Student Perception Surveys as a Component of a Comprehensive Teacher Evaluation System: A

Case Study of Elementary Principals

Matthew M. Kaiser

Accepted by the School of Education Faculty, Indiana University, in partial fulfillment of the requirements for the Degree of Doctor of Education

Doctoral Committee:

_____

Chad Lochmiller, Ph.D.

_____

Patricia Kubow, Ph.D.

_____

Christopher Lubienski, Ph.D.

Date of Defense
April 19, 2023

To Chad, thank you for your support, understanding, guidance, and wisdom.

To the members of my committee, thank you for your patience.

To Kathy, The Colonel, Barb, and Frank thank you for your unwavering faith and support.

To Emily, thank you for expecting me to be who I should be, not who I am.

To Will and Lucy, thank you for making me work harder based on all you have accomplished.

To COVID-19, bye Felicia.

Student Perception Surveys as a Component of a Comprehensive Teacher Evaluation

System: A Case Study of Elementary Principals

While multiple states have implemented a multiple measures approach to teacher evaluation, this approach has not significantly improved the ability of evaluations to identify variations in teacher performance. The purpose of this research study was to understand what elementary school principals in one Midwestern school district believe are the relative strengths, weaknesses, and potential affordances of using student perception data in the context of the teacher evaluation system in their district. To understand teaching evaluations and the impact of satisfactory performance ratings, the study examined the perspective of school principals working in a district that considers the integration of student perspectives in its teacher evaluation system. A qualitative interview study was conducted to gather information through 10 semi-structured interviews with school personnel, including school principals and district central office administrators. District artifacts were also analyzed. The study shows that the participants perceived that student perception data could be useful in informing teacher evaluations and may influence administrative practices related to the evaluation of teacher quality. However, the participants argued that the benefits of these data depend on the implementation of the system and the conditions for its use that were established by the school district.

_____

Chad Lochmiller, Ph.D.

_____

Patricia, Kubow, Ph.D.

_____

Christopher Lubienski, Ph.D.

Table of Contents

Tables   **Page**

Curriculum Vitae

Chapter 1

Introduction

Teacher evaluation has received considerable attention in school districts throughout the United States since the passage and subsequent implementation of the No Child Left Behind Act of 2001. Between 2000 and 2018, multiple states responded to the changing federal expectations by enacting legislation that restructured teacher evaluation systems. In 2009, Race to the Top (RTT) used a series of incentives to encourage public education agencies to use rigorous teacher evaluations to measure teacher effectiveness on an annual basis (U.S. Department of Education [USDE], 2009). To incentivize new teacher evaluation practices, the USDE offered waivers to states that allowed recipients exemptions from certain accountability provisions if the state agreed to implement more rigorous teacher evaluation systems.

These incentives spurred many state legislatures to pass new state laws that required annual teacher evaluations that included measures of student achievement. The most common measures included student achievement or growth as measured by student performance on state standardized tests aligned with the state's accountability framework. In 2009, only 15 states required teacher evaluations to include student growth as a factor. By 2015, the number had increased to 43 states, which suggests that the waiver application process motivated the states to adopt new evaluation practices (Aragon, 2018).

After the passage of the Every Student Succeeds Act (ESSA) of 2015, the use of state standardized tests as a primary component of teachers' evaluations has started tapering off. The transition to the ESSA has potentially accelerated this process by removing federal incentives that previously weighted standardized test scores as a significant component of teacher evaluation (Brown et al., 2016). Instead, states appear increasingly interested in the use of

multiple measures of student performance. In an analysis of 158 bills proposed in 33 states from January 2015 to May 2018, Croft et al. (2018) found that 41 bills nationwide removed or reduced the use of student achievement or growth as a measure of teacher performance. In this study, a focus on teachers was considered vital, as assessments are an essential component to many educators and their yearly evaluation. However, many administrators have shared that assessments and previous assessment data are not good measures to evaluate the effectiveness of educators (Reid, 2017).

The changing policy landscape has compelled school districts to revise their existing evaluation systems and adopt new, multi-level measures to assess teacher performance. This shift has raised important questions about what measures should be used and with what level of emphasis. Classroom observations, principal evaluations, instructional artifacts, portfolios, teacher self-report measures, value-added models, and student perception surveys are all potential sources of valid information for evaluating teacher effectiveness (Goe et al., 2008). Each measure also comes with inherent flaws; thus, relying solely on one measure is an ill-advised approach. For example, classroom observations represent a fraction of the teacher's actual time with students. Likewise, instructional artifacts and portfolios may only show a small sample of the best students or be skewed to reflect the teacher's best lessons (Little, 2009). Teachers' self-reported measures can reflect their biases regarding their individual performances (Temam et al., 2019). Value-added models may be influenced by factors that are well outside the performance of the teacher in the classroom (Rothstein, 2016). Finally, student perception surveys may reflect students' personal feelings about their teachers rather than their instructions (Wallace et al., 2016).

Changes in evaluation practices also align with calls from researchers, many of whom have begun to advocate for multiple measures and a more robust approach to teacher evaluation. In the fall of 2009, the Bill and Melinda Gates Foundation launched the Measures of Effective Teaching (MET) project to develop new approaches to measuring effective teaching. The study presumed that teacher evaluation should include student achievement gains, that evaluation components should be related to these gains, and that evaluation feedback should target specific aspects of teachers' practices, which supports their growth and development. The measures used in the study included student achievement gains on different assessments, classroom observations and teacher reflections, teachers' pedagogical content knowledge, student perceptions of the classroom instructional environment, and teachers' perceptions of working conditions and instructional support in their schools. In discussing the need for this study, the researchers made the following conclusion:

> The public debate over measuring teacher effectiveness usually portrays two options: the status quo (where there is no meaningful feedback for teachers) and a seemingly extreme world where test scores alone determine a teacher's fate. Our results suggest that's a false choice. (MET, 2012)

Perceptions, observations, and reflections were an important part of the essential feedback for teachers. The researchers concluded that either of these extremes was unlikely to provide an accurate picture of teacher effectiveness. A major conclusion from this exploration of teacher evaluation is that neither current systems nor systems relying solely on student test scores are effective in evaluating teaching performance. This study specifically recommended the use of at least three data points to evaluate effective teaching. Accurate value-added measure data from assessments should be gathered from the teacher's students, confidential student surveys should

be collected to gain feedback on the students' perceptions of their teacher's practices, and meaningful classroom observations should be conducted by school administrators (MET, 2012). Martínez et al. (2016) supported the use of this multiple measures approach, which has given new life to the idea of utilizing student perceptions of teacher performance as part of teacher evaluation systems. Many districts have begun to explore the value of a multiple measures approach as one component of their evaluation system (Stosich et al., 2018).

Statement of the Problem

While many states have implemented a multiple measures approach to teacher evaluation, this approach has not significantly improved the ability of evaluations to identify variations in teacher performance (Kraft & Gilmour, 2017). Despite the increasing recognition that teacher effectiveness contributes to student achievement outcomes, scholars have concluded that as many as 98% of teachers receive satisfactory performance ratings (Weisberg et al., 2009). The No Child Left Behind Act of 2001 emphasized the need for highly qualified educators, which created a demand for hiring teachers with impressive evaluations (Brownell et al., 2018). The failure to differentiate teacher quality substantially has been documented in several research studies (Donaldson, 2009; Toch & Rothman, 2008; Wechsler et al., 2007). Some scholars have concluded that "to date, there exists little systematic evidence about the degree to which teacher evaluation reforms have fundamentally changed the distribution of teacher performance ratings" (Kraft & Gilmour, 2017, p. 4). Given this context, the purpose of this research study was to understand what elementary school principals in one Midwestern school district in the United States believe are the relative strengths, weaknesses, and potential opportunities of using student perception in the context of the teacher evaluation system in their district.

4

Research Questions

To understand educator evaluations and the impact of satisfactory performance ratings, this study addressed the following research questions:

1. What do principals feel are the strengths, weaknesses, and potential opportunities of student perceptions as part of their evaluation ratings for classroom teachers under their direct supervision?

2. From principals' perspective, how, if at all, will the integration of student perspectives in teacher evaluations affect an administrative approach to instructional leadership and supervision?

    a. What, if anything, do principals suggest school districts should incorporate in student perception surveys as a component of their comprehensive teacher evaluation system?

Limitations

All studies have limitations that a researcher must address. A qualitative study design, which favors a small sample size, can limit the generalizability of the study. Vasileiou et al. (2018) noted that qualitative study samples are typically small to support the depth of details that a case-oriented analysis provides. Although small sample sizes in qualitative research limit the number of participants, a generalization of information is not the goal of the methodology. As the goal of qualitative research is to explore perspectives for understanding, the detailed information from a small sample size is appropriate for qualitative studies. In addition, as this study focused on the perceptions of principals at the elementary level, this may specifically limit its generalizability to the perceptions of secondary-level school administrators. Also, the administrators who participated in this study all work within non-departmentalized elementary

schools. This may limit the ability to apply the findings of this study to elementary school

settings that engage in departmentalization, which has become a particularly common practice

given the pressures for standardized testing.

Chapter 2

Literature Review

This chapter presents a literature review that supports the focus and purpose of the study. For this study, academic databases and journals were searched and examined to identify and understand the subject of teacher evaluations, the theoretical framework that would support this study, and how previous research has approached this educational concern. This chapter reviews the empirical and theoretical ideas that provide the foundation for the study. Specifically, the literature review was conducted to explore teacher quality as an important factor in student achievement and how school districts are becoming increasingly strategic in their evaluation of teacher quality. Moreover, this section examined the critical role that principals play in the evaluation of teacher quality and the multiple methods they may be employed in evaluating teachers. Finally, it examined the current research surrounding the relatively new practice of including student perspectives as a component of annual teacher evaluations.

Strategic Human Capital Management and Teacher Quality

This study was based on the assumption that organizations such as school districts can both manage and develop the quality of their workforce through the development and use of a coherent human capital management strategy. Wright and McMahan (1992) defined strategic human capital management as "the pattern of planned human resource deployments and activities intended to enable an organization to achieve its goals" (p. 298). Nordhaug and Gronhaug (1992) observed that the knowledge, skills, and values held by individuals in organizations form a "portfolio" of competencies that help organizations achieve their objectives. Dunford et al. (2001) referred to these competencies as stocks of skills and strategically relevant knowledge and behaviors. Organizations can achieve their goals and perform better when they

maximize their ability to blend the various competencies of their employees and direct those competencies toward the aims and objectives of the organization.

In education, researchers such as Hargreaves and Braun (2013) have shared that data are used more for accountability than for improving instruction. However, Custer et al. (2018) indicated that high-quality assessment is an essential tool for evaluating learning strategies and programs and can measure effectiveness. Hanushek and Woessmann (2020) explained that "ignoring differences in the quality of education significantly distorts the picture of how educational and economic outcomes are related" (p. 171). On the basis of recent research studies, Hanushek and Woessmann (2020) concluded that quality learning and the knowledge of students are essential to economic growth. They observed that educational policies focus more on the quality of schools than on the number of schools. Thus, in education, many efforts to manage human capital focus on strategies aimed at improving the quality of teaching and classroom teachers.

A study by Dudek et al. (2019) evaluated the effectiveness of the Classroom Strategies Assessment System (CSAS)-Observer form, an educator assessment tool used by administrators. In their study, Dudek et al. sampled 35 teachers and 829 third- through eighth-grade students. The participants were from six different high-poverty charter schools in New Jersey. Using the CSAS observation tool, six administrators conducted three observations for each teacher. Dudek et al. (2019) found evidence of the usefulness of measuring teacher practices to inform student achievements in high-poverty settings. Specifically, Dudek et al. noted that assessment-based feedback from the administration that supports identifying practice strengths and areas for improvement promotes effective teaching, which impacts student achievement.

School Principals as Primary Evaluators

Administrators carry out important human capital management tasks; in particular, they supervise classroom instructions and evaluate teachers, who provide instructions to students (Rousmaniere, 2007). These responsibilities have taken on new importance, given the pressing need for standards and accountability. In the early 2000s, the passage of the No Child Left Behind Act (NCLB, 2001) further expanded this emphasis on the role of the principal in school success. The NCLB Act led to the passage of the American Recovery and Reinvestment Act (Civic Impulse, 2018), which led to RTT grants. The RTT grant program was designed to encourage administrative recipients to improve teacher effectiveness through evaluation measures.

