted R-squared in the model summary. Significance of the regression model was based on the alpha value in the ANOVA table and an alpha value of .05 was used to determine significance.

Interviews

Interviews were recorded and transcribed. Transcriptions were first coded using the three central ideas of organizational culture, observer power, and observer expertise. Coded quotations were then categorized as positive, negative, or neutral based on the context and delivery of the quotation. The number of comments in each category was than calculated as a percentage of the total number of comments.

Data Aggregation

Finally, findings from the surveys were further analyzed using the teachers' perspectives from the detailed information gained from the interviews. This allowed the most complete understanding of the inter-relationships among the three research questions.

Summary

This study used a mixed-methods approach in data collection. A survey was created to measure teacher perceptions of observer expertise, observer power, school culture, teacher growth, teacher implementation, and student learning. The survey was administered in two high schools in a suburban school district. Survey data was analyzed using regression analysis with perceived observer expertise, observer power, and school culture as independent variables and perceived teacher growth, teacher implementation, and student learning as dependent variables. Interviews were conducted to explore school context and themes found in the regression analysis. Interviews were coded, based on the independent variables, and quantified.

Chapter Four: Findings

Introduction

This research study was focused on three research questions as listed in previous chapters. During the course of the data analysis, notably during the interviews, findings indicated the need to modify the original questions as new discreet themes emerged. The modified questions are presented at the end of this chapter.

The concepts of expertise and power used in this study were based on the frameworks and definitions of Berliner (2004) and Michelson (2001). Berliner (2004) studied expertise in teachers and identified an extensive list of habits and characteristics of expert teachers. Reading this list shows that expert teachers have acquired specialized knowledge and skills and have worked to seamlessly incorporate these habits into practice. Based on the work of Michelson, expertise is comprised of two separate components of personal power: knowledge and effort. Michelson's definition aligned with Berliner's definition to build survey items to measure teacher perceptions of expertise in terms of both knowledge and effort. Michelson also identified five components of positional power: centrality, criticality, flexibility, visibility, and relevance. Centrality and criticality are described as being located near the work flow and having a good communication network. Flexibility, as it relates to power, is having the ability to make adjustments in routines and processes. Visibility is how a person in power makes his or her presence noticeable in the organization. Finally, relevance is how a leader works in connecting various pieces of an organization. Individual survey items were combined into rating scales based on knowledge, effort, centrality, criticality, flexibility, visibility, and relevance.

The survey also contained a rating scale for school culture. The scale for school culture was designed to determine if a school had qualities to make it more easily adaptable to change, since teachers in these settings are more likely to view feedback and change in a positive manner (Garmston & Wellman, 1999; York-Barr, Sommers, Ghere, & Montie, 2006: Kruse, Louis, & Bryk, 1994).

Surveys were administered at two schools at the end of staff meetings. One hundred thirty total surveys were completed at these sessions. Interviews were conducted with four staff members from each school approximately one month after the survey was administered.

The questions in the survey focused on four central ideas: organizational culture, observer power, observer expertise, and evaluation outcomes. As previously mentioned, two of these central ideas are further divided into subcategories. Observer expertise had two components: knowledge and effort. As described in the following section on Interview Data, these two components emerged as discrete components. Evaluation outcomes also had subcomponents. In this context, evaluation outcomes were determined by whether or not the teacher who was observed perceived that she or he had implemented ideas from the evaluation, experienced professional growth, or noticed an increase in student achievement. Using linear regression model summaries and ANOVA tables generated by SPSS, the components of professional growth, implementation of new ideas, and student learning were the dependent variables, and the culture, power, and expertise components were the independent variables. An alpha value of .05 was used to determine significance.

For each of the four component areas, responses were assigned a value. Strongly disagree was assigned the value of 1, and disagree, agree, and strongly agree were assigned 2, 3, and 4 respectively. Responses in each component area were summed into a composite variable for that area and these composite scores were examined in a regression analysis to determine the association between the independent variables, power, expertise, and culture and the dependent variables, growth, implementation, and student learning, as self-reported by the respondent.

Interviews were originally coded based on the independent variables of culture, observer power, and observer expertise. During coding, the concept of expertise as described by Berliner (2004) and Michelson (2001) differentiated into discrete aspects of knowledge and effort. A full discussion of these emerging themes and details about the coding process is provided in the next section concerning the interview data. The qualitative data was used to provide context in examining the relationships among observer effort, observer knowledge, observer power, school culture and teacher perceptions of growth, implementation, and student learning

Survey Data Analysis

Teacher Growth Regression

The first area of analysis used teacher growth as the dependent variable. As aforementioned, regression was done with four separate independent variables; knowledge, effort, power, and culture, to generate a model summary for R and Adjusted R-squared values and ANOVA to generate a p value. These p values were compared to an alpha of .05 to determine significance.

Observer knowledge. The first independent variable was observer knowledge. Linear regression with the knowledge variable yields the following model summary:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.653[a]	.426	.421	.480

a. Predictors: (Constant), Knowledge

Table 1 – Model summary. Teacher growth/observer knowledge.

The R value of .653 indicates a moderate positive relationship between observer knowledge and teacher professional growth. The adjusted R-squared indicates that 42.1% of the variability in teacher growth can be predicted by observer knowledge.

In addition to the data from the model summary, the analysis of variance produced the following results:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.369	1	21.369	92.782	.000[b]
	Residual	28.789	125	.230		
	Total	50.157	126			

a. Dependent Variable: helped_grow
b. Predictors: (Constant), Knowledge

Table 2 – ANOVA. Teacher growth/observer knowledge.

From the Sig. column, $p < .001$, which is less than .05 and the result is significant and observer knowledge can be used to predict teacher growth.

Observer effort. The second area of regression analysis used observer effort as the independent variable. The R value in the model summary below indicates a moderate positive relationship between this independent variable and teacher growth.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.664[a]	.441	.436	.471

a. Predictors: (Constant), Effort

Table 3 – Model summary. Teacher growth/observer effort.

Additionally, the adjusted R-squared value indicates that 43.6 % of the variance in teacher growth can be predicted by observer effort.

ANOVA results, as listed below, indicate that observer effort can be used to predict teacher growth since $p < .001$ and the result is significant given the previous stated alpha value.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.504	1	21.504	96.897	.000[b]
	Residual	27.296	123	.222		
	Total	48.800	124			

a. Dependent Variable: helped_grow

b. Predictors: (Constant), Effort

Table 4 – ANOVA. Teacher growth/observer effort.

Observer power. The next independent variable used in the regression analysis was power. This analysis yielded the following model summary:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.659[a]	.435	.430	.479

a. Predictors: (Constant), Power

Table 5 – Model summary. Teacher growth/observer power.

As with the previous independent variables, the regression indicates a moderate positive relationship between observer power and teacher growth with 43% of the variance in teacher growth predicated by observer power.

Similar results were found in the ANOVA as can be seen in the following table:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.382	1	22.382	97.656	.000[b]
	Residual	29.107	127	.229		
	Total	51.488	128			

a. Dependent Variable: helped_grow

b. Predictors: (Constant), Power

Table 6 – ANOVA. Teacher growth/observer power.

Again, $p < .001$ and the result is significant given the alpha value of .05.

School culture. The final independent variable used in the regression analysis of teacher growth is culture.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.310[a]	.096	.089	.601

a. Predictors: (Constant), Culture

Table 7 – Model summary. Teacher growth/school culture.

Unlike the other independent variables, the R value in this model indicates there is a small positive relationship between this component and teacher growth. Furthermore, with an adjusted R-squared value of .089, only 8.9% of the variance in teacher growth can be predicted by the culture variable. Both of these values are in contrast with the previous three independent variables.

The ANOVA does show a p value of .001 which is still less than the alpha value of .05. However, it is notable that this is the only value not rounded to 0.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.625	1	4.625	12.784	.001[b]
	Residual	43.416	120	.362		
	Total	48.041	121			

a. Dependent Variable: helped_grow

b. Predictors: (Constant), Culture

Table 8 – ANOVA. Teacher growth/school culture.

Teacher Implementation Regression

The next set of regressions used the Teacher Implementation outcome as the dependent variable and again used the same four independent variables of observer knowledge, observer effort, observer power, and school culture. Overall, results are similar to the previous analysis.

Observer knowledge. When looking at the relationship between observer knowledge and teacher implementation, the model summary showed an r value of .623 showing a moderate positive relationship between the independent and dependent variable. The adjusted R-square indicates that 38.4% of the variability in teacher implementation can be predicated by observer knowledge.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.623[a]	.389	.384	.526

a. Predictors: (Constant), Knowledge

Table 9 – Model summary. Teacher implementation/observer knowledge.

The ANOVA results listed below showed p value less than .001 which is significant compared to an alpha value of .05

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.817	1	21.817	78.832	.000[b]
	Residual	34.318	124	.277		
	Total	56.135	125			

a. Dependent Variable: implemented
b. Predictors: (Constant), Knowledge

Table 10 – ANOVA. Teacher implementation/observer knowledge.

These findings are similar to the affect that the independent variable of observer knowledge had on the dependent variable of teacher growth.

Observer effort. The second independent variable considered in regression on the dependent variable of teacher implementation is again the variable of observer effort. The following table shows the model summary for this regression:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.619[a]	.383	.378	.520

a. Predictors: (Constant), Effort

Table 11 – Model summary. Teacher implementation/observer effort.

The R value listed above, .619, shows a moderate positive relationship between observer effort and teacher implementation. Additionally, 37.8% of the variability in teacher implementation can be predicted by observer effort.

The regression between teacher implementation and observer effort also produced an ANOVA table. The table below shows that $p < .001$, significant because it is less than .05, the alpha level.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.489	1	20.489	75.854	.000[b]
	Residual	32.954	122	.270		
	Total	53.444	123			

a. Dependent Variable: implemented
b. Predictors: (Constant), Effort

Table 12 – ANOVA. Teacher implementation/observer effort.

Observer power. The third independent variable used in regression was observer power. It also showed a moderate positive relationship with the dependent variable, teacher implementation. The calculated adjusted-R square indicates that 37.1% of the variability in this dependent variable can be predicted by observer power.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.613[a]	.376	.371	.524

a. Predictors: (Constant), Power

Table 13 – Model Summary. Teacher implementation/observer power.

The table below provided the ANOVA data for the analysis of observer power and teacher implementation. Based on $p < .001$, observer power is a significant variable and is a predictor of teacher implementation.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.834	1	20.834	75.795	.000[b]
	Residual	34.635	126	.275		
	Total	55.469	127			

a. Dependent Variable: implemented
b. Predictors: (Constant), Power

Table 14 – ANOVA. Teacher implementation/observer power.

School culture. The final independent variable is school culture. The R value from the model summary below indicates a small positive relationship between this independent variable and a dependent teacher implementation variable. Additionally, the

summary indicates that 9.9 % of the variability in teacher implementation can be predicated by school culture.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.326[a]	.107	.099	.625

a. Predictors: (Constant), Culture

Table 15 – Model summary. Teacher implementation/school culture.

The following analysis of variables showed a p value less than .001 which indicates that school culture as a variable is significant and is a predictor of teacher implementation.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.542	1	5.542	14.197	.000[b]
	Residual	46.458	119	.390		
	Total	52.000	120			

a. Dependent Variable: implemented

b. Predictors: (Constant), Culture

Table 16 – ANOVA. Teacher implementation/school culture.