Changes in state laws sparked by the NCLB Act often specifically singled out the principal's role in school accountability, many of which have identified the removal and replacement of school principals as a step in the required reform of schools designated as failing. According to Crow (2006), this type of scrutiny "has added to the complexity of the principal's job, requiring principals to be entrepreneurial, to be more focused on student outcomes and instructional processes, and to be more connected with their communities" (p. 316). In a 2010 survey, school district administrators, policy advisors, and other educators named "principal leadership" as second only to teacher quality when asked to rank 21 education issues ranging from special education to dropout rates and school violence (Simkin et al., 2010).

In light of the shifting accountability environment, the role of the principal has gradually become focused on instructional leadership, within which instruction supervision and teacher evaluation are key (Hallinger et al., 2020). With this transition to instructional leadership, principals are expected to effectively evaluate teachers and improve instructional quality. In

exploring the role of principals, Cosner et. al (2015) found that "school principals have traditionally had, and will in most cases continue to have, primary responsibility for evaluating the 3.7 million public school teachers nationwide" (p. 76)). Firestone and Donaldson (2019) examined how new systems of teacher evaluations might facilitate the improvement of educators' teaching performances. Firestone and Donaldson also shared that many teachers and administrators who were using new observational tools reported that the process produced information that teachers could use to enhance their lessons. In their instructional leadership role, principals are responsible for providing this type of feedback.

Principals routinely conduct observation and feedback cycles as part of new evaluation systems (Herlihy et al., 2014). In many cases, principals are given full responsibility for determining teachers' overall summative evaluations. Derrington (2016) explained that principals must navigate various state and federal regulations and policies while providing leadership in the education process for teachers to incorporate into their classrooms. Derrington also shared that formal leaders such as principals facilitate learning and teaching through the influence of practical instructional applications.

As instructional leaders, principals are important actors in the teacher evaluation process. Their perspective is helpful in considering how changes in evaluation might improve the benefits of evaluation processes. As critical actors in the teacher evaluation process, school principals are often the person most solely responsible for the administration of student perception surveys, the analysis of survey results, and the inclusion of these surveys as part of teacher evaluation. On the basis of their substantial influence and central role in the creation, implementation, and/or analysis of these surveys, understanding principals' beliefs regarding student perception surveys can provide rich information for school districts seeking to explore their use.

Research Exploring the Links Between Teacher Quality and Student Achievement

A significant body of research has examined the links between teacher quality and student achievement. This empirical link has elevated the importance of the practice of teacher evaluation and its potential as a lever for educational improvement. Research has widely reported that teachers contribute to and exercise a direct influence on student achievement (Brown et al., 2016; Brownell et al., 2018; Darling- Hammond, 2000; Derrington, 2016). Whitehurst (2002) evaluated the performance of math teachers in grades three, four, and five in two urban school districts in Tennessee and calculated the average amount of academic growth of students in their classrooms. From these data, teachers were identified and grouped into those who were most effective (the top 20%) and those who were least effective (the bottom 20%). The progress of the students assigned to both the most effective and least effective teachers was tracked over 3 consecutive years. The results revealed that at the end of the fifth grade, math students assigned to the high-performing teachers scored in the 83rd percentile, whereas those assigned to the low-performing teachers scored in the 29th percentile (Whitehurst, 2002). Similar effects on student performance were found in a study by Fischer et al. (2018), who reported that teachers' professional development participation is positively associated with classroom practice. Their findings indicated that changes in instruction (based on evaluation feedback) align with challenging curricular goals that improve student performance.

Research studies have indicated that teacher quality more strongly influences differences in student performance than does the race, class, or school of the student (Kraft & Gilmour, 2017; Nye et al., 2004, Sanders et al., 1997). Researchers have also found that achievement gains from having an effective teacher could be almost three times greater for black students than for white students, even when comparing students with the same prior school achievement

11

(Sanders & Rivers, 1996; Stronge, 2018). While 1 year of effective teaching can have a substantial effect, the effects of teacher quality accumulate over time (Polgampa & Huang, 2017). Fifth-grade math students in Tennessee who had three consecutive highly effective teachers scored between 52 and 54 percentile points better than students who had three consecutive teachers who were less effective, even though both groups had the same achievement rates prior to entering second grade (Sanders & Rivers, 1996). A similar study conducted in Texas showed a difference of 34 percentile points in reading and 49 percentile points in math (Jordan et al., 1997). All these research findings indicate a relationship between student success and the quality of learning received from effective teachers.

Research Exploring Teacher Evaluation Methods

While principals are interested in identifying and developing effective teachers, multiple theories present different data sources that principals might use to determine how effective teachers are in the classroom. The common approaches to evaluating teacher quality include principal evaluation of the teacher using classroom observations, instructional artifacts, and portfolios, as well as teacher self-measured reports. In addition, measures include student performance in standardized assessments and student perceptions. Many researchers have advocated for evaluation systems that incorporate a combination of measures, although little consensus exists regarding their relative weight (Geiger & Amrein-Beardsley, 2019; MET, 2012; Reid, 2017; Stosich et al., 2018). Each method of identifying effective teachers has relative strengths and weaknesses in determining their effectiveness. These methods are listed in the following sections. Appendix B contains a chart that briefly summarizes the evaluation components.

Classroom and Administrator Observations

Classroom observations are one of the most common forms of teacher evaluation (Brandt et al., 2007). Although commonly conducted by principals or vice-principals, teacher observations can also be assigned to peer teachers or other designated professional educators. Teacher observations can take many forms and measure multiple aspects of the teaching process. However, observations often have a high degree of variation in how they are implemented. In some cases, school districts develop observation systems by using purchased products, such as those developed by Robert Marzano or Charlotte Danielson. Districts also develop their evaluation criteria for classroom observations. In some instances, observations occur somewhere between once and a few times during the school year.

Some observations may encompass only a single lesson, which might occur at a time agreed upon by the teacher and evaluator conducting the observations (van der Lans et al., 2016). One benefit of classroom observations is that they allow administrators to provide formative and summative feedback. Researchers have linked classroom observations to increases in student achievement (Gallagher, 2004; Milanowski, 2004). In addition, in their study with teachers in Cincinnati Public Schools, Taylor and Tyler (2012) found that observation-based evaluation not only improved teacher performance in that school year but also increased teacher performance in the following years. An additional strength of teacher observations is that they offer teachers a sense of validity. This type of evaluation is common enough that many teachers and administrators have faith in it as a reliable measure of teacher quality (Goe et al., 2008).

However, in exploring the use of classroom observations to evaluate teacher quality in Steinberg and Sartain (2020) found evidence of a race gap in teacher performance ratings in Chicago Public Schools. This speaks to the potential limitations and weaknesses of the existing

models. In exploring this phenomenon, Steinberg and Sartain (2020) found that "teachers teaching in more economically and academically disadvantaged classes receive lower observation scores on average" (p. 70). They also found that teachers receive higher observation scores on average in schools with better instructional leadership, more resources to support instruction, and more supportive relationships with students and parents. This work was further supported by Jiang and Sporte (2016) who found similar evidence of lower observation scores earned by Chicago teachers in high -poverty schools.

When considering classroom observations as a tool for teacher evaluation, data-driven information on the observational tool being used is helpful for administration (Sartain et al., 2011). As school districts seek to implement effective teacher observations, they must address interrater reliability. Researchers in the MET study found that the use of video and calibration protocols was effective in improving consistency across raters (MET, 2012).

In exploring the use of teacher observation to evaluate quality teaching, the MET project utilized five observation protocols, including the Classroom Assessment Scoring System, the Framework for Teaching, the Mathematical Quality of Instruction, the Protocol for Language Arts Teaching Observations, and Quality Science Teaching. The project evaluated these teacher observation systems for reliability by examining the proportion of the variance in instrument scores that reflect consistent differences in practice between individual teachers rather than variations attributable to the particular observer, the group of students taught, or even the particular lesson. In addition, researchers have reported an association between observation ratings and a range of student performance scores for each observation system. These student outcomes included achievement gains on state tests, achievement gains on other assessments, and student self-reports of enjoyment in the class.

The study suggested improving the effectiveness of classroom observations in the following ways: First, the researchers advised that achieving high levels of reliability would require several quality assurances, including observer training and certification, system-level audits using a second set of instructional observers, and the use of multiple observations. Moreover, the researchers emphasized that classroom observation may potentially identify strengths and address weaknesses in teachers' practices. The study of Dudek et al. (2019) supports the idea that evaluations can help administrators identify teaching strengths and weaknesses to improve classroom learning. While classroom observations are an almost universally accepted source of data for the purposes of teacher evaluation, they may be limited in scope. Over the course of a school year, a teacher might spend hundreds of hours delivering instructions to students in the classroom. Of the total time spent in teaching, only 2 to 3 hours might be observed by an evaluator. Therefore, while the data from observations can be a valuable component, administrators may consider an evaluation tool that is not limited to one form of evaluation (Custer et al., 2018).

Principal Evaluations

Principal evaluations often include a pre-observation meeting with a teacher, an official observation, and a post-observation discussion with the teacher to go over the results. They are utilized for both summative and formative purposes and can be an informal drop-in visit by the principal to develop a brief impression of the classroom (Heneman et al., 2006). Brandt et al. (2007) found that most principal evaluation events were conducted for the purposes of high-stake employment decisions such as determining the probationary status of beginning teachers, dismissal of ineffective teachers, or conferring tenure to more experienced teachers.

Two studies examined the accuracy of principals in rating the effectiveness of teachers (Harris & Sass, 2007; Jacob & Lefgren, 2008). In both studies, the principals were asked to rate the effectiveness of teachers on a uniform rating scale. The ratings were based on the principal's overall impression of the teacher's effectiveness. Principal ratings were then correlated to value-added measures of student achievement for the same teachers. Jacob and Lefgren (2008) found that principal ratings were significantly weakly correlated with teacher value-added scores. As Jacob and Lefgren (2008) observed, "principals were found to be fairly accurate at identifying teachers in the top or bottom effectiveness but were less successful at distinguishing between teachers in the middle" (p. 115).

Instructional and Classroom Artifacts

Instructional artifacts generated in the classroom represent another approach to evaluating teacher effectiveness. These artifacts include classroom-based materials such as lesson plans, teacher assignments, assessments, scoring rubrics, and examples of student work. These artifacts are typically selected by the teacher and then presented to the evaluator as evidence of their instructional practice. Several protocols and rubrics have been developed to provide evaluators with guidance in the evaluation of instructional artifacts. For example, the Instructional Quality Assessment (IQA) was developed by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and provided guidance for both evaluations of artifacts and methods of providing teacher feedback. The IQA is used to evaluate the quality of discussion, rigor of lesson activities, and quality of expectations for students (Matsumura et al., 2006).

Likewise, the Intellectual Demand Assignment Protocol (IDAP) was developed by the Consortium on Chicago School Research to evaluate the degree to which classroom assignments involve the construction of knowledge, promote disciplined inquiry, and exhibit value beyond

school. Newmann et al. (2001) found that classroom teachers whose assignments received high IDAP scores showed student learning gains on the Iowa Test of Basic Skills that were 20% higher on average than those of classroom teachers with lower IDAP scores. The Scoop Notebook, developed by Borko et al. (2005), evaluated handouts, writings on the board, student classwork, homework, and classroom projects. Teachers in middle school classrooms provided example assignments and student work, took pictures of the classroom, and answered reflective questions. Borko et al. (2007) found that Scoop Notebook ratings correlated with observational measures but did not demonstrate any links to student achievement. Instructional and classroom artifacts have both strengths and weaknesses in relation to teacher evaluations. Artifacts such as student writing responses and tests can demonstrate students' understanding of concepts learned. However, each student learns differently, so a student's project may demonstrate understanding compared with a test taken by the same student (Goldschmidt & Congdon, 2018).

Teaching Portfolios

Teaching portfolios are defined as purposeful collections of carefully selected artifacts and reflections on teaching (Ouellett, 2007; Sung et al., 2009). The critical difference between portfolios and instructional artifacts is that while instructional artifacts can come from any source, portfolio artifacts are specifically curated by the teacher and meant to display what they consider to be exemplary work. In addition, portfolio materials are often selected specifically to meet certain predetermined standards. Artifact selection often requires the teacher to defend why specific artifacts are included and how they relate to specific standards (Painter, 2001).