Student Learning Regression

The final dependent variable for regression is Student Learning. The model summaries and ANOVA tables for Student Learning are below with regression done using observer knowledge, observer effort, observer power, and school culture as separate independent variables.

Observer knowledge. Observer knowledge was the first independent variable examined in this group. As shown in the model summary below, there is a moderate positive relationship with the student learning dependent variable. Furthermore, 39% of

the variability in the dependent variable can be predicated by the observer knowledge independent variable.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.629[a]	.395	.390	.515

a. Predictors: (Constant), Knowledge

Table 17 – Model Summary. Student learning/observer knowledge.

In the ANOVA table below the p value is less than .001, and observer knowledge as an independent variable is significant. As such, it is a predictor of student learning.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.632	1	21.632	81.710	.000[b]
	Residual	33.093	125	.265		
	Total	54.724	126			

a. Dependent Variable: improved_learning

b. Predictors: (Constant), Knowledge

Table 18 – ANOVA. Student learning/observer knowledge.

Observer effort. The second component of expertise, observer effort, was the next independent variable analyzed using the student learning dependent variable. With an R value of .602, the table showed a moderate positive relationship. Additionally, the observer effort variable predicted 35.7% of the variability.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.602[a]	.362	.357	.526

a. Predictors: (Constant), Effort

Table 19 – Model summary. Student learning/observer effort.

The ANOVA table for this dependent and independent variable pairing showed a p value < .001 indicating that observer effort is a predictor of student learning.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	19.336	1	19.336	69.801	.000[b]
	Residual	34.072	123	.277		
	Total	53.408	124			

a. Dependent Variable: improved_learning

b. Predictors: (Constant), Effort

Table 20 – ANOVA. Student learning/observer effort.

Observer power. The third independent variable related to the observer is observer power. This variable had an R value of .611 and showed a moderate positive correlation to the dependent variable of student learning. Observer power had an adjusted R Square indicating that 36.8% of the variability in student learning can be predicted by observer power.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.611[a]	.373	.368	.526

a. Predictors: (Constant), Power

Table 21 – Model summary. Student learning/observer power.

Observer power is also a predictor of student learning as the calculated p value is less than .001 as demonstrated in the ANOVA table below:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.880	1	20.880	75.473	.000[b]
	Residual	35.135	127	.277		
	Total	56.016	128			

a. Dependent Variable: improved_learning

b. Predictors: (Constant), Power

Table 22 – ANOVA. Student learning/observer power.

School culture. The final independent variable analyzed was school culture. The R value for this variable and the student learning variable was .367 and indicated a small positive relationship. The adjusted R square showed that 12.8% of the variability in student learning could be predicted by the school culture variable.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.367[a]	.135	.128	.620

a. Predictors: (Constant), Culture

Table 23 – Model summary. Student learning/school culture.

The p value for this pairing was p < .001 and showed that school culture is a predicator of student learning as is shown in table 24.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.192	1	7.192	18.699	.000[b]
	Residual	46.153	120	.385		
	Total	53.344	121			

a. Dependent Variable: improved_learning

b. Predictors: (Constant), Culture

Table 24 – ANOVA. Student learning/school culture.

Summary of Survey Findings

Table 25 below is a summary of specific data from the ANOVA and Model Summary tables contained in the data analysis section. In addition to compiling the information, the table is coded so that the independent variable with the highest R value and is in bold and the independent variable with the lowest R value for each of the dependent variables italicized.

Dependent Variable	Independent Variable	**R value**	**Adjusted R-Squared**	**Sig.**
Teacher Growth	Observer Knowledge	0.653	0.421	0.000
	Observer Effort	**0.664**	**0.441**	**0.000**
	Observer Power	0.659	0.430	0.000
	School Culture	*0.310*	*0.089*	*0.001*
Teacher Implementation	**Observer Knowledge**	**0.623**	**0.384**	**0.000**
	Observer Effort	0.619	0.378	0.000
	Observer Power	0.613	0.371	0.000
	School Culture	*0.326*	*0.099*	*0.000*
Student Learning	**Observer Knowledge**	**0.629**	**0.390**	**0.000**
	Observer Effort	0.602	0.357	0.000
	Observer Power	0.611	0.368	0.000
	School Culture	*0.367*	*0.128*	*0.000*

Table 25 – Regression summary

Dependent Variables

Teacher growth. All 4 independent variables have a p value < .05 for the teacher growth dependent variable and are significant. Teacher growth has moderate positive correlation with 3 of the 4 independent variables: observer knowledge, observer effort, and observer power. Each of these independent variables is more strongly correlated to teacher growth than they are to the two other dependent variables. The R values in this set of 3 have a range of .011. Observer effort is the independent variable with the highest correlation to teacher growth with an R value of .664. This is the highest R value for any

independent variable with any dependent variable. Teacher growth has a small positive correlation with school culture. This R value, .310, is the smallest correlation between any independent variable and dependent variable. Overall, the range of R values is .354

Teacher implementation. As with the previous dependent variable, all four independent variables have a p value < .05 for this dependent variable and are significant. Teacher implementation, like teacher growth, has a moderate positive correlation with the same set of 3 independent variables: teacher growth, teacher implementation, and student learning. R values in this set have a range of .010, which is the smallest range for this set of independent variables and a dependent variable. Teacher implementation has a small positive correlation with school culture. The overall range for the entire set of R values is .297.

Student learning. Like the prior dependent variables, all four independent variables are significant and have p values < .05. There is a moderate positive correlation between student learning and the set of 3 observer related independent variables: knowledge, effort, and power. The range of R values in this set is .018, which is the greatest range between this set and any dependent variable. As with the two other dependent variables, school culture has a small positive correlation with student learning. The correlation between student learning and school culture is greater than between school culture and the other two dependent variables. The range of .262 for the entire set of R values is the smallest for all dependent/independent variable sets.

Independent Variables

School culture is the least predictive of all the independent variables for all dependent variables and never accounts for more than approximately 13% of variability.

Observer knowledge is the most predictive independent variable and has the highest correlation with both the teacher implementation and student learning dependent variable. For teacher growth as an independent variable, observer effort is the most predictive and correlated. Observer power was found never to be either the most or least predictive and correlated, and was the only variable to never be at either extreme.

Interview Data

To reiterate, the interview questions used in this study are as follows:

- Describe the students in your school.
- Describe how staff interacts in your school.
- What does your school value?
- Describe the process of your last observation cycle.
- What is your relationship with your observer?
- Describe your observer's knowledge on the evaluation rubric.
- Describe your observer's knowledge on the evaluation classroom management.
- Describe your observer's knowledge on the evaluation pedagogy.
- Describe your observer's knowledge on the evaluation of (your content area).
- How did this observation cycle impact your teaching?

Interviews were audio-taped, transcribed, and coded based on the four concepts: Culture, Observer Effort, Observer Power, and Observer Knowledge. The original study design considered expertise as an important concept as defined by Michelson (2001) that was comprised of both knowledge and effort. However, during the course of the interviews, participants talked extensively and specifically about knowledge and effort to the extent that these two aspects of expertise emerged as important concepts and were subsequently coded individually.

Of these aforementioned four concepts, respondents were only directly asked about culture. As mentioned earlier in the chapter, schools with a more adaptive culture were more likely to have teachers that were open and accepting of feedback and changing practice. For these teachers, the source of the feedback would not be as important as reflecting on that feedback. It was necessary to directly establish the respondent's perception of the school culture to allow culture to serve as a meditating variable. Respondents made 21 comments about school culture and had "flat affect" when describing the school culture. That is, respondents had little variation in tone, speed, volume, or inflection in their delivery of these statements and any gestures were small and smooth. Statements on the other 3 variables occurred spontaneously as respondents described the workings of the observation process. Respondents commented 81 times on these other three variables, and used more descriptive language and varied more in volume, tone, and pitch when speaking about these items.

Observer Effort. The most frequently mentioned variable was observer effort. Respondents mentioned observer effort 49 times in the interviews. This is approximately 61% of the total comments for the three observer variables. Positive comments about effort accounted for 25, or roughly 51%, of these comments. There were 17, or 35%, negative comments and 7, or 14%, neutral comments. During the analysis of the interviews, the concept of effort became focused on the thoroughness of the observer and the extent to which the observer conducted the observation with fidelity to process. This was expressed either as a function of time, feedback, or preparedness. Sample quotes from the interviews are included below and highlight the importance of these three aspects of effort.

Preparedness.

- She got all excited, and she brought in stuff, and she did it with me and we both got excited about the results.
- We have rubrics out. They were pulled up. We were both on screens. We were looking through the look-fors, rubrics, all of those things that I was confident that she knew where to access them, and she knew how to read them, and she had done her work. When I arrived at my meeting, it was all done, and we talked through each of the components. I felt like she knew what to do, it was done, and we were able to have a conversation about it.
- She liked to share, and she wanted to help us all be better teachers, which is good.
- Super positive, super engaged. I always felt he legitimately cared and the questions that he was asking were authentic and gauged for him to understand where I was coming from and how he could support that.
- She would always try to get more suggestions of how you could do things differently.
- I would definitely agree with that. I think with even just doing my first observation this year, I felt like she knew the criteria that I was supposed to meet in way greater detail and gave me way greater detail of feedback compared to last year
- I feel like one of the misconceptions is that they're judging us. And really, she just has so much data and it's surprising to me how fast she can collect it, just based on what I did, which was cool.

- I'm finding that the reviewer has never read it which is disheartening too because you took a lot of time to write it. Read it and ask me a question about it. I wrote some kind of cool things that I know my reviewer didn't read

Time.

- She would actually offer her time instead of observing us. She would come in and cover my class and I could go to another teacher's.
- Our pre-observation meetings lasted probably forty minutes both times. Then he came in and observed the post observation. We went through his script. He showed me some evidence where things were coming from and how he highlighted different things.
- Of all the observations that I've had from administrators I've only had one that I would say is a good one. We knew when I was hired there that I wasn't coming back the next year because of the budget, but he still did probably about five formal observations, tons of drop-ins, drop-ins at conferences. For each of those we would talk for probably eighty or ninety minutes.
- Probably an hour ... not even, maybe 15 minutes pre-observation, a full hour observation, and then 15 to 20 minutes post-observation. And I feel like her feedback was really motivating.
- That was, "I'm an administrator at the back of your classroom on a computer, typing, and made a comment at the start, 'I'm kinda swamped today, so I'm gonna be doing some emailing and things during the observation.' Right there, to me, the value of that has just gone out the window."

- Every other administrator that I've had has flaked out or cancelled one of either the post or the pre or the actual observation itself. At another school I needed to have nine observations and twenty-seven total parts, I probably had ten total. Ten of the twenty-seven.
- I got an email last night saying that I don't have to meet for a post observation unless I want to. Knowing the admin who is doing it, I went in as this is a thing to check off because I have to do it.
- I just think admin doesn't have the time to spend when they're given however many people that they're supposed to be evaluating. I think it just ends up being a time issue more than, "I could do it and am knowledgeable enough to do it, but really I don't have the time to do it."

Feedback.

- I still very much felt like she was comfortable giving me constructive feedback. She didn't sugarcoat things, or change things just because she and I had a relationship prior to her being my observer.
- I feel like it's always been pretty positive for me. Not a whole lot of, "You could do this differently." Which I don't know if it's supposed to be more critical or informative, but typically it's more like, "Here's what I saw, here's how you met these things." It's not a whole lot of, "Here's what you could do better."
- I feel like he's just like, "Oh, you're great. You're doing everything wonderful." I'm like, "There's always things I can improve on." I didn't feel like there was a ton of feedback on areas of improvement.

- I would be fine with someone like if you were going to review me, I give you the paperwork ahead of time, you come review me, and then we talk afterwards.
- I don't recall getting anything specific from the feedback that I wasn't aware of. If he noticed Jimmy asked a low level question, a process question, I was aware of that. I might not have had the data or the tallies of how many questions were asked but I had general sense of it.

Observer Knowledge. The second most mentioned observer variable was observer knowledge. Respondents made 31 comments, approximately 38% of the total comments, related to observer knowledge. Of these 31 responses, 13 were positive, 9 were negative, and 9 were neutral. Respondent's comments were, therefore, 42% positive, 29% negative, and 29% neutral. Respondents commented on the specific content knowledge of their observers or the general pedagogical knowledge of their observer. Some of the contents on general pedagogy were related to the use of the observation rubrics. As Danielson (2007) notes regarding her framework for teaching, on which the district's rubric is based, "The framework applies to virtually every setting. It describes those aspects of teaching that occur in some form in every context" (p. 16) She adds, Each of the four domains of the framework refers to a distinct aspect of teaching. . .Of course, there are many points of connection across domains" (p 29). Therefore, an observer who is knowledgeable about the rubric is knowledgeable about the underlying pedagogical research and assumptions on which the rubric is based. Below are some notable insights from the interviews on content and pedagogical knowledge.

Comments on Content Knowledge.

- She knew the lingo. She knew what to write. She could spot things, and say things, and she was encouraging.
- I enjoyed that I had somebody who's in my area; because I think it's different when you have an observer who is in your content area. I think they see things and understand things differently, because the area was, at one point anyway, their area of expertise.
- There are benefits, of course, by having somebody not in your science area, because they don't look and listen at content, they look at you as the teacher, I think, less so than getting hung up on specifics of content.
- Usually people that come in from the outside are kind of, and they don't know enough. They don't really know if I'm doing a good job teaching details of my content because they don't know what the details are. You're all like educating that person in your field too.
- They don't know. You're probably going to get a higher, less accurate score. We were all afraid at first we'd get lower scores, but what we're realizing and hearing from other people is that they're probably scoring you higher because they don't understand your content.
- I will say when I have had an administrator before who is probably as far out of my department as you can get, that, to me, held significantly less value.
- I feel like it's still hard being observed by somebody who doesn't know the content. When you're coming into my class and observing me, or talking about things you probably haven't seen in years or maybe never have seen, I

feel like how I teach them, if you don't really know the content, it's harder to give me feedback on how I'm teaching that content if you're not familiar with the content.

- Which is helpful to a point, but I don't think that's what this review is supposed to be about. That was a little bit, it was helpful but I didn't feel like we were.
- I've always laughed and wondered how an observation goes for a Spanish teacher who's speaking the whole thing in Spanish, and their observer is non-Spanish speaking. How does that work?

Comments on Pedagogical Knowledge.

- I had no doubt that she knows what she's doing. I feel very confident that she understands the pieces that she was looking for. I felt very confident in that way.
- Yep. I remember an instance, so the first observation students were coming up with questions for Socratic seminar and I remember in our post observation how he took me through the rubric and said, this is what makes distinguishing questioning and students are coming up with the questions rather than teacher generating them. That was an example of him walking through that.
- With a peer I feel like they're not just talking the talk, they're walking the walk with you. They've been there. They've been in your shoes. They've done what you're doing. And not that an administrator hasn't, but it seems more far removed.

- She had a really great visual of where my level of questioning was and where I was going with it, and the level.
- You can observe how students are engaged and you can observe how students are interacting with each other and with me.
- I don't think that he had specific knowledge of it but he was able to draw on some parallels from the other content courses that he's taught when we were talking about the content.
- A lot of the days that I had him come in were not content heavy but they were skill based days for the purpose of demonstrating the rubric. It wasn't necessarily needing that content expertise.
- With newer teachers, I know he had given them suggestions on classroom management stuff, but I've never really received that kind of feedback from him because classroom management hasn't been an issue for me.
- I didn't ever have that moment of, "Does she know what she's ..." There was no question that she knew the process. That has not always been the case with some others I've had along the way, so I recognize that as a bonus.

Observer Power. Only one comment was made that related to observer power for 1% of the total comments. The comment was negative:

> I always felt like I was in trouble in a staff meeting. They'd [the administrators] make everyone feel terrible because one teacher ordered food and instead of going to that teacher and talking to them about it they would just yell at the whole staff. I never ordered food here in my life, but I left the meeting feeling really ashamed and like I should never have ordered food.

Summary of Quantitative and Qualitative Data Findings

Taken together, the findings from the survey administered to teachers and the eight individual teacher interviews have shown that all three of the observer characteristics have a positive relationship with the evaluation outcomes, but expertise, particularly observer effort, is the most important factor in the observation process. Survey results demonstrated that observer characteristics, as a whole, had the most significant relationship with teacher growth, then teacher implementation, and finally, perceived student learning. However, the two expertise variables, knowledge and effort, both had a more significant relationship with the outcomes than observer power had with the outcomes. The relationship between observer effort and teacher growth was the most positive result of all regression pairs.

It was important to examine the greater context for these survey results because the surveys focused on teacher perceptions. The value of expertise, and specifically effort, was addressed more specifically in the interviews results based on both the quantity and quality of the comments made regarding effort, knowledge, and power. Ninety-nine percent of all comments were related to effort or knowledge. The majority of these comments were positive perceptions. Overall, comments on effort were both the most numerous and focused on the observer prioritizing evaluation meetings over other obligations, connecting specific feedback to personal growth plans, providing new resources, and being prepared for meetings. Comments on knowledge focused on the observer's knowledge of the content, pedagogy, and the evaluation process and rubrics.

Chapter Five: Conclusion

Overview

The 2001 reauthorization of the Elementary and Secondary Education Act, commonly known as No Child Left Behind (NCLB) is often credited with starting the accountability movement in education. While NCLB did raise the profile of accountability, its roots reach deeper into the past. The legal decision in Brown v. Board of Education (1954) planted the seed of equity that would germinate into the passage of the original Elementary and Secondary Education Act (ESEA) in 1965. As part of President Lyndon B. Johnson's War on Poverty, the ESEA initiated the concept of addressing the achievement gap. It is the call to eliminate the gap in performance between different racial and socioeconomic groups that is at the core of accountability.

School systems are looking for ways to answer this challenge and the research base points to one factor as the most important school-based factor in student achievement: teacher quality (Louis et al., 2010; Leithwood, Louis, Anderson, & Wahlstrom, 2004; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2000; Rowan, Correnti & Miller, 2002; Wright, Horn, & Sanders, 1997). Teacher quality has historically been measured through the observation process. The observation process has historically been used summatively, to assess teacher performance, and formatively, to provide feedback for teacher growth. Several studies have examined ways to more reliably observe teacher quality (Kane, T. J; Staiger, D. O., 2010 ; Kane, T. J; Staiger, D. O. ,2012; Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O., 2013), and have suggested ways to improve summative observations.

School leaders, however, need to do more than assess quality; they, then, need to use the assessment information to improve teacher quality.

Towards that end, this study focused on two observer characteristics, power and expertise, that could affect whether teachers perceived the observation process had the following outcomes: it increased their growth, made it more or less likely they would implement suggestions, or perceived that it increased student learning. The results of the interviews presented an emerging distinction between two separate aspects of expertise, resulting in the concept of expertise being redefined as observer knowledge, content and pedagogy, and observer effort. The original research questions to address those issues were, therefore, modified to reflect the findings. The modified questions became:

1. What role does the perceived knowledge, both content specific and general pedagogical, of the observer play in teacher observations for the observer and the observed?
2. What role does the perceived effort of the observer play in teacher observations for the observer and the observed?
3. How does the perceived power relationship between observer and observed shape teacher observations?

It was noted in Chapter 3 that an observer can be perceived to be high or low in each of these characteristics, and possible combinations of the original concepts were included in Figure 1 in that chapter. Teachers were asked in surveys and interviews about these characteristics, as well as questions about school culture. The responses from both the survey and the interviews were analyzed to ascertain their perceived impact on

three outcomes listed above and a new visual representation of these combinations is included in the Discussion of Findings which follows.

The Role of the Observer

The study focused on three qualities, or characteristics, of the observer: power, effort, and knowledge. Each of these characteristics was significantly positively associated with all three dependent variables, or outcomes: teacher growth, teacher implementation, and student learning. However, not all of these observer characteristics have the same impact on the outcomes, nor do they impact the evaluation outcomes to the same degree. These distinctions warrant discussion because of the impact they can have for the quality of the evaluation and observation process. However, before discussing the impact of the characteristics on the outcomes, it is important to review what specific elements make up each of the observer characteristics.

Overview of Observer Characteristics

Observer power might seem to be the most easily understood of the three characteristics because of the assumption that it is based on position. However in this study, power was a combination several elements based on the work of Michelson (2001) who identified five factors that contribute to positional power: centrality, criticality, flexibility, visibility, and relevance. Three factors, criticality, flexibility, and visibility, are based on how leaders are perceived during the process of making building-level decisions Therefore, the district's structure that has the peer evaluators' offices located in the central administration office building and not in the individual school sites removes the peer evaluators from participating in building-level decision-making. It also means that the only positional power factors that could be directly observed by the teachers,

related to peer evaluators, were issues of centrality and relevance. Specifically, the survey measured centrality by asking if the evaluator was accessible and gave enough time. The survey measured the concept of relevance by asking if the observer fostered growth.

Similarly, observer knowledge is a combination of factors related to knowledge about instruction. Teachers were asked the surveys and interviews if their observers were knowledgeable in their particular content area, as well as general pedagogy, classroom management, and the evaluation rubric. Additionally, teachers were asked in the survey if their observer provided new resources related to any of the aforementioned areas. This range of knowledge of factors related to instruction would allow observers to be strong in several areas, but perhaps not as strong in others.

The final observer characteristic is observer effort. In this system, teachers complete a pre-observation form that includes personal goals and lesson objectives, and teachers and evaluators also have a meeting before and after the actual observation. Therefore, when asked to consider the evaluator's effort, teachers were asked in both the survey and interviews if their evaluator understood lesson objectives, gave detailed feedback connected to the lesson objectives and personal goals, provided adequate meeting times, and made time as needed outside scheduled meetings.

Discussion of Observer Characteristics

It is helpful to briefly examine two summary charts before discussing the implications for this study. First, it is important to review the data collected in the survey due to the significance that each of the independent variables had on the dependent variables. Table 26 provides a summary of these data which shows that each independent

variable is significant for each dependent variable, based on a significance level of .05. Therefore, each independent variable is a predictor for changes in each dependent variable. Furthermore, Table 26 shows that the variables related to observer characteristics, knowledge, effort, and power, have, at a minimum, a moderate positive relationship with the dependent variables of teacher growth, teacher implementation, and perceived student learning. In addition to compiling the information, the table is coded so that the independent variable with the highest R value and is in bold and the independent variable with the lowest R value for each of the dependent variables italicized.