The most widely recognized portfolio system in the United States is the National Board for Professional Teaching Standards (NBPTS) certification. The primary component of earning an NBPTS certification is a portfolio submission. The contents of this portfolio include videos of

17

instructional practices, teacher-student interactions, and student work examples. Each entry is accompanied by a reflection on and analysis of the part of the teacher explaining why the item was selected and included. Portfolios are then reviewed by trained assessors who have met training and qualification requirements; many assessors are teachers who have received the certification themselves. Several studies have found links between NBPTS certification and increased student achievement (Cavalluzzo, 2004; Clotfelter et al., 2006). However, studies have reported inconsistent findings, as other studies have found no such relationship (Cunningham & Stone, 2005; Sanders et al., 2005). Hakel et al. (2008) found that NBPTS certification is successful in identifying high-performing teachers, but evidence is lacking regarding whether the process itself leads to improvement or whether teachers who choose to engage in the process are simply already high-performing.

Attinello et al. (2006) conducted a study in a southeastern district of the United States to explore teachers' and administrators' perceptions of teaching portfolios. The study surveyed 752 teachers who had created portfolios and 46 administrators who had been involved in the support and evaluation of these portfolios. Semi-structured interviews were also conducted with 10 teachers and four administrators. The findings showed that 85% of the teachers and 93% of the administrators believed that the teaching portfolios revealed an accurate and comprehensive picture of teachers' learning and performance. While portfolios are easy to administer and teachers and administrators feel they are accurate measures of teaching (Tucker et al., 2003), some caution is required, as they are too teacher driven to be used as a stand-alone assessment in high-stake decision-making (Johnson et al., 2000).

Teacher Self-report Measures

Self-reported measures from teachers often take the form of surveys, instructional logs, or interviews (Little et al., 2009). While the format may vary, all teacher self-report measures are designed to gather information from the teachers themselves about their classroom activities and students' performances. Large-scale surveys have been created for this purpose, although the survey used may vary depending on the aspects of teaching and learning it seeks to explore (Little et al., 2009). Examples designed for use in all classrooms include those developed by the National Center for Education Statistics and Surveys of Enacted Curriculum (SEC). While they may be better known for their student tests, both the Trends in International Mathematics and Science Study and the California Learning Assessment contain teacher self-report surveys. The Surveys of Enacted Curriculum may be the most thoroughly studied teacher self-report tool. Blank et al. (2001) described how SEC data could be used in schools, and Blank et al. (2001) identified potential inconsistencies or inaccuracies in teacher answers due to differing interpretations of terminology and a lag in reporting time.

Instructional logs are another derivation of a teacher self-reported measure and require teachers to keep detailed records of their teaching. In studying these types of measures, Camburn and Barnes (2004) examined teacher log responses and compared them with the responses of third-party observers. Thirty-one teachers in eight schools were observed for a single school day. Following the observation, both the teacher and the observer filled out the same instructional log. Camburn and Barnes (2004) found that classroom logs exhibited higher levels of consistency between the different observers than between the teachers and the observers. They attributed this to the fact that because the observer was not embedded in that classroom, they may not be aware of all contextual information and may perceive classroom events differently than teachers.

Interviews have also been used as a form of teacher self-report measure to gather information on teaching and learning. Interviews can be either highly structured or open-ended, and they are often locally designed (Ford et al., 2017). When used in evaluations, teacher responses to interview questions are evaluated by a trained evaluator using a scoring rubric. In both cases, it was also recommended that administrators provide teachers with the interview procedure and standards to be implemented before the interview so that they can prepare materials and develop clarifying questions.

Teacher self-report measures in evaluations may represent a unique opportunity that allows an evaluator to evaluate a teacher's intent in ways that other methods of evaluation do not. They are also typically cost-effective and relatively easy to administer. In the case of surveys, evaluators invest little time in the data gathering and most of the time evaluating the results. This stands in stark contrast with other methods such as classroom observations, where typically much more time is spent in data gathering than in data evaluation. However, caution should be exercised when making high-stake decisions based on teacher self-report measures. Even a well-crafted teacher self-report measure is susceptible to an individuals' desire to present themselves in the best light. In addition, teachers can be highly selective about what they choose to report and what they choose to omit, which may lead to a less-than-complete evaluative picture.

Value-added Models

Value-added modeling (VAM; also known as value-added analysis or value-added assessment) is a teacher evaluation method that measures a teacher's contribution in a given year by comparing the current scores of the teacher's students with those of the same students in the previous school year on the same or similar assessment (Guarino et al., 2015; Wessel-Powell et al., 2019). Regardless of the statistical methods used, the goal of all value-added models is to

create an environment where "teachers are not held responsible for their students' incoming achievement, but rather are evaluated by how much they contribute to their students' learning" (Lipscomb et al., 2010, p. 7).

In 2015, Guarino et al. argued that VAM is a substantially more objective measure of teacher performance that allows school districts to define a teacher's impact more clearly regardless of the incoming abilities of the students and obtain a more accurate picture of the true impact of the teacher on student learning. Hershberg (2005) argued that value-added assessments provide school leaders with rich diagnostic information that they can use in assigning personnel, allocating resources, and identifying mentor teachers and coaches. In 2019, Guarino et al. reported that both value-added measures and beat-the-odds measures can be useful for measuring school effectiveness. Guarino et al. (2019) indicated that "student-level regressions that include prior test scores, other student characteristics, and school fixed effects are more desirable when school effectiveness measures used in high-stakes accountability policies place a high degree of importance on achievement" (p. 463).

Critics of the use of value-added models in teacher evaluation have expressed that administrators should use caution in interpreting the meaning of results from value-added measures because judgments of teacher effectiveness may vary simply depending on the test selected (Lockwood et al., 2007). In 2008, Goldhaber and Hansen evaluated value-added scores for teachers in North Carolina. They concluded that estimates of teacher effectiveness are not the same across years in reading or math (Goldhaber & Hansen, 2008), making them inaccurate or unstable (Ravitch, 2010). Similarly, Koretz (2008) questioned whether it is even possible to compare gains from 1 year to the next using tests that do not measure the same content. Strauss (2013) went as far as to state that there were some problems with value-added measures and

concluded that these measures should not be used for high-stake decisions, given their questionable reliability.

Teacher value-added models are attractive to administrators and educators as measures of teacher performance, as they present a more objective measure than many of the other methods detailed here (Guarino et al., 2015). They are designed to remove the subjective opinions of the evaluator or teacher that can lead to, for example, unreliability in classroom observations and teacher self-report measures. They also require limited work on the part of the evaluator, as the score is typically generated by an outside agent using statistical analysis.

While value-added models do not require a great deal of work or time on the part of the administrator, once received, they require substantial statistical formulation that may be beyond the skill set of local administrators (Braun, 2005; Heinrich & Good, 2018). As a result, school administrators may receive value-added scores for teachers in state-evaluated subjects such as math or English, but it may be difficult to create similar models for their teachers who do not teach state-tested subjects. As such, some teachers may receive value-added scores from tests that have been thoroughly statistically evaluated with complex controls, while others receive scores on less sophisticated locally created tests with little or no statistical validation of their value (Amrein-Beardsley et al., 2016). For example, an English teacher may receive a value-added evaluation score based on scores from a state-required achievement test that compares the growth performance of her seventh grade students to the growth of all other seventh graders in the state. Conversely, a colleague who teaches science may receive a value-added score based on the performance of their students on a locally created and graded pre/post-test for seventh grade science. The former meets a strict definition of a value-added measure; the latter clearly does not.

Student Perception Surveys as a Component of Teacher Evaluation

One emerging approach to teacher evaluation is the use of student perception surveys to inform evaluation ratings. Goldhaber (2019) shared that early exploration of the use of student perception surveys as a tool for evaluating teaching was dominated by studies in higher education. In exploring the perceptions of college and university students, multiple studies have found student perception surveys to be a reliable measure of teacher quality. Aleamoni (1999) found strong correlations between the use of student perception surveys and evaluations of the performance of instructors in colleges and universities. A meta-analysis of studies on student ratings conducted by Renaud and Murray (2005) led to similar conclusions based on their finding of an average correlation of .43 between the mean student ratings of instructors and the mean student performance on common final exams in multi-section courses. They also found positive correlations between student feedback ratings and those assigned by external observers: "The weight of evidence from research is that student ratings of teacher effectiveness validly reflect the skill or effectiveness of the instructor" (Renaud & Murray, 2005, p. 930). However, there are studies in the higher education space that question the validity of student perception surveys as a measure of instructor performance. Krautmann and Sander (1999) found results "consistent with the hypothesis that instructors can 'buy' better evaluations through more lenient grading. Kane (2012) found that higher education uses student evaluations almost universally and that they are often the only data collected on the instruction of college professors.

Geiger and Amrein-Beardsley (2019) documented six student perception surveys (see Appendix C) commonly available to schools and provided descriptive information for each. This descriptive information included the cost, constructs or domains assessed, number of items, response option types, and grade levels where it can be administered. They identified several

possible benefits of utilizing student perception surveys, including the fact that they target the population with whom teachers interact the most, that they can provide valid and reliable data on classes not measured by state standardized tests, that they are relatively quick and easy to administer, and that they can gather information on multiple facets of teaching.

Unlike those on higher education, studies that examined the use of student surveys in K-12 education are much fewer and less developed. However, researchers have examined the potential of using student perception surveys as more K-12 institutions begin to experiment with them as a component of teacher evaluation systems. In the early 1990s, it was estimated that student perception surveys were being used in under 5% of school districts nationwide (Loup et al., 1996). The use of student perception surveys began to increase in the United States in the early 2000s, with the development of the Tripod survey, which would eventually become the most widely used survey in the country (LaFee, 2014). Dr. Ronald Ferguson (2001) developed Tripod when other measures failed to provide insight into the reasons for the disparities among student surveys. The resulting survey has grown and developed into the current Tripod survey.

A renewed interest in the use of student perception surveys emerged with the publication of the Widget Effect (Weisberg et al., 2009). In this work, Weisberg et al. (2009) highlighted various shortcomings in most teacher evaluation systems. One of their recommendations for increasing the rigor of teacher evaluation systems was the inclusion of student perception surveys as a component of teacher evaluation systems.

Wilkerson et al. (2000) examined student perception surveys from 1,976 students in Wyoming and compared the data gathered from their student perception surveys with student performance in reading and math. They found that student perception survey results not only strongly but also much more strongly correlated with student achievement in reading (.75) than

with teacher self-ratings (.21), principal ratings (.09), and principal summative evaluations (.34).

Similarly, Peterson et al. (2000) evaluated results from 9,765 surveys of students ranging from

kindergarten to high school. They also found these surveys to be reliable measures of teacher

performance. These types of research studies add to an emerging body of literature that has

begun to "provide convincing evidence that student ratings of teaching are worth considering for

inclusion in teacher evaluation systems" (Goe et al., 2008, p. 40). Ferguson (2010) indicated that

student perception surveys can capture key dimensions of the classroom and teaching practices

as students experience them.

The Measures of Effective Teaching (MET) project investigated the relationship between

several different measures of teaching and value-added estimates of student achievement. The

study investigators utilized 36 questions from the Tripod survey. Survey items were chosen to

assess the degree to which students viewed the classroom environment as engaging, demanding,

and supporting their intellectual growth. In their initial findings, Kane and Cantrell (2010) found

a significant correlation between a teacher's score on the student survey and value-added

achievement on tests in both math and language arts. This suggests a potentially powerful

connection between what students say their teachers do and how well they perform on

standardized achievement tests.

In addition to the validity and reliability of the feedback provided to teachers and

evaluators about a teacher's ability in the classroom, supporters of the use of student perception

surveys of teacher performance also point to some additional reasons that they should be

included as part of a teacher's evaluation. They pointed to the fact that student perception surveys

are low-cost and require little work on the part of the evaluator (Schulz et al., 2014). While it

might take hours for an evaluator to observe multiple teachers or evaluate their artifacts, student

perception surveys allow evaluators to gather a substantial amount of data on each teacher they evaluate in less than an hour. The inclusion of student perception survey information can also reduce the halo or horn effect in teacher evaluations. The halo effect tends to be used to describe the influence that a generally positive impression on one aspect of performance has on specific performance criteria. After raters cast their halo around their subject, they are so dazzled by its radiance that they cannot differentiate the subject's separate qualities (Johnson & Vidulich, 1956). Closely related is the horn effect, or the tendency of an unsatisfactory rating on one aspect to influence the rater to give a person a similar or lower-than-deserved rating on other specific criteria (Henderson, 1984; Landy & Farr, 1983).