Dependent Variable	Independent Variable	R value	Adjusted R-Squared	Sig.
Teacher Growth	Observer Knowledge	0.653	0.421	0.000
	Observer Effort	**0.664**	**0.441**	**0.000**
	Observer Power	0.659	0.430	0.000
	School Culture	*0.310*	*0.089*	*0.001*
Teacher Implementation	**Observer Knowledge**	**0.623**	**0.384**	**0.000**
	Observer Effort	0.619	0.378	0.000
	Observer Power	0.613	0.371	0.000
	School Culture	*0.326*	*0.099*	*0.000*
Student Learning	**Observer Knowledge**	**0.629**	**0.390**	**0.000**
	Observer Effort	0.602	0.357	0.000
	Observer Power	0.611	0.368	0.000
	School Culture	*0.367*	*0.128*	*0.000*

Table 26 – Regression summary

The second table to consider is a summation of the interview data that was collected. The interviews were coded using the independent variables. Also, representative responses of the interviewees were rated as positive perceptions, negative perceptions, or neutral perceptions based on the content and context. Table 27 summarizes the findings of the coding process.

Independent Variable	Total items Coded	Positive Perceptions	Negative Perceptions	Neutral Perceptions
Observer Effort	49	51%	35%	14%
Observer Knowledge	31	42%	29%	29%
Observer Power	1		100%	
School Culture	21			100%

Table 27 – Survey coding summary

Together, these tables provide a summary of the data analyzed in more detail in the preceding chapter. These are included here to provide additional context for discussing the implications of this study.

Discussion of Findings and Implications for Practice

Observers Have an Impact

The first implication from the data is that observers can have an impact on the evaluation. All three of the observer characteristics had a significant relationship with all three of the teacher outcomes. That is, the greater the perception of observer power, knowledge, or effort, the more teachers perceived their observation led to increased teacher growth, teacher implementation, and student learning. Therefore, all three observer characteristics are important tools for observers to consider. Most importantly, the fact that each of these observer characteristics are measured by several items in the rating scale means that perceived improvement on even one item raises overall perception of knowledge, effort, or power. Therefore, observers can focus on a wide variety of items to change the way they are perceived by teachers.

A key characteristic that emerged from the data analysis was observer effort. Observer effort had the most significant relationship with teacher growth and the most

significant relationship of any independent variable on any dependent variable. Additionally, it has the second largest impact across dependent variables. The actionable traits of observer effort are completely under the control of the observer. For example, he or she can take time to read the lesson plan in advance to get a better understanding of the lesson objectives and better prepare to ask probing questions during the pre-observation meeting. Likewise, the observer can read the teacher's personal growth plan to more fully understand what the teacher wants to target for his or her own professional development, and thus can seek to better understand how the teacher wants feedback on these targets. Once the observer understands the objectives and growth targets, he or she can give specific feedback on these items to the teacher. The observer needs to allocate adequate time for the meetings with the teacher being observed and needs to prioritize meetings with teachers being observed over other demands on the evaluator's time. Finally, the observer can check in at other times with teachers under an observation schedule, doing so in person or electronically, to demonstrate availability. Unless an observer is already doing all of these things, he or she can adapt one or more of these strategies to increase the perception that he or she is making a significant effort towards the observation process.

An observer also has multiple avenues to increase the perception that he or she is a knowledgeable observer. Observer knowledge was most significant factor in determining if a teacher would implement suggestions on student achievement, as reported by teachers in this study. It also had the largest combined overall effect on the dependent variables. At the secondary level where buildings are organized into content-area departments instead of grade levels, knowledge of the content area by the observer is

the most difficult of the knowledge elements to address. Fortunately, this is only one of the key elements for being an effective observer and improving in the other elements can compensate for lack of content knowledge.

For example, an observer can make sure that he or she understands the evaluation rubric and criteria. While knowledge of the rubric is, itself, one element, this knowledge also impacts the element of general pedagogy. In the schools participating in this study, the evaluation rubric is based on Charlotte Danielson's framework (Danielson, 2007). This framework has four domains, three of which are based on pedagogy. Therefore, understanding the rubric requires an understanding of the Danielson's instructional philosophy. Additionally, observers in this study had viewing access to shared documents that defined "look-fors" for many content areas. These documents were created by expert teachers and they list examples of observable teacher behavior that demonstrates proficiency in each domain. Other districts would benefit from emulating this practice. Another element that transfers across content areas is classroom management. Attuning to classroom behaviors during observations provides an avenue to give feedback and strategies to teachers. Finally, observers can stay current on developments in pedagogy and classroom management, so that they can provide resources to teachers struggling in these areas.

Expertise is more significant than power

Taken together, the two areas, observer knowledge and observer effort, comprise expertise as defined by this study. In addition to expertise, the second scale that this study examined was observer power. Power had the least significant relationship with the dependent variables. Furthermore, power was not the largest contributor to any

individual dependent variable; one or both expertise characteristics, observer knowledge or observer effort, had a larger significance for each of the outcomes.

Previously, four possible combinations of observer expertise and power were noted in the 2X2 matrix for Figure 1 shown previously in Chapter 3. Given the findings from this study, and in light of the relative significance of expertise, now split between observer knowledge and observer effort, and high and low power, it is possible to rank these four combinations in terms of their significance for effective teacher evaluations. A modification of Figure 1 is shown in Figure 2 below.

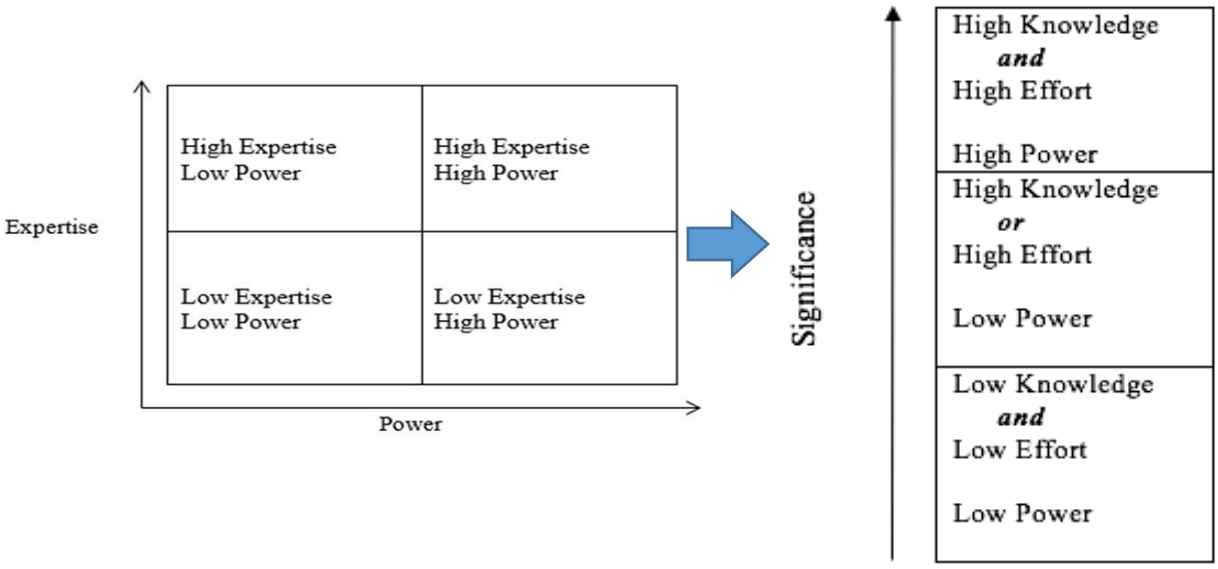


Figure 2 – Ranked significance of power, knowledge, and effort combinations

Given the ranking provided above, observers could find greater benefits for teachers when they place a high priority on their observation strategies and use of available time, knowing that expertise (knowledge and effort) has greater impact on perceived benefits of observations, instead of the assumption that power has greater influence.

Systems can help increase teacher quality

School districts and individual buildings can also make changes in their systems to allow the observation and evaluation process to increase teacher quality. Systems generally address and allocate power by assigning positions and roles to individual employee categories, such as principal or teacher. These positions are defined contractually and roles within those positions are difficult to change without negotiations. However, expertise across roles can be more easily influenced by district and building leadership and, as found in this study, can also be more significant. Two specific ways of aligning expertise can be accomplished, one through the manner in which observers are assigned to which teachers and second, how observers are trained.

Different observers can be assigned to different staff to maximize their expertise. Maximizing expertise can be accomplished by focusing on the knowledge characteristic and/or the effort characteristic. There are several ways to improve pairings to increase observer knowledge characteristics. One way to accomplish this in a building is to assign administrators and peer observers to teachers with the same content expertise whenever possible. Where this is not possible, a building may choose to assign observers in similar cross-content areas, such as a math/science pairing or an English/social studies pairing. These areas are similar because they use more similar instructional strategies. On a larger scale, the optimal pairing for purposes of sharing greater expertise can be facilitated by the district by pairing observers across buildings. Traditionally, administrators have only been assigned to observe teachers in their own buildings. Allowing administrators to observe teachers in other buildings potentially increases the number of content-matched pairings.

Of course, principals want to be involved in the observations for their own staff members when evaluations are to be used for summative or retention purposes. At such times, they may be skeptical of using a principal from another building. However, as research from the Gates Foundation MET study found, the best way to increase the validity of an evaluation is to have more observers rather than more observations (Ho and Kane, 2013). One final way to increase knowledge is to have an observer develop longevity in observing specific content areas for which the building and/or district cannot make a same or similar content match. This might be in a special area, such as art, music or a world language. The observer, over time, can see different strategies in action and gauge their effectiveness by periodic reviewing of the summative assignment or assessment.

Expertise can also be maximized by focusing on the effort characteristic. Systems that allow for greater effort would allow observers to prioritize time spent on evaluations over time spent on other assigned duties. One way to do this would be to have more peers doing evaluations. As several interviewees noted, “With a peer I feel like they're not just talking the talk, they're walking the walk with you. They've been there. They've been in your shoes. They've done what you're doing. And not that an administrator hasn't, but it seems more far removed.”

Another way to allow for greater effort is to ensure that observers are assigned only as many teachers as they can fully accommodate with reasonable performance expectations. That is, determine the optimal caseload based on the time the average observer takes for an observation cycle. Also, additional staff can be added or duties reassigned to allow more time for the observation process, such as using a system of

deans to handle low-level discipline issues and other basic management tasks that can otherwise absorb the limited time that school leaders have to conduct evaluations.

Finally, another means by which to assist the improvement of the observation process would be to increase knowledge of observers via professional development. Several teachers who were interviewed noted how much they valued specific feedback. For example, one teacher noted "I still very much felt like she was comfortable giving me constructive feedback." Others also mentioned how effective observers connected their feedback to the rubric, such as "I felt like she knew the criteria that I was supposed to meet in way greater detail and gave me way greater detail of feedback compared to last year." Given these insights, training observers to better understand the rubric and to provide more specific, targeted feedback would make observers more knowledgeable. On a positive note, the district participating in this study does provide extended professional development for observers by using recorded lessons as a training tool for inter-rater reliability.

A Note on the School Culture Variable

One of the encouraging results of this study was the minimal significance of the school's culture on a teacher's reported growth, implementation of new strategies, and student learning. The use of culture as a mitigating variable was predicated on the concern that if a school that has a more innovative culture, or has a staff that embraced what we currently call a "growth mindset", the teachers might be more receptive to change and growth than a school culture that is perceived to be more resistant. That is, the impact on teacher growth, teacher implementation, or student achievement in some schools might have been due more to individual differences in staff or school cultures

than it was to observer power or expertise. However, given the finding of this study that there was only a small level of significance related to school culture, the significance of the power and expertise of the observer is increased.

This is not to say that a school's culture is not a factor to be considered in the future. The two schools that were studied are part of the same school system and use the same evaluation system, and yet, it is likely that they each have a somewhat different culture from one to the other. Nevertheless, the commonalties may diffuse any differences in culture that may exist between them. The peer evaluators in the district for this study have a significant level of training in observation, feedback, and peer coaching. Likewise, the administration uses the same evaluation tools, rubrics, and receives much of the same training. Additionally, the state requires that teachers are to be evaluated by an administrator once every three years. Continuing contract teachers are evaluated by a peer during the other two years of the cycle. The district also participates in the state's teacher performance pay program, which requires annual observations and goal setting. The fact that both the peers and administrators use the same processes and tools means that the only significant systemic difference between peer and administrative cycles is that the administrative cycle ends with an additional summative performance appraisal. While culture might be a more significant variable if two or more schools operating in a different system were studied, in this study, people mattered more.

Areas for Further Study

This study has provided evidence to support the importance of the roles that perceived power and expertise of the observer, by the teacher, plays in teacher observations. Achieving greater understanding of why that is the case could be

strengthened and be made more generalizable through further study. Some possible areas for further study include the use of more demographic data, missing power elements, and alternate settings, as described below.

The first unexplored area would be to examine additional demographic information of the teachers and observers that might impact the relationship between the independent and dependent variables. Gathering more demographic data about the teachers and the observers would allow for a deeper analysis along these lines. Some specific data to examine might be years in a system, years of teaching experience, and the gender of the observer and teacher. Looking at these specific areas could suggest additional ways to optimize observer assignments and time allocation.

In addition to collecting demographic data of teachers, another area for further study would be to explore the elements of power from Michelson's model (Michelson, 2001) that were not examined in this study. In brief, Michelson's survey included centrality, relevance, criticality, flexibility, and visibility (see Table 28 below). As mentioned earlier in the chapter, the setting of this study precluded the use criticality, flexibility, and visibility. These three elements were initially part of the original survey developed for this study, but were removed at the request of the cooperating district. Adding these elements would give a more complete picture of the perception of the observer's power.

Elements of Positional Power	Description
Centrality	Relationship between positions in a communication network
Criticality	Relationship between tasks performed in a work flow process
Flexibility	Amount of discretion vested in a position
Visibility	Degree to which a task performance is seen by influentials in the organization
Relevance	Relationship between a task and organizational priorities

Table 28 – Michelson's elements of positional power. (Michelson, 2001. p. 195)

A final area of further study would include conducting similar research as was done here, but in different settings. There are two possibilities for adjusting the setting of additional studies. First, a similar study could be replicated in secondary schools in different districts and/or states. Breaking away from the homogeny of the system might give further insight into the role of culture in influencing the role of power and expertise in the evaluation system. The second alternative study would be to examine the roles of power and expertise of observers in elementary settings. Whereas the secondary level is organized into content-specific departments, elementary schools are organized into grades. This distinction potentially could have an interesting effect on the area of content expertise, as elementary teachers are licensed to teach all content areas in all elementary grades.

Conclusion

This study sought to discover the role of power and expertise in the teacher observation process. It was found that both power and expertise were positively associated with teacher's perceptions of their growth. Moreover, expertise was found to be more

significant than power, with observer effort being the most significant characteristic. This suggests that observers should prioritize being thorough in the observation process over their other competing obligations. Ultimately, teachers will benefit from this effort by having more reflective conversations and targeted feedback on their pedagogy, which may increase the likelihood they will implement suggested improvements. Consequently, students will benefit from having better teachers. Learning is the goal for all participants in education and access to quality teachers is a component of making sure education is equitable. Thus, society will benefit from increased observer effort.

References

Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology and Society. 24*(3), 200–212.

Bill & Melinda Gates Foundation. (2010). Learning about teaching: Initial findings from the Measures of Effective Teaching Project. *MET Project Research Paper*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf

Council of Chief State School Officers. (2016). Comparison of select elements of ESEA. Retrieved from http://www.ccsso.org/Documents/2016/ESSA/CCSSOComparisonofSelectElementsofESEA3.31.16.pdf

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: ASCD.

Darling-Hammond, L. (2012). *Creating a comprehensive system for evaluation and supporting effective teaching*. Stanford, CA. Stanford Center for Opportunity Policy in Education.

Education Code Prekindergarten – Grade 12. Minn. Stat,, § 122A.40, Subd. 8 (2016).

Epstein, J. L. (1985). A question of merit: Principals' and parents' evaluations of teachers. *Educational Researcher, 14*(7), 3-10.

Evertson, C.M., & Burry, J.A. (1989). Capturing classroom context: The observation system as lens for assessment. *Journal of Personnel Evaluation in Education, 2,* 297-320.

French, J. R. P. & Raven, B. (1959) The bases of social power. In D. Cartwright and A. Zander. *Group Dynamics* (pp. 259-269). New York, NY: Harper & Row.

Garmston, R., & Wellman, B. (1999). The adaptive school: A sourcebook for developing collaborative groups. Norwood, MA: Christopher.

Heneman, H.G., & Milanowski, A.T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(2), 173-195.

Ho, A. D., & Kane, T. J. (2013). The reliability of classroom observations by school personnel. *MET Project Research Paper*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/MET_Reliability%20of%20Classroom%20Observations_Research%20Paper.pdf

Johnson, L., & Shields, J. (2007). Lessons from management-union partnership in teacher performance appraisal in the New South Wales public education system. *International Journal of Human Resource Management, 18*(7), 1214-1227.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. *MET Project Research Paper.* Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/MET_Validating_Using_Random_Assignment_Research_Paper.pdf

Kane, T. J; Staiger, D. O. (2012). Gathering feedback for teaching: Combing high-quality observations with student surveys and achievement gains. *MET Project Research Paper.* Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from

http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Kruse, S., Louis, K. S., & Bryk, A. (1994). Building professional community in schools. *13 Parameters: A Literacy Leadership Toolkit, Research Resource Book*. *Issues in restructuring schools*, *6*, 3-6.

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis*, *25*(3), 287-298.

Laws of Minnesota 2011, 1st Spec. Sess. chapter 11, article 2, section 14. Retrieved from https://www.revisor.mn.gov/laws/?id=11&doctype=Chapter&year=2011&type=1

Leithwood, K., Louis, K.S., Anderson, S., & Wahlstrom, K. (2004). *Learning from leadership: A review of the literature*. Published by the Center for Applied Research and Educational Improvement, University of Minnesota, Minneapolis, MN.

Leonard, M. (2003). The A-Z of Social Research. In J. D. Miller, R. L. Brewer (Ed.), *The A-Z of Social Research* (pp. 167–172). London, UK: SAGE Publications, Ltd. doi:10.4135/9780857020024

Louis, K.S., Leithwood, K., Wahlstrom, K., Anderson, S., Michlin, M., Mascall, B., Gordon, M., Strauss, T., Thomas, E., & Moore, S. (2010). *Learning from leadership: Investigating the links to improved student learning. Final report of research.* Center for Applied Research and Educational Improvement. Minneapolis, MN: University of Minnesota. Also, The Wallace Foundation, New York, NY.

McCaffrey, J. R., Lockwood, D. F., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value added models for teacher accountability* [Monograph]. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf

Minnesota Department of Education. (2013). *The Teacher Development, Evaluation, and Peer Support Model: Implementation Handbook*. Retrieved from http://education.state.mn.us/MDE/dse/edev/mod/

Minnesota Office of the Legislative Auditor. (2009, February). *Q comp: Quality Compensation for Teachers.* Retrieved from http://www.auditor.leg.state.mn.us/ped/pedrep/qcomp.pdf

Michelson, B. J. (2001). Leadership and power base development: using power effectively to manage diversity and job-related interdependence in complex organizations. In R. I. Lester & A. G. Morton (Eds.), Concepts for Airforce Leadership (pp. 193–199). Maxwell Air Force Base, Alabama: Air University Press. Retrieved from http://www.au.af.mil/AU/AWC/awcgate/au-24/michelson.pdf

Mihaly, K., McCaffrey, D. F., Staiger, D. O., & Lockwood, J. R. (2013). *A Composite Estimator of Effective Teaching*. Arlington, VA: RAND Corporation. Retrieved from http://www.metproject.org/downloads/MET_Composite_Estimator_of_Effective_Teaching_Research_Paper.pdf

Nelson, T. L. (2008). Perception Question. In *Encyclopedia of Survey Research Methods*. (pp. 580-582) Thousand Oaks, CA: SAGE Publications.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Ovando, M.N., & Harris, B.M. (1993). Teachers' perceptions of the post-observation conference: Implications for formative evaluation. *Journal of Personnel Evaluation in Education, 7,* 301-310.

Ovando, M.N., & Ramirez, A., Jr. (2007). Principals' instructional leadership within a teacher performance appraisal system: Enhancing students' academic success. *Journal of Personnel Evaluation in Education, 20,* 85-110.

Peterson, K. D. (1987). Teacher evaluation with multiple and variable lines of evidence. *American Educational Research Journal*, *24*(2), 311-317.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2000). *Teachers, schools, and academic achievement* (Working Paper W6691). Cambridge, MA: National Bureau of Economic Research.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record, 104,* 1525-1567.

Steffy, B. E., Wolfe, M.P., Pasch, S.H., & Enz, B.J. (Eds.). (2000). *Life cycle of the career teacher*. Corwin Press.

Taut, S., Santelices, V., Araya, C., & Manzi, J. (2010). Theory underlying a national teacher evaluation program. *Evaluation and Program Planning, 33,* 477-486.

U.S. Department of Education. *Race to the Top Program Executive Summary.* Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project. Retrieved from http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Wise, A.E., Darling-Hammond, L., McLaughlin, M.W., Bernstein, H. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA. RAND.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teachers and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67

York-Barr, J., Sommers, W. A., Ghere, G. S. & Monte, J., (Eds.). (2006). *Reflective practice to improve schools: An action guide for educators* (2nd ed.). Thousand Oaks, CA: Corwin.

Appendix A: Instruments

Observation Cycle Reflection

Please answer questions based on your last full cycle of observations.

Context
Who administered your last observation cycle?