The use of student perception surveys as a component of teacher evaluation is hardly universally accepted, and there are detractors who advise caution when employing them. While some researchers have found that student perception surveys can provide valid and reliable data about teaching when correlated with other measures, others have found much more modest correlations. Tripod is a leading provider of educational assessments (Liu et al., 2015). Its evaluations include insights into teaching practices, student engagement, and school climate. Polikoff (2015) found that the overall stability of the Tripod survey was low to moderate. An additional concern is that while students may provide feedback on whether a teacher is engaging or responsive to their needs, they may be limited in their capacity to provide feedback on a teacher's understanding of a certain set of standards or performance objectives (Peterson et al., 1998).

Student input can provide educators with feedback on lesson engagement and understanding. However, some researchers have also expressed concern that it may be difficult to convince teachers that student surveys provide reliable and actionable information to inform

their evaluations (Rothstein, 2010). They pointed to the fact that teachers may fear that student evaluations will be influenced by their personal feelings about the teacher (Schulz et al., 2014). Marsh (2007) created a list of these potentially biasing factors as expressed by teachers, and they ranged from course rigor to class size and to whether a course was an elective or a required course. The nonrandom placement of students into classes may be another potential issue with the use of student perception surveys (Rothstein, 2010). They argued that classrooms might be homogenized based on factors such as race/ethnicity, socioeconomic status, gender, or achievement level (Desimone et al., 2010). The strategic placement of students could have a substantial negative impact on the reliability and validity of student perception surveys collected from those classes.

Teacher Attitudes About Student Perception Surveys

While little research has explored the principal beliefs about student perception surveys, there has been a small but emerging body of related research exploring teacher attitudes about student perception surveys and their inclusion as part of comprehensive teacher evaluation systems. Kauchak et al. (1985) interviewed and surveyed teachers about various evaluation methods, including student perception surveys, and found teachers to be equally divided among three different views of teacher perception surveys. The most positive third of teachers felt that these types of surveys can provide valuable information but that professional judgment should be exercised when evaluating the results. The middle third was highly skeptical of the data from student perception surveys and felt that any information gathered from students should be interpreted with a great deal of caution. The final third actively opposed the use of student perceptions surveys and doubted that students could provide valid or reliable information about teaching performance. When the results were disaggregated by level, it was found that

27

elementary school teachers were more likely than secondary teachers to express a negative view

regarding the use of student evaluations. Many of these teachers expressed doubts that

elementary-age students could understand the complexities of teaching and differentiate between

competent and incompetent behaviors. Although policies have changed since this research was

conducted, a study by Geiger and Amrein-Beardsley (2019) supports many of the concerns

expressed in the study of Kauchak et al. (1985). In Geiger and Amrein-Beardsley's study,

educators expressed concern that not all students can evaluate all aspects of teaching.

     Schwab and Iwanicki (1988) found similar results when they explored a performance-

based salary program for teachers. In their research, they found that 41% of teachers indicated

that they did not support the inclusion of student feedback in the evaluation process, 47%

supported the inclusion of student feedback with reservations, and only 11% strongly supported

the inclusion of student feedback. Schwab and Iwanicki (1988) also found that high school

teachers were more in support of including student feedback than junior high or elementary

school teachers.

     Drawing upon research conducted in 16 rural school districts, the researchers examined

both teachers' and administrators' beliefs about the utility of student perception surveys as a basis

for teacher evaluation. Specifically, the study sought to determine whether teachers and

administrators believed student perception surveys could provide principals with information that

enabled them to assign a valid rating to teacher performance. Table 1 summarizes the findings of

a research study by researchers at the University of Minnesota (Dretzke et al., 2015). This study

offered some indication of the potential that student perception surveys have for being a

component of teacher evaluations.

Table 1

Ratings of the Minnesota Model Student Survey

Rating	Summative Evaluator Percent Rating	Classroom Teacher Percent Rating
Poor	14%	43%
Fair	36%	36%
Good	36%	16%
Very Good	14%	4%
Excellent	1%	0%

Note: From the study of Dretzke et al. (2015).

Schulz et al. (2014) found similar concerns expressed by teachers when they reviewed the use of student perception surveys across states. They identified gaining teacher buy-in and support as the first challenge school districts face as they began including student perception surveys as a component of a comprehensive teacher evaluation system. School districts in "Georgia, for example, originally planned to require student surveys to count as 10% of a teacher's evaluation" (Schulz et al., 2014, p. 11). The researchers also pointed to other districts such as those in Connecticut, where teachers raised concerns regarding survey benchmarking, measurable goals from surveys, and what constitutes movement from fall to spring surveys. According to Schulz et al., "Connecticut leaders struggled to respond to these questions and therefore determined that student surveys would be an optional, rather than required, component of teacher evaluation" (2014, p. 11).

Summary

Multiple previous studies have found student perception surveys to be a reliable measure of teacher quality in higher education; however, inadequate information is available concerning how administrators perceive the use of student perception surveys at the K-12 level. Although

research supports the use of evaluation tools for educator assessments (Dudek et al., 2019; Firestone & Donaldson, 2019), there is a gap in the literature concerning how administrators in K-12 schools perceive the value of student perspectives in the teacher evaluation process. This chapter presents a literature review of both previous and current studies that support the many facets of teacher evaluations and the relative components that are part of the teaching evaluation process. Information on teacher quality and student achievement, and subsections on observations, principal evaluations, instructional artifacts, portfolios, teacher self-reporting, value-added models, and student perceptions surveys were presented. Information was also presented on the role of the school principal and teacher attitudes toward student surveys. Information was presented on the theoretical framework and limitations of the research. The following chapter will include information on the methodology of the research study.

Chapter 3

Research Methodology

This qualitative study draws data from semi-structured interviews and district artifacts to investigate participants' perceptions of the potential opportunities of using student perception surveys in the teacher evaluation system in the district under investigation.

Research Setting

The research took place in a suburban school district located near a large Midwestern city. The district was formed by parents in 1955 who had moved outside the city limits and envisioned a strong local school district that would meet the unique needs of their children. Tables 2 and 3 present an overview of the district demographics. As illustrated in Table 2, the ethnicity of the students in the identified school district was predominantly black, representing approximately 40% of the student population. As illustrated in Table 3, the identified school district did have a high concentration of poverty, as represented by the percentage of student qualification for free and reduced lunches. Approximately 57% of the students in this district qualified for free or reduced lunches. The district currently offers the International Baccalaureate (IB) curriculum (an inquiry-based learning style that is research based and motivates students to succeed in academics) to a diverse population of 11,200 students enrolled in kindergarten through 12th grade. The district is currently made up of seven elementary schools, three middle schools, and one high school. It employs approximately 786 teaching staff members.

Table 2

District Ethnicity by Enrollment

Student Ethnicity	Students Enrolled	Total Percent of Student Enrollment (%)
American Indian	0	0
Asian	410	3.7
Black	4,473	40.2
Hispanic	2,219	19.2
Multiracial	694	6.2
Native Hawaiian or another Pacific Islander	5	0

Table 3

District-free and Reduced Lunch Status by Student Enrollment

Category	Students Enrolled	Percent of Total Student Enrollment (%)
Free Meals	5,553	49.8
Paid Meals	4,701	42.2
Reduced Meals	886	8

Research Participants

This study included 10 purposefully selected school and district administrators employed in a suburban district in the Midwest. I selected principals from eight elementary schools based on their participation in a longstanding discussion about the use of student perspectives as one factor in determining a teacher's overall evaluation rating. All the principals included in the study had been involved in the collaborative development of a comprehensive teacher evaluation system and had utilized the evaluation system to assign scores to classroom teachers. District

personnel (assistant superintendent, human resources director, and director of elementary education) were selected on the basis of their direct involvement in the creation and application of the district evaluation instrument or in overseeing principals who used the instrument to evaluate teachers. Table 4 presents the participants who were invited to participate in the study.

Table 4

Participant Role, Background, and Education Level

Participant Pseudonym	Gender	Race	Current Role	Level	Served on Evaluation Committee	Years of Experience in Current Role	Highest Degree Obtained
Mr. Washington	Male	White	Principal	Elementary	No	4	Masters
Ms. Harrison	Female	Black	Principal	Elementary	Yes	1	Masters
Ms. Tyler	Female	Black	Principal	Elementary	No	4	Masters
Ms. Monroe	Female	White	Principal	Elementary	No	13	Masters
Ms. Madison	Female	Black	Principal	Elementary	Yes	6	Masters
Ms. Adams	Female	Multi	Director of Elementary Education	Central Office	No	7	Masters
Ms. Van Buren	Female	White	Principal	Elementary	Yes	1	Masters
Mr. Polk	Male	White	Principal	Elementary	No	2	Masters
Mr. Jefferson	Male	White	Assistant Superintendent	Central Office	Yes	1	Doctorate
Ms. Jackson	Female	White	Principal	Elementary	No	2	Masters

Positionality

I am a former principal and current administrator in the same school district where this research was conducted. Thus, I have many preexisting relationships with the study participants

and other stakeholders who participated in district conversations about this role. While the topic of student surveys of teacher performance has not been discussed in the district, I have participated in conversations about teacher evaluation practices in the district over the last several years. Therefore, my understanding of the evaluation practices in the district will significantly and directly inform my interpretation of the data. Furthermore, my own biases about teachers and teacher evaluations will significantly shape my views.

In addition, as a former principal, I have experience utilizing the evaluation system and my own opinions regarding the values of its components. My experience as a classroom teacher who was evaluated also informs my perceptions of student surveys and their inclusion as an element of evaluation. I surveyed students annually about my instruction and found value in the information I received.

Data Collection

Data collected from the interviews and evaluation documents were prepared for analysis. For the interviews, each interview audio recording was converted into a Microsoft Word transcript. Participant identification was removed and replaced with pseudonyms. I listened to the audio recording three times while reading through the transcript to verify that the transcription was accurate. Once the transcripts were verified to be accurate, the data and the collected evaluation documents were ready for analysis. For this study, the participants included school and district administrators employed in a suburban district in the Midwest. Table 4 (page 33) presents the participants' demographics. Before the collection of any data, the participants were informed of the study and asked to sign a consent form (see Appendix E) that included information about the study and researcher and the contact information.

Semi-structured Interviews

I conducted semi-structured interviews with 10 school personnel, including school principals and district central office administrators, to examine their beliefs about the inclusion of student perception surveys in the teacher evaluation system in the district. I used interview questions to examine principals' perspectives about the value of student perception surveys, their level of comfort in including this type of data as a part of formal evaluation, their desire to utilize an externally or internally developed student survey instrument, and the extent to which they feel that student survey data will align with previously used measures of teacher effectiveness and student achievement data. The interview questions (see Appendix F) were based on the research questions. By having an interview guide to follow, I maintained consistent interviews with each participant and reduced the potential for bias (Creswell & Poth, 2017). The interviews lasted 45 to 60 minutes and were recorded using virtual meeting software. For convenience, the interviews were conducted via Zoom, a virtual meeting platform. Before the analyses, the audio files were transcribed.

Documents

In this study, I collected documents relevant to the teacher evaluation process in the school district. The documents that may be accessed for analysis include teachers' informal and formal observations, evaluation guidebooks, teacher evaluation rubrics, and other documents that may prove useful and relevant to the study's topic (see Appendix H). These documents allow for tracking changes in practice over time and provide a resource to triangulate findings from other data sources.

Data Analysis

A thematic analysis was used to identify themes from the data that could answer the research questions (Braun & Clarke, 2006). The process included the refinement of the qualitative data such as the interview transcripts used in the study. Stake (1995) defined data analysis as "a matter of giving meaning to first impressions as well as to final compilations" (p. 71). In many ways, it requires a researcher to take apart observations and impressions to make meaning. In a case study, the researcher must understand why participants take a specific stance on a topic or react to procedures in a particular way (Yin, 1984). To gain better understanding, I employed "certain protocols that help them draw systematically from previous knowledge and cut down on misperception. Still, there is much art and much intuitive processing to the search for meaning" (Stake, 1995, p. 72).