☐ Principal/Asst. Principal ☐ Peer Evaluator

Based on my last full observation cycle: **My last observer:**	Strongly Disagree	Disagree	Agree	Strongly Agree
Was knowledgeable in my content area	☐	☐	☐	☐
Was knowledgeable about general pedagogy	☐	☐	☐	☐
Was knowledgeable about classroom management	☐	☐	☐	☐
Was knowledgeable about the evaluation rubric	☐	☐	☐	☐
Provided new resources	☐	☐	☐	☐
Understood my lesson objectives	☐	☐	☐	☐
Wrote detailed feedback	☐	☐	☐	☐
Connected feedback to details from observed lesson	☐	☐	☐	☐
Connected feedback to my personal growth plan	☐	☐	☐	☐
Provided adequate meeting time to discuss feedback	☐	☐	☐	☐
Was available outside of scheduled meetings	☐	☐	☐	☐
Allotted enough time for me	☐	☐	☐	☐
Was easily accessible	☐	☐	☐	☐
Fostered professional growth	☐	☐	☐	☐

Outcomes:	Strongly Disagree	Disagree	Agree	Strongly Agree
My last observation helped me grow	☐	☐	☐	☐
I implemented suggestions from my last evaluation	☐	☐	☐	☐
Strategies I implemented from my last evaluation improved student learning	☐	☐	☐	☐

Organizational Culture	Strongly Disagree	Disagree	Agree	Strongly Agree
I know our school goals	☐	☐	☐	☐
I develop my own goals	☐	☐	☐	☐
My school is open to new ideas	☐	☐	☐	☐
My collaborative team has a common vision	☐	☐	☐	☐
My collaborative team shares successes and failures	☐	☐	☐	☐
My collaborative team plans together	☐	☐	☐	☐
My collaborative team considers how our instructional changes might affect our colleagues throughout our school	☐	☐	☐	☐

Demographics

Total years teaching including this year _____

Content area(s) taught during last observation _____

Interview questions

1. Describe the students in your school.
2. Describe how staff interacts in your school.
3. What does your school value?
4. Describe the process of your last observation cycle.
5. How would characterize your interactions with your observer (e.g., formal, relaxed)?
6. Describe your observer's knowledge of the evaluation rubric.
7. Describe your observer's knowledge of classroom management.
8. Describe your observer's knowledge of pedagogy.
9. Describe your observer's knowledge of your content.
10. How did this observation cycle impact your teaching?

Appendix B: Study Approval

Institutional Review Board Approval

Michael Bury <bury0009@umn.edu>

1603E85342 - PI Bury - IRB - Exempt Study Notification

1 message

irb@umn.edu <irb@umn.edu> Fri, Apr 8, 2016 at 10:53 AM
To: bury0009@umn.edu

TO : wahls001@umn.edu, bury0009@umn.edu,

The IRB: Human Subjects Committee determined that the referenced study is exempt from review under federal guidelines 45 CFR Part 46.101(b) category #2 SURVEYS/INTERVIEWS; STANDARDIZED EDUCATIONAL TESTS; OBSERVATION OF PUBLIC BEHAVIOR.

Study Number: 1603E85342

Principal Investigator: Michael Bury

Title(s):
Examination of Teacher Observation Dynamics: Power and Expertise

---

This e-mail confirmation is your official University of Minnesota HRPP notification of exemption from full committee review. You will not receive a hard copy or letter.

This secure electronic notification between password protected authentications has been deemed by the University of Minnesota to constitute a legal signature.

The study number above is assigned to your research. That number and the title of your study must be used in all communication with the IRB office.

Research that involves observation can be approved under this category without obtaining consent.

SURVEY OR INTERVIEW RESEARCH APPROVED AS EXEMPT UNDER THIS CATEGORY IS LIMITED TO ADULT SUBJECTS.

This exemption is valid for five years from the date of this correspondence and will be filed inactive at that time. You will receive a notification prior to inactivation. If this research will extend beyond five years, you must submit a new application to the IRB before the study's expiration date. Please inform the IRB when you intend to close this study.

Upon receipt of this email, you may begin your research. If you have questions, please call the IRB office at (612) 626-5654.

You may go to the View Completed section of eResearch Central at http://eresearch.umn.edu/ to view further details

on your study.

The IRB wishes you success with this research.

We value your feedback. We have created a short survey that will only take a couple of minutes to complete. The questions are basic, but your responses will provide us with insight regarding what we do well and areas that may need improvement. Thanks in advance for completing the survey. http://tinyurl.com/exempt-survey

Cooperating District Approval

A copy of a letter from the district involved in this study is included on the following page. The letter has been redacted to mask the identity of the district. This redaction was necessary to increase the confidentiality of the survey respondents and interview participants. Survey and interview participants were informed that their responses would be confidential and allowed questions to probe more sensitive issues.

[redacted]
District Letterhead

March 1, 2016

Hello Mike,

We are pleased to inform you that your research project through the University of Minnesota has been approved for completion within [redacted] School District pending your Confidentiality/Non-disclosure Agreement being signed.

However, as part of our process and procedures, the principals at [redacted] and [redacted] principal at [redacted] have approved your request to proceed with your study at their high schools. At this time, [redacted] principal at [redacted] is declining participation at this time.

I will pass your information along to [redacted] who will contact you to set up a plan for signing the Confidentiality/Non-disclosure Agreement once she has drawn up the agreement. Once we have received a signed copy of this agreement you are free to begin your research.

Upon completion of your research, please submit a copy of your findings to the Research, Evaluation, and Testing Department. We are looking forward to hearing of your results.

If there is anything else you need, please let me know.

Thank you,

[redacted]
Signature

Achievement Analyst
Research, Evaluation and Testing

[redacted]
Contact information

[redacted]
Signature

Director, Research, Evaluation and Testing

[redacted]
Contact information

[redacted]
District Footer

Appendix C: District Process Documents

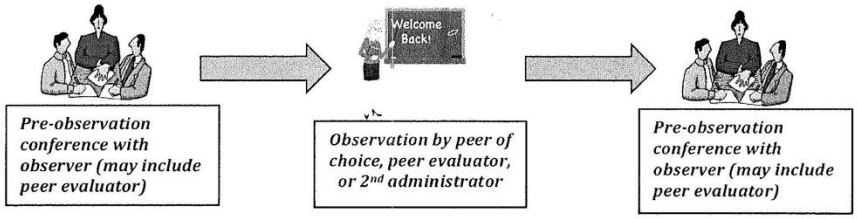
This appendix contains documents used by the administrators and peer evaluators in the district. The first three pages are the summary documents that clarify the observation cycles and differences between them for teachers in high, low, and probationary cycles. The remaining pages contain the evaluation rubric.

High Cycle at a Glance

Observations – Three cycles of observations collect evidence of proficiency in twenty-two component areas.

Observation Cycle 1 – Teachers self-assess all 22 components. INSTRUCTIONAL domain components are the scored, but assigned administrator also collects data in other components.

Observation Cycle 2 – Formative observation conducted by a **peer of choice, peer evaluator, or a different administrator**. Focused on a targeted area of growth.



Observation Cycle 3 – CLASSROOM ENVIRONMENT domain is scored if not previously completed. Summative observation / evaluation conducted by assigned administrator.

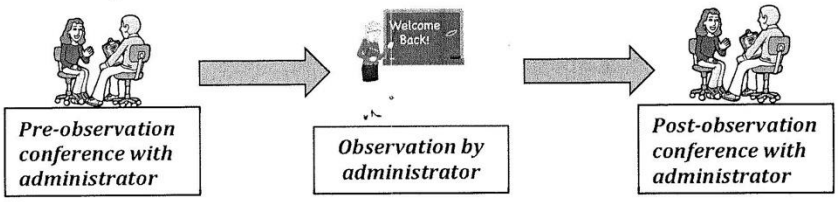
Probationary Cycle at a Glance

Observations – Three cycles of observations collect evidence of proficiency in twenty-two component areas.

Observation Cycle 1 – Teachers self-assess all 22 components. Administrator focuses on INSTRUCTIONAL domain, but evidence can also be collected in other domains.

Observation Cycle 2 – Additional evidence collected by administrator. Focus is on the CLASSROOM ENVIRONMENT domain components, but evidence is also collected in other areas.

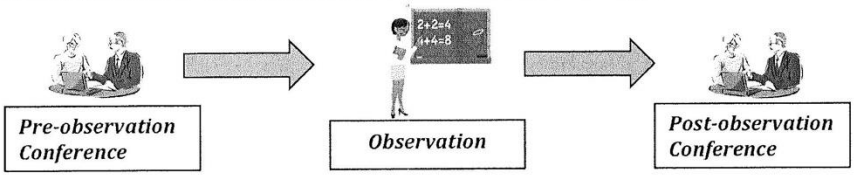
Observation Cycle 3 – Summative observation and evaluation conducted by administrator.



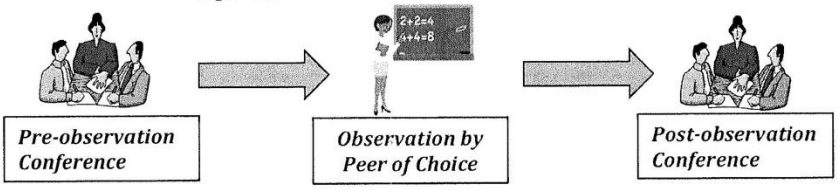
Low Cycle at a Glance

Observations – Three cycles of observations collect evidence of proficiency in five component areas.

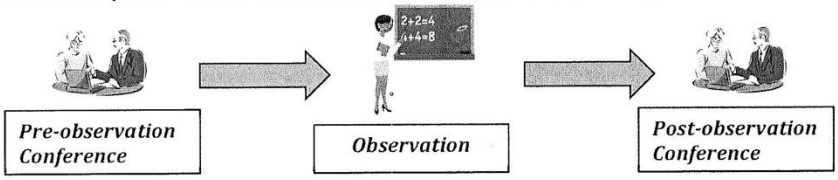
Observation Cycle 1 – Data is collected in all five component areas by peer evaluator to set a baseline.



Observation Cycle 2 – Formative observation conducted by a peer of choice. Focused on a targeted area of growth.



Observation Cycle 3 – Summative observation / evaluation conducted by peer evaluator.