For this study, I utilized Dedoose, a web-based qualitative analysis tool, to organize the analytic process. I read, coded, and wrote memos on each of the interview transcripts to note comments relevant to the inclusion of student perception surveys as a component of teacher evaluation systems. I focused specifically on comments related to what the administrators feel could be gained by including these types of surveys and the areas of concern they have surrounding their inclusion. In this analysis phase, I searched for illustrative instances. Using basic coding, I identified illustrative comments and flagged them for further analysis. In the first coding cycle (Saldana, 2009), the items coded ranged in magnitude from a single word to a full page of text. The primary function of these codes was to organize and narrow the data set, as they are relevant to the research questions and focus of the study. These generic codes included terminologies such as an administrator comment or other similar phrases and keywords.

Following this initial review, I reviewed previously identified passages and conducted a second coding cycle to assign descriptive codes (Saldana, 2009) to those that were directly informed by the theoretical perspectives in the study. These codes were developed a priori, and I specifically highlighted administrator comments on student perception surveys related to reliability, alignment with current instructional ratings of teachers, school culture impact of using student perception surveys as a component of the evaluation, and logistical concerns of implementing student surveys. For the coding process, I considered Strauss and Corbin's (1998) work to organize and refine all the categories until distinct themes emerged. While these codes were developed a priori, coding can be a fluid and reactive research process. While two coding cycles were planned for this research, the potential need for additional cycles as the research unfolded was acknowledged. In accordance with the focus of this research design, recoding and recategorizing were necessary, as codes became increasingly refined throughout the process. In the process of movement from the first to the second cycle, the coding data were rearranged and reclassified. The goal in this process was that through each coding cycle, codes and categories become more refined. Once all coding cycles were complete, all final descriptive codes were applied. Table 5 presents codes that were grouped, the categories under which they fall, and the number of times these codes occurred throughout the transcripts. I then grouped the codes into broad categories that will ultimately become the basis for final themes. Labeling the themes was the final step in the data analysis. Table 6 presents the six themes identified using the codes.

Table 5

Codes from Transcripts

Categories	Codes	Relevant Research Question	Count
Survey Integration Benefits	Effectiveness of Evaluation on Important Characteristics	RQ1	5
	Integration Positives	RQ1	7
	Potential Opportunity	RQ1	3
	Positive Primary and Secondary Measures	RQ1	6
	Surveys More Accurate than Observations	RQ1	7
Problems with Survey Integration	Inter-rater Reliability in Evaluation	RQ1	9
	Importance of Being in the Classroom without Evaluations	RQ1	2
	Improper Use of Artifacts	RQ1	3

Categories	Codes	Relevant Research Question	Count
Problems with Survey Integration	Integration Negatives	RQ1	2
	Gotcha Evaluations	RQ1	5
	Primary and Secondary Measure Concerns	RQ1	8
	Student Survey Concerns	RQ1	5
	Rating Favored Teachers Higher	RQ1, RQ2	8
	Transiency	RQ1, RQ2, RQ3	3
Survey Feedback Can Help Shape Classroom Instruction	Artifacts	RQ2	16
	Elementary Student Feedback	RQ2	3
	Middle School Feedback	RQ2	5
	High School Feedback	RQ2	3
Surveys Help Identify Student Needs	Positive Student Survey	RQ1, RQ2	13
	Surveys Relative to Measure	RQ2, RQ3	5
	Survey Students Personally	RQ2	6
Ongoing Evaluation of Survey is Beneficial	Formats of Perception Survey	RQ3	5

Weight of Surveys in Teacher Evaluation	Weight of Student Perception Surveys	RQ3	9

Table 6

Themes of Principal Perceptions Concerning Teacher Evaluations

Research Question	Theme
Research Question 1	Principals See Potential Benefits of Student Survey Integration
	Principals Share Potential Problems with Student Survey Integration
Research Question 2	Feedback Shapes Instruction
	Student Surveys Help Identify Student Needs
Research Question 3	Ongoing Evaluation to Understand Value of Survey
	Considerations for Weight of Survey

Analytical memos were also written to create an audit trail and to identify any locations where codes are added or changed. I used analytical memos to document how I refined my coding scheme and developed my data interpretation. My memos included but were not limited to frustrations in the research process, new connections, sudden epiphanies, or any other thoughts surrounding the research process.

	Principals Share Potential Problems with Student Survey Integration
Research Question 2	Feedback Shapes Instruction
	Student Surveys Help Identify Student Needs
Research Question 3	Ongoing Evaluation to Understand the Value of Survey
	Considerations for the Weight of Survey

I also analyzed my observation field notes. Observation field notes should be distinguished from analytical memos. Researchers use interviews and observation field notes to capture not just the words expressed by participants but also broad social interactions. As such, they include observations surrounding body language, tone, pace of speech, and other aspects of social interaction. Observational notes helped me infer how a participant felt about answering specific questions. Body language may suggest that the participant has more to share, and this type of observation can help prompt the participant to share more information. While analytical memos are designed to capture the changing thoughts and perceptions of the researcher concerning the research questions, observation field notes are designed to capture the researcher's social interactions with the research participants.

Document analysis is a systematic procedure for reviewing or evaluating both printed and electronic (computer-based and internet-transmitted) documents. The analytical procedure entailed finding, selecting, appraising (making sense of), and synthesizing the data contained in the documents. The document analysis yielded data such as excerpts, quotations, or entire passages. These data were then organized into major themes, categories, and case examples, specifically through a content analysis (Labuschagne, 2003). I used feedback from the

participants to answer the research questions, which helped me understand how student

perspectives were viewed by administrators as part of the teacher evaluation process.

Triangulation is often used in research to assure validity and involves using more than

one data collection method (Creswell & Poth, 2017). For this research study, interviews were

triangulated with observations and documentation. The triangulation of data through interviews

and document analysis resulted in a "confluence of evidence that breeds credibility" (Eisner,

1991, p. 110). I used the information of the transcripts and compared that information with my

document analysis to confirm relevancy to the research questions and focus of the study.

Limitations

All studies have limitations. For this case study, as an Indiana University doctoral

student, I ensured a position of neutrality in all communications related to the research study to

separate the work from my work related to my role in the district. One limitation to the study is

the small sample size, which limits the generalizability. An additional limitation is the study

population, most of whom were predominantly responsible for teacher evaluations at the

elementary school level. This limitation may limit the generalization to the secondary level, as

information gathered may reflect unique considerations related to using student perceptions

surveys with older students and the perceptions of secondary school principals on their use. On

the basis of these limitations, it will be incumbent to continually reinforce throughout all

research findings that the data and conclusions generated in the study stem from the participants'

statements. This is thoughtfully reinforced in the title of the research study and throughout all

documents related to the research.

Chapter 4

Results

In 2011, the Indiana General Assembly passed Senate Enrolled Act 001 (Public Law 90) as part of a comprehensive education reform package focused on improving student outcomes and educational practices. The act significantly changed how classroom teachers in Indiana were evaluated and compensated for. As part of this reform, school districts in Indiana were directed to select from several approved teacher evaluation systems or to develop their own teacher evaluation systems and submit them for state approval. The school district where this study was conducted chose to develop their own teacher evaluation model. Representatives from the Indiana Teacher Appraisal and Support System facilitated a district committee that combined elements of the RISE, Danielson, and Marzano teacher evaluation frameworks to develop the district's own modified teacher evaluation system. The district implemented the new system in the 2012–2013 school year. In 2016, the school district was recognized by the Indiana State Board of Education and the Indiana Department of Education as having achieved exemplary status in the development of a teacher evaluation system. Despite the praise received by the district, significant discussions continued about how best to configure the system to support teacher growth. One set of questions relates to the potential use of student perception surveys as a core component of the evaluation process. This study reports on the perceptions of individuals within the district about the potential use of student surveys and further describes the perceived benefits, drawbacks, and challenges of using these data as part of the evaluation process compared with the existing processes in the district. In all, seven of the 10 participants in this study were involved in some form of the committee to create, review, or enhance the teacher evaluations system.

Description of the District's Evaluation System

Before discussing the participants' perceptions, it is valuable to discuss the evaluation system developed by the district in response to the education reform legislation. As initially conceived, the school district's teacher evaluation system included data from multiple sources such as classroom observations, teacher reflections, and samples of student work. This information is used by a school principal to establish the final performance rating for a classroom teacher. Administrators conduct informal and formal observations of classroom instruction and attach rubric scores to events such as classroom lessons that are observed in person throughout the academic year. Teacher- and administrator-created artifacts also receive rubric scores. Finally, the system also included primary and secondary value-added measure scores. These value-added measures vary by teaching assignment, and most are locally created pre-post or achievement tests, nationally normed assessments (predominantly NWEA), or mandatory state assessments. In total, classroom observations by administrators and artifacts rated by administrators determine 80% and value-added measures determine 20% of a teacher's final yearly evaluation rating. These combine to determine a teacher's evaluation category for the year. The categories are "Ineffective," "Improvement Necessary," "Effective," and "Highly Effective."

Finding Number One: Reported Strengths of Student Perception Data in Teacher Evaluation

The participants in the study recognized that current data sources were sufficient to evaluate lesson design and delivery but did not provide principals with information about students' experiences with the lesson, their interactions in the classroom, or the informal

exchanges that occur between the students and the teacher. Thus, the participants expressed the view that collecting student perceptions survey data might be one way for principals and the district to gain additional insight into student experiences that were not considered by the current evaluation system. The interview analysis found that six of the 10 participants considered this as a potential strength of including student perception survey data. The participants thought that this survey information provided data on the individualization of instruction that principals cannot currently access via observations, artifacts, and measures. The participants spoke about the possible benefits of asking students how their teachers had tailored instruction to reach their unique learning needs. In describing this possibility, Ms. Jackson stated:

> If they (students) can communicate with us and share, "I don't see myself represented in the instruction in the books or in the way that you respond," those voices can shape our instruction and what we do as teachers and with our curriculum... But I think it could be done, and it would be very impactful for sure.

This comment illustrates how student perception surveys might enable administrators to understand how issues of representation and identity are addressed within the teacher's instruction. This information might then help administrators gauge the students' sense of belonging and the teacher's ability or intention to foster this sense among students. Mr. Jefferson echoed this statement when he stated:

> I think they could tell our teachers about how they feel from an identity perspective in that class, from a relational perspective in that class... Again, I'm speaking broad strokes, but I do think our kids could speak to the community.

This sense of belonging or community was an important element of teacher quality that the participants noted was not currently captured in their teacher evaluation system. Ms. Jackson and Mr. Jefferson both articulated, as did other participants, that the survey data might reflect information about the sense of belonging expressed by students or created by the teacher through their instructional practices. Such information is unlikely to appear through classroom observation alone. In total, these quotes underscore the significant opportunity these two participants believe student surveys would provide to capture a whole class or group's perception of belonging within a teacher's class. The comments were also representative of other participants.

Beyond examining issues related to the student's sense of belonging, the participants also noted that student perception surveys might afford them opportunities to examine how students were engaged in the learning environment. Indeed, this was seen as a more important piece of information produced by the evaluation system. For example, Ms. Tyler felt that student perception surveys could provide evaluators with a window into the effectiveness of teacher lesson design and differentiation and how the choices made by teachers contribute to engagement in the classroom, as she stated:

> Did they (the students) feel like it needed to be more engaging? Do they feel like they're needed to be more opportunity for conversation or hands-on opportunities? Do they understand, were you clear, was your pacing at a good rate? Did they have opportunity for mastery?

This quote demonstrates that the principal felt that student perception survey data would yield information about engagement and differentiation in the classrooms they oversee. Her statement,

as well as others like it, also reflects how Ms. Tyler views the student perception data being used.

It focuses on the kinds of learning experiences, the student's engagement, and the support the

student received to achieve mastery. Ms. Tyler's comments demonstrate a belief on the part of

administrators that student perception surveys would provide data to them on individual student

interactions with instruction that current teacher evaluation components are unable to provide.

This point was reinforced by other participants in the study. For example, Ms. Madison stated:

> I think it would tell a lot about teaching style. Like students could speak to the way the
>
> teacher teaches, and they, you know, they can share, well, this is the way I learn, but you
>
> didn't teach it like this. "So, it was hard for me to understand everything you were
>
> teaching or, you know, like you talk too much, I need to see stuff more often, you know,
>
> visual more than you talk. I need to be able to talk to my neighbors more, you know, to
>
> do my thinking."