Performance Appraisal System
Regular Education
Rubrics
COMPONENTS OF PROFESSIONAL PRACTICE

Domain 1: Planning and Preparation

Component 1A:	Demonstrating Knowledge of Content and Pedagogy
Component 1B:	Demonstrating Knowledge of Students
Component 1C:	Setting Instructional Outcomes
Component 1D:	Demonstrating Knowledge of Resources
Component 1E:	Designing Coherent Instruction
Component 1F:	Designing Student Assessments

Domain 2: The Classroom Environment

Component 2A:	Creating an Environment of Respect and Rapport
Component 2B:	Establishing a Culture for Learning
Component 2C:	Managing Classroom Procedures
Component 2D:	Managing Student Behavior
Component 2E:	Organizing Physical Space

Domain 3: Instruction

Component 3A:	Communicating Clearly and Accurately
Component 3B:	Using Questioning and Discussion Techniques
Component 3C:	Engaging Students in Learning
Component 3D:	Using Feedback and Assessment
Component 3E:	Demonstrating Flexibility and Responsiveness

Domain 4: Professional Responsibilities

Component 4A:	Reflecting on Teaching
Component 4B:	Maintaining Accurate Records
Component 4C:	Communicating with Families
Component 4D:	Participating in the Professional Community
Component 4E:	Growing and Developing Professionally
Component 4F:	Showing Professionalism

DOMAIN 1: Planning and Preparation

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
1A. Demonstrating Knowledge of Content and Pedagogy	In planning and practice, the teacher makes content errors or does not correct errors made by students. The teacher displays little understanding of prerequisite knowledge important to student learning of the content. The teacher displays little or no understanding of the range of pedagogical approaches suitable to student learning of the content.	The teacher is familiar with the important concepts in the discipline but displays a lack of awareness of how these concepts relate to one another. The teacher indicates some awareness of prerequisite learning, although such knowledge may be inaccurate or incomplete. The teacher's plans and practice reflect a limited range of pedagogical approaches to the discipline or to the students.	The teacher displays solid knowledge of the important concepts in the discipline and how these relate to one another. The teacher demonstrates accurate understanding of prerequisite relationships among topics. The teacher's plans and practice reflect familiarity with a wide range of effective pedagogical approaches in the subject.	The teacher displays extensive knowledge of the important concepts in the discipline and how these relate both to one another and to other disciplines. The teacher demonstrates understanding of prerequisite relationships among topics and concepts and understands the link to necessary cognitive structures that ensure student understanding. The teacher's plans and practice reflect familiarity with a wide range of effective pedagogical approaches in the discipline and the ability to anticipate student misconceptions.
1B. Demonstrating Knowledge of Students	The teacher displays minimal understanding of how students learn—and little knowledge of their varied approaches to learning, knowledge and skills, special needs, and interests and cultural heritages—and does not indicate that such knowledge is valuable.	The teacher displays generally accurate knowledge of how students learn and of their varied approaches to learning, knowledge and skills, special needs, and interests and cultural heritages, yet may apply this knowledge not to individual students but to the class as a whole.	The teacher understands the active nature of student learning and attains information about levels of development for groups of students. The teacher also purposefully acquires knowledge from several sources about groups of students' varied approaches to learning, knowledge and skills, special needs, and interests and cultural heritages.	The teacher understands the active nature of student learning and acquires information about levels of development for individual students. The teacher also systematically acquires knowledge from several sources about individual students' varied approaches to learning, knowledge and skills, special needs, and interests and cultural heritages.
1C. Setting Instructional Outcomes	The outcomes represent low expectations for students and lack of rigor, and not all of these outcomes reflect important learning in the discipline. They are stated as student activities, rather than as outcomes for learning. Outcomes reflect only one type of learning and only one discipline or strand and are suitable for only some students.	Outcomes represent moderately high expectations and rigor. Some reflect important learning in the discipline and consist of a combination of outcomes and activities. Outcomes reflect several types of learning, but the teacher has made no effort at coordination or integration. Outcomes, based on global assessments of student learning, are suitable for most of the students in the class.	Most outcomes represent rigorous and important learning in the discipline and are clear, are written in the form of student learning, and suggest viable methods of assessment. Outcomes reflect several different types of learning and opportunities for coordination, and they are differentiated, in whatever way is needed, for different groups of students.	All outcomes represent high-level learning in the discipline. They are clear, are written in the form of student learning, and permit viable methods of assessment. Outcomes reflect several different types of learning and, where appropriate, represent both coordination and integration. Outcomes are differentiated, in whatever way is needed, for individual students.

DOMAIN 1: Planning and Preparation, continued

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
1D. Demonstrating Knowledge of Resources	The teacher is unaware of resources to assist student learning beyond materials provided by the school or district, nor is the teacher aware of resources for expanding one's own professional skill.	The teacher displays some awareness of resources beyond those provided by the school or district for classroom use and for extending one's professional skill but does not seek to expand this knowledge.	The teacher displays awareness of resources beyond those provided by the school or district, including those on the Internet, for classroom use and for extending one's professional skill, and seeks out such resources.	The teacher's knowledge of resources for classroom use and for extending one's professional skill is extensive, including those available through the school or district, in the community, through professional organizations and universities, and on the Internet.
1E. Designing Coherent Instruction	Learning activities are poorly aligned with the instructional outcomes, do not follow an organized progression, are not designed to engage students in active intellectual activity, and have unrealistic time allocations. Instructional groups are not suitable to the activities and offer no variety.	Some of the learning activities and materials are aligned with the instructional outcomes and represent moderate cognitive challenge, but with no differentiation for different students. Instructional groups partially support the activities, with some variety. The lesson or unit has a recognizable structure; but the progression of activities is uneven, with only some reasonable time allocations.	Most of the learning activities are aligned with the instructional outcomes and follow an organized progression suitable to groups of students. The learning activities have reasonable time allocations; they represent significant cognitive challenge, with some differentiation for different groups of students and varied use of instructional groups.	The sequence of learning activities follows a coherent sequence, is aligned to instructional goals, and is designed to engage students in high-level cognitive activity. These are appropriately differentiated for individual learners. Instructional groups are varied appropriately, with some opportunity for student choice.
1F. Designing Student Assessments	Assessment procedures are not congruent with instructional outcomes and lack criteria by which student performance will be assessed. The teacher has no plan to incorporate formative assessment in the lesson or unit.	Assessment procedures are partially congruent with instructional outcomes. Assessment criteria and standards have been developed, but they are not clear. The teacher's approach to using formative assessment is rudimentary, including only some of the instructional outcomes.	All the instructional outcomes may be assessed by the proposed assessment plan; assessment methodologies may have been adapted for groups of students. Assessment criteria and standards are clear. The teacher has a well-developed strategy for using formative assessment and has designed particular approaches to be used.	All the instructional outcomes may be assessed by the proposed assessment plan, with clear criteria for assessing student work. The plan contains evidence of student contribution to its development. Assessment methodologies have been adapted for individual students as the need has arisen. The approach to using formative assessment is well designed and includes student as well as teacher use of the assessment information.

DOMAIN 2: The Classroom Environment

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
2A. Creating an Environment of Respect and Rapport	Patterns of classroom interactions, both between teacher and students and among students, are mostly negative, inappropriate, or insensitive to students' ages, cultural backgrounds, and developmental levels. Student interactions are characterized by sarcasm, put-downs, or conflict. The teacher does not deal with disrespectful behavior.	Patterns of classroom interactions, both between teacher and students and among students, are generally appropriate but may reflect occasional inconsistencies, favoritism, and disregard for students' ages, cultures, and developmental levels. Students rarely demonstrate disrespect for one another. The teacher attempts to respond to disrespectful behavior, with uneven results. The net result of the interactions is neutral, conveying neither warmth nor conflict.	Teacher-student interactions are friendly and demonstrate general caring and respect. Such interactions are appropriate to the ages, cultures, and developmental levels of the students. Interactions among students are generally polite and respectful, and students exhibit respect for the teacher. The teacher responds successfully to disrespectful behavior among students. The net result of the interactions is polite, respectful, and businesslike, though students may be somewhat cautious about taking intellectual risks.	Classroom interactions between teacher and students and among students are highly respectful, reflecting genuine warmth, caring, and sensitivity to students as individuals. Students exhibit respect for the teacher and contribute to high levels of civility among all members of the class. The net result is an environment where all students feel valued and are comfortable taking intellectual risks.
2B. Establishing a Culture for Learning	The classroom culture is characterized by a lack of teacher or student commitment to learning, and/or little or no investment of student energy in the task at hand. Hard work and the precise use of language are not expected or valued. Medium to low expectations for student achievement are the norm, with high expectations for learning reserved for only one or two students.	The classroom culture is characterized by little commitment to learning by the teacher or students. The teacher appears to be only "going through the motions," and students indicate that they are interested in the completion of a task rather than the quality of the work. The teacher conveys that student success is the result of natural ability rather than hard work, and refers only in passing to the precise use of language. High expectations for learning are reserved for those students thought to have a natural aptitude for the subject.	The classroom culture is a place where learning is valued by all; high expectations for both learning and hard work are the norm for most students. Students understand their role as learners and consistently expend effort to learn. Classroom interactions support learning, hard work, and the precise use of language.	The classroom culture is a cognitively busy place, characterized by a shared belief in the importance of learning. The teacher conveys high expectations for learning for all students and insists on hard work; students assume responsibility for high quality by initiating improvements, making revisions, adding detail, and/or assisting peers in their precise use of language.

DOMAIN 2: The Classroom Environment, continued

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
2C. Managing Classroom Procedures	Much instructional time is lost due to inefficient classroom routines and procedures. There is little or no evidence of the teacher's managing instructional groups and transitions and/or handling of materials and supplies effectively. There is little evidence that students know or follow established routines.	Some instructional time is lost due to partially effective classroom routines and procedures. The teacher's management of instructional groups and transitions, or handling of materials and supplies, or both, are inconsistent, leading to some disruption of learning. With regular guidance and prompting, students follow established routines.	There is little loss of instructional time due to effective classroom routines and procedures. The teacher's management of instructional groups and transitions, or handling of materials and supplies, or both, are consistently successful. With minimal guidance and prompting, students follow established classroom routines.	Instructional time is maximized due to efficient and seamless classroom routines and procedures. Students take initiative in the management of instructional groups and transitions, and/or the handling of materials and supplies. Routines are well understood and may be initiated by students.
2D. Managing Student Behavior	There appear to be no established standards of conduct, or students challenge them. There is little or no teacher monitoring of student behavior, and response to students' misbehavior is repressive or disrespectful of student dignity.	Standards of conduct appear to have been established, but their implementation is inconsistent. The teacher tries, with uneven results, to monitor student behavior and respond to student misbehavior.	Student behavior is generally appropriate. The teacher monitors student behavior against established standards of conduct. Teacher response to student misbehavior is consistent, proportionate, and respectful to students and is effective.	Student behavior is entirely appropriate. Students take an active role in monitoring their own behavior and/or that of other students against standards of conduct. Teacher monitoring of student behavior is subtle and preventive. The teacher's response to student misbehavior is sensitive to individual student needs and respects students' dignity.
2E. Organizing Physical Space	The classroom environment is unsafe, or learning is not accessible to many. There is poor alignment between the arrangement of furniture and resources, including computer technology, and the lesson activities.	The classroom is safe, and essential learning is accessible to most students. The teacher makes modest use of physical resources, including computer technology. The teacher attempts to adjust the classroom furniture for a lesson or, if necessary, to adjust the lesson to the furniture, but with limited effectiveness.	The classroom is safe, and students have equal access to learning activities; the teacher ensures that the furniture arrangement is appropriate to the learning activities and uses physical resources, including computer technology, effectively.	The classroom environment is safe, and learning is accessible to all students, including those with special needs. The teacher makes effective use of physical resources, including computer technology. The teacher ensures that the physical arrangement is appropriate to the learning activities. Students contribute to the use or adaptation of the physical environment to advance learning.