In this quote, Ms. Madison offered her belief that student perception survey data would allow

principals to evaluate how students are receiving instruction in a new way. They would provide

principals with data on how that instruction is perceived and received by students. In doing so,

these data expands upon traditional methods of teacher evaluation, which often focus on how

effectively instruction is designed and delivered. She believes that student perceptions survey

data would be provide principals with valuable data about the match between lesson design (i.e.,

what they do) and student learning styles (i.e., what students need). Specifically, this would

allow principals to review data on how instructions are received by students on an individual

basis and therefore might contribute to discussions with teachers about the ways in which they

are approaching instructional activities that are more specific and tailored than other evaluation practices.

In addition, six of 10 participants in the study also identified ways in which evaluative dialogue between administrators and teachers would be influenced or expanded by principal access to student perception survey data. Specifically, the participants stated that data from student perception surveys would provide a previously unavailable element for dialogue between administrators and teachers in the evaluation process, improving both the volume and quality of this dialogue. In describing how the addition of student perception surveys might impact administrative dialogue, Ms. Jackson felt that the incorporation of these surveys would provide administrators with tangible feedback that they could use to help teachers genuinely assess the extent of students' sense of belonging in their classroom and whether the teacher's attempts at differentiation are hitting the mark. She states that administrators and teachers could "let the students tell us the way that they're feeling and what they need from us, like it could drive a lot of our instruction." This quote demonstrates her belief that the inclusion of concrete feedback from students about their classroom environment and the way they interact with instruction would provide her with key elements for teacher dialogue that have not previously been at her disposal. Ms. Madison put this more bluntly by saying, "I really feel like some teachers are clueless on how students feel about them and perceptions (they have)." Ms. Madison believes that introducing student perception surveys could provide her additional tools to improve the substance and quality of her dialogue with teachers, requiring them to confront their lack of knowledge, evaluate the climate in their classroom, and better understand the effectiveness of their differentiation. The inclusion of student perception surveys could provide her with direct feedback from students that would change this dialogue substantially, shifting the conversation

away from what the teacher is attempting to do to create classroom culture and differentiate

instruction and focusing instead on the extent to which students express that these efforts to

create classroom culture and differentiate instruction are effective from the perspective of

students. Ms. Van Buren envisioned such steps that lead to a dialogue "worded in such a way

that teachers weren't defensive, you know, so they use it in a really productive way… it isn't just

like a time for people to say lots of negative feedback." This statement is indicative of the

administrators' expressed feelings that administrator and teacher dialogue around student

perception data will differ from traditional teacher evaluation dialogue because the material for

the dialogue was not created by the teacher or the evaluator. The participants expressed that this

may present an opportunity for both administrators and teachers to de-center themselves in a way

that is not possible with current evaluation evidence generated by either the teacher or the

administrator.

Finding Number Two: Reported Weaknesses and Concerns with Student Surveys in Teacher Evaluation

Despite strongly supporting student surveys in teacher evaluation, the participants also

noted concerns about the use of student data in evaluating classroom teachers. One specific

concern they expressed was that student surveys could induce personal sentiments about an

instructor in the evaluation process. The participants believed that these sentiments could

undermine the information gathered through the surveys and invite concerns on the part of

teachers about the fairness of the evaluation. The participants expressed concern that student

sentiments, which they described as opinions or reactions to teacher behaviors or the content of a

course, could influence how students rated teacher performance on the survey. The participants

were concerned that some students would not be able to limit themselves to reporting about the teacher's professional dispositions toward students, communication, or classroom management.

The analysis indicated that seven participants viewed students' tendency to (dis)like teachers as being something that could influence student survey ratings. This influence could be positive, as would be the case when a teacher was well liked, or it could be negative, as would be the case when the teacher was not well liked. Specifically, the seven participants indicated that students might rate teachers higher on the basis of the teacher's personal characteristics, which might have little or nothing to do with their skill as an instructor. For example, Mr. Polk indicated that teachers might be rated higher because they "asked students about football." Ms. Madison noted that they would give higher rates to teachers whom they perceive to be "fun." Mr. Jefferson referenced students rating teachers higher who were "funny" or "easy graders," and Ms. Jackson stated that "a lot of elementary kids still think you're a good teacher because you're pretty or (because of) what you wear." As illustrated by these comments, the teacher traits that administrators felt might influence student ratings ranged widely and are indicative of participant concerns that the personal characteristics, not the instructional characteristics, of teachers might influence student ratings of teacher quality, even in areas where students were asked to provide feedback about instructional practice.

Classroom management practices were also viewed by the participants in the study as an important consideration and one that could potentially influence how teachers were rated on the perception surveys. Two participants expressed concerns that students might allow their personal feelings about the teacher as classroom manager or disciplinarian to influence their responses regarding teacher quality. In illustrating this concern, Ms. Jackson provided a potential scenario

that could result in a teacher's score being lowered by a student's negative feelings toward a recent disciplinary event:

A kiddo that came back to school today that had been suspended and felt down this morning, I came up there to intervene, and her first response was, "You're the person that sent me home for five days." So if she was to rate me today, I probably wouldn't get a good rating, but she and I usually are like this (makes a hand gesture of two fingers crossed together). So it's kind of like on what day am I getting (evaluated)?

Ms. Jackson's statement reflects concern that a teacher might receive very different scores from a student on a student perception survey if that perception survey is given to the student within a few days of a negative disciplinary event. Ms. Monroe expanded on this concern and provided her own example of a teacher she felt might receive scores from students who were lower than she deserves because of her high expectations for student behavior, as she noted:

I have a teacher, (Teacher Name Omitted), she's excellent, and people love her. I get tons of requests for her. She is hardcore. I mean, don't mess with me. Take no prisoners, but she loves her kids. So she will rip you, but then she will hug you after, you know, that kind of person, but she's not for everybody. So while they got amazing instruction and she's got great communication with parents, I believe there are kids who would've said she was horrible. "I didn't like her at all." And "She didn't like me." And it's because she probably actually was the best teacher they ever had.

Ms. Monroe's explanation illustrates how student feedback about instruction might be influenced by students' feelings about the teacher as a classroom manager or disciplinarian. In her estimation, this could lead to teachers with high expectations for behavior being rated lower than

52

they should be instructionally because of students' personal feelings about those high behavioral expectations. This has long-term implications for a teacher whose strong and assertive management style might lead to lower student instructional ratings throughout the year. Both administrators ultimately believed that students may not have the emotional maturity to accurately rate teachers' effectiveness in all areas; thus, this could be a potential consideration for the inclusion of their perspectives in teacher evaluation processes.

Finally, the participants also observed that the subject matter could influence how the student rated the teacher. As one principal expressed, students might rate teachers based on whether they liked or disliked the content being taught. Mr. Jefferson said, "It could be that a student says taking a STEM class may really love STEM. And it doesn't matter who the heck the teacher is." Mr. Jefferson's comment reflects a concern that students might rate a teacher as a more effective instructor simply because they are highly engaged in the content they are learning from that teacher. In this case, the students' positive feelings about the content of the course could lead to an inaccurately high rating of the teacher as an instructor. Conversely, Mr. Jefferson also shared a potential scenario that could result in a teacher score being lowered by a student's negative feelings about a course where they struggle with the content. As Mr. Jefferson later reflected, a student might be feeling that "math is hard, you know, and I don't like math and I'm in a class with a teacher who isn't letting me disengage." In this case, the student's negative feelings about the content of the course could lead to an inaccurately low rating of the teacher as instructor. Mr. Jefferson's comments illustrate his concerns about the undue influence that a students' personal feelings about course content might have on student perception surveys. He asserts that students in a class where they have a strong affinity for the content might rate the

instructor much higher than they deserve to be rated simply because they love the content being taught, and conversely, students who have strong negative feelings toward the content of a course might rate the instructor much lower than they deserve simply because they cannot stand the material being taught.

Two participants in the study expressed opinions that contrasted with the other eight in reference to the influence that student's personal feelings might have on their ratings of teacher quality. They viewed the inclusion of a student's perceptions of the instructor as a person as a much more valid and legitimate measure of teacher quality. For example, Mr. Polk shared his belief that "if a kid feels like their teacher cares about them, that they are going to give them a little something extra." Similarly, Ms. Adams also felt that the personal feelings that students might have about a teacher or their class was a legitimate factor that should influence a student's rating of the teacher's instruction. She stated:

> They would rate higher teachers who have a command of a classroom, make learning fun, who engage kids and what they're doing. And (where) they see some relevance, which are all things we (principals) want.

These two participants did not deny that students' personal feelings about a teacher could influence their responses to teacher perception surveys, but they saw this as a potential strength rather than a potential weakness of gathering student perception data from students. Unlike their colleagues, these two administrators viewed students' personal feelings about their instructor as an important element of the data gathered by student surveys, rather than as a potential bias in the data being gathered. As expressed above, these participants felt that if students rated teachers lower due to positive or negative feelings about the teacher, this was a reflection of their lived

experiences as students in that teacher's class and is therefore valuable data for an administrator to use when determining teacher evaluation ratings.

An additional area of concern found in this study was an apparent lack of clarity on the part of administrators regarding how access to student perception survey data would actually lead to any change in administrative actions in the evaluation process. The participants struggled to identify how data collected from student surveys would change administrative actions in teacher evaluation in any way at all. As described in Finding One, administrators believed that principals' access to student perceptions surveys will have a positive impact on current evaluative processes such as administrator-teacher dialogue. However, they could not articulate new evaluative practices in which administrators might engage with data from these surveys.

The only participant who identified a previously unused administrative action that principals might engage in if they had access to student perception survey data was Mr. Jefferson. He felt that administrator evaluative actions might change to include a more customer or client service orientation in evaluation that is not currently present in administrative leadership and supervision. Specifically, he stated:

> From an ideology, it's hard to argue because students are consumers. And so if I'm thinking about a business model, and I want to improve my business, I want to hear from my clients, and I want my clients to give me that feedback.

This customer-service or client-service orientation is not a typical concept applied to educational leadership and supervision, and Mr. Jefferson believed it could introduce new administrative actions in response to student survey data. However, as with those participants who cited an expansion in the area of teacher dialogue, Mr. Jefferson did not provide any specific actions that evaluators would undertake that are not already present in many current evaluative models. His

thoughts on orienting teachers toward a customer or client-service perspective on the teaching and learning process could provide a new lens that would inform current evaluative tasks but stop short of identifying any truly new evaluative actions.

While this one administrator identified a potential new administrative approach to instructional leadership and evaluation, most participants in the study could not identify concrete actions that administrators might undertake if they had access to perception data generated by students. While all 10 participants expressed a desire to have access to this information and felt that it would positively influence a teacher evaluation system in some way, only one referenced a specific way that administrators' behaviors might change as a result of having it. These ideas were still largely vague and unformed. All participants struggled to operationalize these concepts into concrete actions that evaluators would take if they had access to student perception data as instructional leaders.

Finding Number Three: Concerns Related to Survey Implementation

All study participants expressed that they had personal experience surveying students or that they were aware of teachers surveying students. In total, four of the 10 school administrators included in the study had conducted their own surveys of students during their time as teachers in the classroom, and nine of the 10 participants referenced that they were aware of other teachers who surveyed students based on their desire to learn more about their own classroom.

In describing his personal experience with surveying students as a teacher, Mr. Polk described:

> I can't, can't tell you how much it counted towards anything, but I know that I saw them. I know that my boss saw them. I know that the department chair saw them, and we really reflected at that point as, and again, I was a high school math teacher. So as a department

and department to department, like when my teachers are scoring really high in these

areas and what teachers aren't, and at that time it was super unnerving as a new teacher

and it did probably feel bad. It was great to know, like I didn't, I didn't even know like

that guy seems cool, but obviously he's doing something in his room that really is making

a difference.

Similarly, Ms. Tyler described her experience with surveying students at the time that she was

teaching in the classroom as follows:

I chose to do a teacher rating with student surveys. I remember just being in a group,

having a conversation about how does that impact your instruction? How does that

impact you as a teacher?