DOMAIN 3: Instruction

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
3A. Communicating Clearly and Accurately	The instructional purpose of the lesson is unclear to students, and the directions and procedures are confusing. The teacher's explanation of the content contains major errors and does not include any explanation of strategies students might use. The teacher's spoken or written language contains errors of grammar or syntax. The teacher's academic vocabulary is inappropriate, vague, or used incorrectly, leaving students confused.	The teacher's attempt to explain the instructional purpose has only limited success, and/or directions and procedures must be clarified after initial student confusion. The teacher's explanation of the content may contain minor errors; some portions are clear, others difficult to follow. The teacher's explanation does not invite students to engage intellectually or to understand strategies they might use when working independently. The teacher's spoken language is correct but uses vocabulary that is either limited or not fully appropriate to the students' ages or backgrounds. The teacher rarely takes opportunities to explain academic vocabulary.	The instructional purpose of the lesson is clearly communicated to students, including where it is situated within broader learning; directions and procedures are explained clearly and may be modeled. The teacher's explanation of content is scaffolded, clear, and accurate and connects with students' knowledge and experience. During the explanation of content, the teacher focuses, as appropriate, on strategies students can use when working independently and invites student intellectual engagement. The teacher's spoken and written language is clear and correct and is suitable to students' ages and interests. The teacher's use of academic vocabulary is precise and serves to extend student understanding.	The teacher links the instructional purpose of the lesson to the larger curriculum; the directions and procedures are clear and anticipate possible student misunderstanding. The teacher's explanation of content is thorough and clear, developing conceptual understanding through clear scaffolding and connecting with students' interests. Students contribute to extending the content by explaining concepts to their classmates and suggesting strategies that might be used. The teacher's spoken and written language is expressive, and the teacher finds opportunities to extend students' vocabularies, both within the discipline and for more general use. Students contribute to the correct use of academic vocabulary.
3B. Using Questioning and Discussion Techniques	The teacher's questions are of low cognitive challenge, with single correct responses, and are asked in rapid succession. Interaction between the teacher and students is predominantly recitation style, with the teacher mediating all questions and answers; the teacher accepts all contributions without asking students to explain their reasoning. Only a few students participate in the discussion.	The teacher's questions lead students through a single path of inquiry, with answers seemingly determined in advance. Alternatively, the teacher attempts to ask some questions designed to engage students in thinking, but only a few students are involved. The teacher attempts to engage all students in the discussion, to encourage them to respond to one another, and to explain their thinking, with uneven results.	While the teacher may use some low-level questions, he poses questions designed to promote student thinking and understanding. The teacher creates a genuine discussion among students, providing adequate time for students to respond and stepping aside when doing so is appropriate. The teacher challenges students to justify their thinking and successfully engages most students in the discussion, employing a range of strategies to ensure that most students are heard.	The teacher uses a variety or series of questions or prompts to challenge students cognitively, advance high-level thinking and discourse, and promote metacognition. Students formulate many questions, initiate topics, challenge one another's thinking, and make unsolicited contributions. Students themselves ensure that all voices are heard in the discussion.

DOMAIN 3: Instruction, continued

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
3C. Engaging Students in Learning	The learning tasks/activities, materials, and resources are poorly aligned with the instructional outcomes, or require only rote responses, with only one approach possible. The groupings of students are unsuitable to the activities. The lesson has no clearly defined structure, or the pace of the lesson is too slow or rushed.	The learning tasks and activities are partially aligned with the instructional outcomes but require only minimal thinking by students and little opportunity for them to explain their thinking, allowing most students to be passive or merely compliant. The groupings of students are moderately suitable to the activities. The lesson has a recognizable structure; however, the pacing of the lesson may not provide students the time needed to be intellectually engaged or may be so slow that many students have a considerable amount of "downtime."	The learning tasks and activities are fully aligned with the instructional outcomes and are designed to challenge student thinking, inviting students to make their thinking visible. This technique results in active intellectual engagement by most students with important and challenging content and with teacher scaffolding to support that engagement. The groupings of students are suitable to the activities. The lesson has a clearly defined structure, and the pacing of the lesson is appropriate, providing most students the time needed to be intellectually engaged.	Virtually all students are intellectually engaged in challenging content through well-designed learning tasks and activities that require complex thinking by students. The teacher provides suitable scaffolding and challenges students to explain their thinking. There is evidence of some student initiation of inquiry and student contributions to the exploration of important content; students may serve as resources for one another. The lesson has a clearly defined structure, and the pacing of the lesson provides students the time needed not only to intellectually engage with and reflect upon their learning but also to consolidate their understanding.
3D. Using Feedback and Assessment	Students do not appear to be aware of the assessment criteria, and there is little or no monitoring of student learning; feedback is absent or of poor quality. Students do not engage in self or peer assessment.	Students appear to be only partially aware of the assessment criteria, and the teacher monitors student learning for the class as a whole. Questions and assessments are rarely used to diagnose evidence of learning. Feedback to students is general, and few students assess their own work.	Students appear to be aware of the assessment criteria, and the teacher monitors student learning for groups of students. Questions and assessments are regularly used to diagnose evidence of learning. Teacher feedback to groups of students is accurate and specific; some students engage in self-assessment.	Assessment is fully integrated into instruction, through extensive use of formative assessment. Students appear to be aware of, and there is some evidence that they have contributed to the assessment criteria. Questions and assessments are used regularly to diagnose evidence of learning by individual students. A variety of forms of feedback, from both teacher and peers, is accurate and specific and advances learning. Students self-assess and monitor their own progress. The teacher successfully differentiates instruction to address individual students' misunderstandings.

DOMAIN 3: Instruction, continued

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
3E. Demonstrating Flexibility and Responsiveness	The teacher ignores students' questions; when students have difficulty learning, the teacher blames them or their home environment for their lack of success. The teacher makes no attempt to adjust the lesson even when students don't understand the content.	The teacher accepts responsibility for the success of all students but has only a limited repertoire of strategies to use. Adjustment of the lesson in response to assessment is minimal or ineffective.	The teacher successfully accommodates students' questions and interests. Drawing on a broad repertoire of strategies, the teacher persists in seeking approaches for students who have difficulty learning. If impromptu measures are needed, the teacher makes a minor adjustment to the lesson and does so smoothly.	The teacher seizes an opportunity to enhance learning, building on a spontaneous event or students' interests, or successfully adjusts and differentiates instruction to address individual student misunderstandings. Using an extensive repertoire of instructional strategies and soliciting additional resources from the school or community, the teacher persists in seeking effective approaches for students who need help.

DOMAIN 4: Professional Responsibilities

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
4A. Reflecting on Teaching	The teacher does not know whether a lesson was effective or achieved its instructional outcomes, or the teacher profoundly misjudges the success of a lesson. The teacher has no suggestions for how a lesson could be improved.	The teacher has a generally accurate impression of a lesson's effectiveness and the extent to which instructional outcomes were met. The teacher makes general suggestions about how a lesson could be improved.	The teacher makes an accurate assessment of a lesson's effectiveness and the extent to which it achieved its instructional outcomes and can cite general references to support the judgment. The teacher makes a few specific suggestions of what could be tried another time the lesson is taught.	The teacher makes a thoughtful and accurate assessment of a lesson's effectiveness and the extent to which it achieved its instructional outcomes, citing many specific examples from the lesson and weighing the relative strengths of each. Drawing on an extensive repertoire of skills, the teacher offers specific alternative actions, complete with the probable success of different courses of action.
4B. Maintaining Accurate Records	The teacher's system for maintaining information on student completion of assignments and student progress in learning is nonexistent or in disarray. The teacher's records for non-instructional activities are in disarray, the result being errors and confusion.	The teacher's system for maintaining information on student completion of assignments and student progress in learning is rudimentary and only partially effective. The teacher's records for non-instructional activities are adequate but inefficient and, unless given frequent oversight by the teacher, prone to errors.	The teacher's system for maintaining information on student completion of assignments, student progress in learning, and non-instructional records is fully effective.	The teacher's system for maintaining information on student completion of assignments, student progress in learning, and non-instructional records is fully effective. Students contribute information and participate in maintaining the records.

DOMAIN 4: Professional Responsibilities, continued

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
4C. Communicating with Families	The teacher provides little information about the instructional program to families; the teacher's communication about students' progress is minimal. The teacher does not respond, or responds insensitively, to parental concerns.	The teacher makes sporadic attempts to communicate with families about the instructional program and about the progress of individual students but does not attempt to engage families in the instructional program. Moreover, the communication that does take place may not be culturally sensitive to those families.	The teacher provides frequent and appropriate information to families about the instructional program and conveys information about individual student progress in a culturally sensitive manner. The teacher makes some attempts to engage families in the instructional program.	The teacher communicates frequently with families in a culturally sensitive manner, with students contributing to the communication. The teacher responds to family concerns with professional and cultural sensitivity. The teacher's efforts to engage families in the instructional program are frequent and successful.
4D. Participating in the Professional Community	The teacher's relationships with colleagues are negative or self-serving. The teacher avoids participation in a professional culture of inquiry, resisting opportunities to become involved. The teacher avoids becoming involved in school events or school and district projects.	The teacher maintains cordial relationships with colleagues to fulfill duties that the school or district requires. The teacher participates in the school's culture of professional inquiry when invited to do so. The teacher participates in school events and school and district projects when specifically asked.	The teacher's relationships with colleagues are characterized by mutual support and cooperation; the teacher actively participates in a culture of professional inquiry. The teacher volunteers to participate in school events and in school and district projects, making a substantial contribution.	The teacher's relationships with colleagues are characterized by mutual support and cooperation, with the teacher taking initiative in assuming leadership among the faculty. The teacher takes a leadership role in promoting a culture of professional inquiry. The teacher volunteers to participate in school events and district projects, making a substantial contribution and assuming a leadership role in at least one aspect of school or district life.
4E. Growing and Developing Professionally	The teacher engages in no professional development activities to enhance knowledge or skill. The teacher resists feedback on teaching performance from either supervisors or more experienced colleagues. The teacher makes no effort to share knowledge with others or to assume professional responsibilities.	The teacher participates to a limited extent in professional activities when they are convenient. The teacher engages in a limited way with colleagues and supervisors in professional conversation about practice, including some feedback on teaching performance. The teacher finds limited ways to assist other teachers and contribute to the profession.	The teacher seeks out opportunities for professional development to enhance content knowledge and pedagogical skill. The teacher actively engages with colleagues and supervisors in professional conversation about practice, including feedback about practice. The teacher participates actively in assisting other educators and looks for ways to contribute to the profession.	The teacher seeks out opportunities for professional development and makes a systematic effort to conduct action research. The teacher solicits feedback on practice from both supervisors and colleagues. The teacher initiates important activities to contribute to the profession.

DOMAIN 4: Professional Responsibilities, continued

LEVEL OF PERFORMANCE

Component	Unsatisfactory	Basic	Proficient	Distinguished
4F. Showing Professionalism	The teacher displays dishonesty in interactions with colleagues, students, and the public. The teacher is not alert to students' needs and contributes to school practices that result in some students being ill served by the school. The teacher makes decisions and recommendations that are based on self-serving interests. The teacher does not comply with school and district regulations.	The teacher is honest in interactions with colleagues, students, and the public. The teacher's attempts to serve students are inconsistent, and unknowingly contribute to some students being ill served by the school. The teacher's decisions and recommendations are based on limited though genuinely professional considerations. The teacher must be reminded by supervisors about complying with school and district regulations.	The teacher displays high standards of honesty, integrity, and confidentiality in interactions with colleagues, students, and the public. The teacher is active in serving students, working to ensure that all students receive a fair opportunity to succeed. The teacher maintains an open mind in team or departmental decision making. The teacher complies fully with school and district regulations.	The teacher can be counted on to hold the highest standards of honesty, integrity, and confidentiality and takes a leadership role with colleagues. The teacher is highly proactive in serving students, seeking out resources when needed. The teacher makes a concerted effort to challenge negative attitudes or practices to ensure that all students, particularly those traditionally underserved, are honored in the school. The teacher takes a leadership role in team or departmental decision making and helps ensure that such decisions are based on the highest professional standards. The teacher complies fully with school and district regulations, taking a leadership role with colleagues.