These personal experiences with student surveys expressed by Mr. Polk and Ms. Tyler

led them to believe that many school districts likely have teachers who are already conducting

surveys of their students. Thus, they feel that if districts simply implement a formal student

survey tied to teacher evaluation without first understanding the level of current surveying of

students by teachers, teachers might resent having to use a new instrument or might not trust an

instrument that is replacing something they perceive to have some value. Ms. Monroe cited an

experience with a past survey she had been required to give: "You get the people who just write,

spiteful, hateful things that you're like, okay, that's not constructive. You didn't give me anything

I can use." Comments like this one are reflective of Ms. Monroe's concerns that if surveys

currently in use by teachers are replaced by district surveys, staff may feel that the district-level

surveys are inferior or untrustworthy as measures of their instructional abilities. This, in turn,

could limit their effectiveness as components of a teacher evaluation system.

The participants in the study also raised concerns that poorly constructed student survey questions could lead to instruments that are not trusted or supported by teachers. All 10 participants expressed that survey questions for elementary students would need to be clearly crafted to solicit the desired information about instruction from young children. Mr. Jefferson stated, "I do think our students, there's a wide range of maturity levels from kindergarten all the way through 12th grade. So with that comes a certain knowledge base that would have to be carefully understood." This is indicative of his concern that surveys will need to be carefully crafted to ensure that students are providing feedback on classroom instruction.

One interesting theme of the information shared by participants as they described the need for thoughtful question selection or design relative to age was concerns regarding middle school students. The participants specifically mentioned the concern that while elementary and high school students may be able to be more objective, middle school students might be developmentally more susceptible to basing an evaluation on their social needs rather than their instructional needs. An example of this was a comment by Ms. Adams:

> I think the difference in middle [school] is [that] developmentally, kids want the social time that if a teacher had lack of preparation and there was more time to socialize, that would not be, that would not come up on a student perception survey. If kids have the time to chat or hang out or be on their phone or play a game on their Chromebook, that teacher will not be rated low.

This comment underscores the concern that all students, especially middle school students, may not provide school administrators with actional information about teacher effectiveness without a thoughtful and tailored question design.

The participants also expressed the implementation concern that teachers may reject and devalue student perception survey data as a component of their overall teacher evaluation score if that component was given too much weight in determining the teacher's final annual evaluation rating. The reasoning expressed was that if scores were made a substantial component of an overall evaluation too early, they would be perceived as too "high-stake" and that teachers would reject them out of fear of their impact. Ms. Harrison stated, "But I think, you know, definitely knowing that trying something out and getting this new vantage point on perspective, it would have to start with a smaller percentage."

Similarly, Ms. Tyler indicated that she also felt that it would be difficult for teachers to accept survey data if it was given too much weight in the evaluative system when first introduced:

> I would probably not rate it very high and kind of get a gist as to how does this pan out.
> Do we get a sense that students are being honest in this? They're really getting a chance
> to share their feelings and thoughts. So I would say 5 to 10%.

Overall, two participants did not provide a percentage, three participants recommended a percentage between 0% and 5%, four participants recommended a percentage between 5% and 8%, and one participant recommended a percentage between 15% and 20%.

Chapter 5

Discussion

This study contributes to our collective understanding of the potential uses of student perception data as part of a comprehensive teacher evaluation strategy. The analysis revealed that administrators see value in the collection of student perception survey data as part of the evaluation process and believe it has the potential to support teacher growth. However, the analysis also revealed that support for this kind of data hinges on the conditions under which it is collected. Administrators were especially concerned that using data from student perception surveys could potentially introduce bias and/or personal experiences that mitigate the value of the data. The findings from this study add to existing academic research and professional information concerning teacher evaluations and assessment tools. While assessments and evaluations may vary from state to state, teacher evaluations may include the use of classroom observations (van der Lans et al., 2016), principal evaluations (Brandt et al., 2007), instructional artifacts (Matsumura et al., 2006), portfolios (Sung et al., 2009), teacher self-report measures (Ford et al., 2017), value-added models (Guarino et al., 2015), and student perception surveys (Geiger & Amrein-Beardsley, 2019). Through an analysis of the data provided by the participants, the findings were organized by research question. The following sections present a discussion of the findings and how the information relates to past research.

The participants saw the greatest value in obtaining information about the classroom experience. This information, they believed, would help teachers improve. This finding supports and largely aligns with the previous research of Wilkerson et al. (2000), who argued that student surveys could be used as reliable measures of teacher performance. Similarly, this study aligns with findings reported by Ferguson (2010), indicating that student perception surveys could

capture essential dimensions of the classroom and teaching practices as students experience them. Thus, in broad strokes, the claims this study makes contribute to the argument supporting the use of student perception surveys in teacher evaluation.

However, the participants also expressed concerns about the use of these data in teacher evaluation. For example, the participants cited concerns about students rating teachers they liked or disliked inaccurately and students failing to grasp the complexities of teaching. These concerns reflect the perceptions of teachers in the research of Kauchak et al. (1985), who voiced doubt that students could provide valid or reliable information about teaching performance, and the finding of Peterson et al. (1998) that students may be limited in their capacity to provide feedback on a teacher's understanding of a certain set of standards or performance objectives. These studies, which involved teachers, closely mirror the administrators in this study, who cited very similar doubts about student objectivity and skill. In all these studies, concerns were most prevalent when surveying students at the elementary level. The reasons for potential bias in student responses expressed in this study mirror almost exactly the reasons found in similar studies with teachers conducted by Schulz et al. (2014) and Marsh (2007). This alignment is perhaps not overly surprising, as eight of the 10 administrators who participated in this study had served as elementary school teachers at some point in their careers.

The participants noted that student surveys could allow administrators access to information from students regarding their experiences in a teacher's class, which would afford principals access to information regarding the students' sense of belonging. This finding extends the study of Finefter-Rosenbluh et al. (2021), where researchers noted that student surveys in diverse student populations help capture the learning process and identify areas in need of improvement. In their research, the participants reported that student feedback surveys could

help teachers identify student struggles, prompting them to adjust their teaching process. Similarly, the finding also supports the work of Cunningham-Nelson et al. (2021) concerning how accessing student comments gives the students a voice and provides educators with a sense of how their lessons are being received. Additionally, the potential improvement in dialogue between administrators and teachers cited by study participants aligns with research conducted in Chicago by Steinberg and Sartain (2015). This study found that structured dialogue between administrators and principals did lead to improved student outcomes.

While the findings in this study align with research about the affordances of using student perception data, there are also some important cautions. The findings urge school districts implementing student perception surveys to begin using them in a low-risk environment until teachers grow comfortable with the data they produce, and evaluators can develop concrete ways of using the data to guide teachers in improving their instruction. These findings are similar to those of Amrein-Beardsley and Geiger (2022), who noted student surveys' benefits but cautioned that more research and evaluation were needed. In addition, van der Lans (2018) and Dretzke et al. (2015) warned that appropriate decisions should be made concerning how the measures of components of teacher evaluations inform one decision, the teacher's final evaluation score.

Recommendations for Practice

The findings of this research study confirm that school administrators at the elementary and district levels see student perception data as a rich source of information about the instruction happening in their classrooms. This understanding should encourage districts to explore the possibility of including student perception surveys in their evaluation systems. Thus, the study offers a few recommendations to practitioners:

Importance of planning. The evidence suggests that in their planning processes, districts will not need to spend extensive time convincing administrators of the value of these data or the potential information about classrooms that it could provide. Process steps related to building consensus for change, establishing the need for change, or building a coalition of willing participants can potentially be relatively short, as it will not take a great deal of effort to convince administrators that the data collected from student perception surveys will have value. Instead, school districts seeking to add student perception surveys can plan to spend much more time and effort in process steps like selecting or developing survey instruments.

Win support for the survey instrument. Districts should understand that administrators are highly concerned about bias and students' understanding of the purpose of surveys. They must spend substantial time selecting the survey instrument, and developing a plan of how surveys will be framed and explained to students will be a critical step of the process that requires substantial time and focus. Interest in collecting survey data will come easily, but faith in the instrument and students' understanding of its purpose will be much harder to earn. Thus, organizations seeking to implement student surveys should plan ample time for surveys to be reviewed and evaluated by multiple stakeholder groups preceding their implementation.

Recognize and manage bias. Concerns about student bias due to personal feelings about instructors, feelings about the content of a course, and students' failure to understand the purpose of the survey should be addressed on an ongoing basis. Work on the front end to address bias concerns is likely to mitigate future concerns surrounding potential bias in the data that will eventually be collected. One concrete step for districts that was revealed

by this research study is that districts looking to address these concerns about bias and student understanding of the purpose of student perceptions surveys may benefit from an initial analysis of the extent to which teachers and principals are already surveying students about instruction within their organization. Many teachers and schools are already gathering student survey data in different ways and at district levels, and the fact that these surveys were individually or locally created may cause them to enjoy some level of inherent trust on the part of the teachers and administrators giving them.

Support principals as evaluators. District-level staff will need to support principals in developing concrete methods of using the data from these surveys to improve instruction once they have the data. This study found ample evidence that while principals believe that strong data can come from student surveys, they have little concept of how to operationalize that data into concrete administrative steps to influence teacher quality. Building-level administrators know they want the insight into how students personally experience instruction that student perception survey data can provide, but principals will struggle with how to use that information to influence instruction. Organizations seeking to implement student perception surveys as a component of their evaluation systems must spend ample time defining these concrete steps that administrators will take with the data because building-level administrators do not inherently know what those steps should be.

Significance and Opportunities for Future Research

Like all reforms, the adoption of student perception surveys as a component of evaluation practices requires careful preparation and planning. Given these concerns related to student bias and the acknowledgment that building leaders will need ample guidance in how to use the data collected, this study offers valuable insights into districts seeking to include student perception

surveys as a component of their overall teacher evaluation system. However, the study also raises important but largely unanswered questions. These, I assert, could be useful in informing future research, which I discuss below:

1. Administrator actions with individualized student data

Quantitative research in the area of student perception studies has predominantly focused on the validity of perception surveys in measuring teacher quality and the correlation between the results of student perception surveys and other measures of teacher effectiveness. Qualitative research has largely focused on teachers' feelings about surveys, how they perceive the results, and their feelings about including those results as part of the teacher evaluation process. However, there has been little scholarly exploration into how the data from student perceptions surveys will be used by administrators once they have been collected and aggregated. Exploring these concrete actions more thoroughly is an important step in understanding the impact that the inclusion of student perception surveys could have on the educational space.

2. Bias in student survey data

The results of this study confirm that school administrators at the elementary level share concerns about student bias that have arisen in studies that investigated teachers' beliefs about the inclusion of student perception surveys in evaluation. There is a legitimate concern that without an ample body of research to counteract the collective belief that students cannot be objective in providing data on instruction, the void will be filled by a general sense that student perception data are substantially influenced by student bias inherent in the process. Initial scholarship in this area has found this to be largely inaccurate (Amrein-Beardsley & Geiger, 2022; Marsh, 2007). However, it is clear that

this research has not reached a "tipping point' that moved the general perceptions of teachers and administrators.

Conclusion

In conclusion, the findings of this study suggest that school districts seeking to add student perception surveys as a component of a comprehensive teacher evaluation system can increase their chances of successful implementation by leveraging support on the part of administrators, thoughtfully designing or selecting survey instruments, developing administrative actions with survey data, and introducing surveys with relatively low influence on the overall score. The efforts of school district personnel seeking to implement surveys could also be supported by further research exploring concrete administrative actions with survey data and additional research regarding bias in student perception survey results.

References

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, *13*(2), 153–166.

Amrein-Beardsley, A., Polasky, S., & Holloway-Libell, J. (2016). Validating "value added" in the primary grades: One district's attempts to increase fairness and inclusivity in its teacher evaluation system. *Educational Assessment, Evaluation and Accountability*, *28*(2), 139–159.

Amrein-Beardsley, A., & Geiger, T. (2022). The breakfast of champions: Student surveys to sustain and cultivate teachers. Phi Delta Kappan, 103(8), 49-53. https://doi.org/10.1177/00317217221100011

Aragon, S. (2018). *Policy snapshot: Teacher evaluations.* Education Commission of the States.

Attinello, J. R., Lare, D., & Waters, F. (2006). The value of teacher portfolios for evaluation and professional growth. *NASSP Bulletin*, *90*(2), 132–152.

Bembry, K. L., Jordan, H. R., Gomez, E., Anderson, M. C., & Mendro, R. L. (1998). Policy implications of long-term teacher effects on student achievement. http://dallasisd.org/depts/inst_research/aer98rm1.html

Blank, R. K., Porter, A., & Smithson, J. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics & science. Results from survey of enacted curriculum project. Final report*. Council of Chief State School Officers.

Borko, H., Stecher, B. M., Alonzo, A. C., Moncure, S., & McClam, S. (2005). Artifact packages

    for characterizing classroom practice: A pilot study. *Educational Assessment*, *10*(2), 73–

    104.

Borko, H., Stecher, B. M., & Kuffner, K. (2007). *Using artifacts to characterize reform-oriented*

    instruction: The Scoop Notebook and rating guide. National Center for Research on

    Evaluation, Standards, and Student Testing (CRESST).

Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). *Examining district*

    guidance to schools on teacher evaluation policies in the Midwest region (Issues &

    Answers Report, REL 2007–No. 030). Regional Educational Laboratory Midwest,

    Institute of Education Sciences, US Department of Education.

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added*

    models. Policy information perspective. Educational Testing Service.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in*

    Psychology, 3(2), 77–101.

    https://www.tandfonline.com/doi/abs/10.1191/1478088706QP063OA

Brown, C., Boser, U., Sargrad, S., & Marchitello, M. (2016). *Implementing the Every Student*

    Succeeds Act: Toward a coherent, aligned assessment system. Center for American

    Progress. https://files.eric.ed.gov/fulltext/ED568079.pdf

Brownell, M. T., Bishop, A. M., & Sindelar, P. T. (2018). Republication of "NCLB and the

    demand for highly qualified teachers: Challenges and solutions for rural schools." *Rural*

    Special Education Quarterly, 37(1), 4–11.

    https://doi.org/10.1177%2F8756870517749604

Camburn, E., & Barnes, C. A. (2004). Assessing the validity of a language arts instruction log

    through triangulation. *The Elementary School Journal*, *105*(1), 49–73.

Cavalluzzo, L. C. (2004). Is national board certification an effective signal of teacher quality?.

    CNA Corporation.

Civic Impulse. (2018). H.R. 1—111th Congress: American Recovery and Reinvestment Act of

    2009. https://www. govtrack.us/congress/bills/111/hr1

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the

    assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778–820.

Cosner, S., Kimball, S. M., Barkowski, E., Carl, B., & Jones, C. (2015). Principal roles, work

    demands, and supports needed to implement new teacher evaluation. *Mid-Western*

    Educational Researcher, *27*(1), 76-95.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among*

    five approaches. Sage.

Croft, M., Guffy, G., & Vitale, D. (2018). *The shrinking use of growth: Teacher evaluation*

    legislation since ESSA. ACT.

Crow, G. M. (2006). Complexity and the beginning principal in the United States: Perspectives

    on socialization. *Journal of Educational Administration*, *44*(4), 310–325.

Cunningham, G. K., & Stone, J. E. (2005). Value-added assessment of teacher quality as an

    alternative to the National Board for Professional Teaching Standards: What recent

    studies say. In R. Lissitz (Ed.), *Value added models in education: Theory and*

    applications (pp. 209–232). JAM Press.

Cunningham-Nelson, S., Laundon, M., & Cathcart, A. (2021). Beyond satisfaction scores: visualising student comments for whole-of-course evaluation. Assessment & Evaluation in Higher Education, 46(5), 685-700.

Custer, S., King, E. M., Atinc, T. M., Read, L., & Sethi, T. (2018). *Toward data-driven education systems: Insights into using information to measure results and manage change*. Center for Universal Education at The Brookings Institution.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives, 8*, 1.

Derrington, M. L. (2016). Implementing teacher evaluation: Lattice of leadership. *Journal of Research on Leadership Education, 11*(2), 181–199. https://doi.org/10.1177%2F1942775116658689

Desimone, L. M., Smith, T. M., & Frisvold, D. E. (2010). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy*, *24*(2), 267–329.

Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Center for American Progress.

Dretzke, B. J., Sheldon, T. D., & Lim, A. (2015). What do K-12 teachers think about including student surveys in their performance ratings? *Mid-Western Educational Researcher*, *27*(3), 185–206.

Dudek, C. M., Reddy, L. A., & Lekwa, A. (2019). Measuring teacher practices to inform student achievement in high poverty schools: A predictive validity study. *Contemporary School Psychology, 23*(3), 290–303. https://doi.org/10.1007/s40688-018-0196-8

Dunford, B. B., Snell, S. A., & Wright, P. M. (2001). *Human resources and the resource based view of the firm* (CAHRS Working Paper No. 01-03). Cornell University, School of

Industrial and Labor Relations, Center for Advanced Human Resource Studies.

http://digitalcommons.ilr.cornell.edu/cahrswp/66

Eisner, E. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice.* Macmillan.

Every Student Succeeds Act of 2015. Pub. L. No. 144-95.

http://dallasisd.org/depts/inst_research/aer98rm1.html

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, *28*(2), 465–498.

Ferguson, R. F. (2010). Peer interaction: The experience of distance students at university level. *Journal of Computer Assisted Learning*, *26*(6), 574–584.

Finefter-Rosenbluh, I., Ryan, T., & Barnes, M. (2021). The impact of student perception surveys on teachers' practice: teacher resistance and struggle in student voice-based assessment initiatives of effective teaching. Teaching and Teacher Education, 106, [103436]. https://doi.org/10.1016/j.tate.2021.103436

Firestone, W. A., & Donaldson, M. L. (2019). Teacher evaluation as data use: What recent research suggests. *Educational Assessment, Evaluation and Accountability, 31*(3), 289–314. https://doi.org/10.1007/s11092-019-09300-z

Fischer, Christian, Fishman, Barry, Dede, Chris, Eisenkraft, Arthur, Frumin, Kim, Foster, Brandon, Lawrenz, Frances, Levy, Abigail Jurist, & McCoy, Ayana. Investigating relationships between school context, teacher professional development, teaching practices, and student achievement in response to a nationwide science reform. Teaching and Teacher Education, 72 (C). Retrieved from https://par.nsf.gov/biblio/10067585. https://doi.org/10.1016/j.tate.2018.02.011

Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017).
Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation
policy in Louisiana. *Educational Policy*, *31*(2), 202–248.

Gallagher, M. A. (2004). A constructivist approach to integrating science, technology, and
engineering into preservice teacher education.

Geiger, T., & Amrein-Beardsley, A. (2019). Student perception surveys for K-12 teacher
evaluation in the United States: A survey of surveys. *Cogent Education*, *6*(1), Article
1602943.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research
synthesis*. National Comprehensive Center for Teacher Quality.
http://wwww.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf

Goldhaber, D. D., & Hansen, M. (2008). *Is it just a bad class?: Assessing the Stability of
Measured Teacher Performance*. Center on Reinventing Public Education.

Goldhaber, D. (2019). Evidence-Based Teacher Preparation: Policy Context and What We
Know. Journal of Teacher Education, 70(2), 90–101.
https://doi.org/10.1177/0022487118800712

Goldschmidt, P., & Congdon, A. (2018). Exploratory analysis of teacher artifacts as evidence of
educator effectiveness implementation fidelity. *Journal of School Administration
Research and Development, 3*(1), 6–18. https://files.eric.ed.gov/fulltext/EJ1190936.pdf

Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). Can value-added measures of
teacher performance be trusted? *Education Finance and Policy*, *10*(1), 117–156.

Guarino, C. M., Stacy, B. W., & Wooldridge, J. M. (2019). Comparing and assessing the
consequences of two different approaches to measuring school effectiveness. *Educational*

Assessment, Evaluation and Accountability, 31(4), 437–463.

https://doi.org/10.1007/s11092-019-09308-5

Hakel, M. D., Koenig, J. A., & Elliott, S. W. (2008). *Committee on Evaluation of Teacher Certification by the National Board for Professional Teaching Standards, assessing accomplished teaching: Advanced-level certification programs*. National Research Council.

Hallinger, P., Gümüş, S., & Bellibaş, M. Ş. (2020). 'Are principals instructional leaders yet?' A science map of the knowledge base on instructional leadership, 1940–2018. *Scientometrics*, *122*(3), 1629–1650. https://doi.org/10.1007/s11192-020-03360-5

Hanushek, E. A., & Woessmann, L. (2020). Education, knowledge capital, and economic growth. In S. Bradley & C. Green (Eds.), *The economics of education* (pp. 171–182). Academic Press. https://doi.org/10.1016/B978-0-12-815391-8.00014-8

Hargreaves, A., & Braun, H. (2013). *Data-driven improvement and accountability.* National Education Policy Center. http://nepc.colorado.edu/publication/data-driven-improvement-accountability/

Harris, D. N., & Sass, T. R. (2007). *Teacher training, teacher quality, and student achievement* (Working Paper No. 3). National Center for Analysis of Longitudinal Data in Education Research.

Heinrich, C. J., & Good, A. (2018). Informed practice improvements: exploring linkages between school district use of research evidence and educational outcomes over time. *School Effectiveness and School Improvement*, *29*(3), 418–445.

Henderson, E. S. (1984). *Independent learning in higher education*. Educational Technology.

Heneman, H. G., Milanowski, A., Kimball, S., & Odden, A. (2006). Standards-based teacher evaluation as a foundation for knowledge-and skill-based pay. https://repository.upenn.edu/cpre_policybriefs/33

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

Hershberg, T. (2005). Value-added assessment and systemic reform: A response to the challenge of human capital development. *Phi Delta Kappan*, *87*(4), 276–283.

Hunter, S. B., & Rodriguez, L. A. (2021). Examining the demands of teacher evaluation: time use, strain and turnover among Tennessee school administrators. *Journal of Educational Administration*, *59*(6), 739-758.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of labor Economics*, *26*(1), 101–136.

Jiang, J. Y., & Sporte, S. E. (2016). Teacher Evaluation in Chicago: Differences in Observation and Value-Added Scores by Teacher, Student, and School Characteristics. Research Report. *University of Chicago Consortium on School Research*.

Johnson, D. M., & Vidulich, R. N. (1956). Experimental manipulation of the halo effect. *Journal of Applied Psychology*, *40*(2), 130.

Johnson, R. L., McDaniel, F., & Willeke, M. J. (2000). Using portfolios in program evaluation: An investigation of interrater reliability. *American Journal of Evaluation*, *21*(1), 65–80.

Jordan, H., Mendro, R., & Weerasinghe, D. (1997). *Teacher effects on longitudinal student achievement* [Conference presentation]. Sixth Annual Evaluation Institute sponsored by CREATE, Indianapolis.

Kane, T. (2012). Capturing the dimensions of effective teaching: Student achievement gains, student surveys, and classroom observations. *Education Next*, *12*(4), 34–42.

Kane, T., & Cantrell, S. (2010). *Learning about teaching: Initial findings from the measures of effective teaching project*. MET Project, Bill & Melinda Gates Foundation.

Kauchak, D., Peterson, K., & Driscoll, A. (1985). An interview study of teachers' attitudes toward teacher evaluation practices. *Journal of Research & Development in Education*, *19*(1), 32–37.

Koretz, D. (2008). A measured approach. *American Educator*, *32*(2), 18–39.

Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, *46*(5), 234–249.

Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review*, *18*(1), 59-63.

Labuschagne, A. (2003). Qualitative research-airy fairy or fundamental? *The Qualitative Report*, *8*(1), 100–103.

LaFee, S. (2014). Students evaluating teachers. *The Education Digest*, *80*(3), 4.

Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance: Methods, theory, and applications*. Academic Press.

Lipscomb, S., Teh, B. R., Gill, B., Chiang, H., & Owens, A. (2010). *Teacher and principal value-added: Research findings and implementation practices. Final Report.* Mathematica Policy Research, Inc.

Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness*. National Comprehensive Center for Teacher Quality.

Liu, K., Springer, J., Stuit, D., Lindsay, J., & Wan, Y. (2015). *The utility of teacher and student surveys in principal evaluations: An empirical investigation.* https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/REL_2016132.pdf

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, *44*(1), 47–67.

Loup, K., Garland, J., Ellett, C., & Rugutt, J. (1996). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest school districts. *Journal of Personnel Evaluation in Education*, *10*, 203–226.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319–383). Springer.

Martínez, J. F., Schweig, J., & Goldschmidt, P. (2016). Approaches for combining multiple measures of teacher performance: Reliability, validity, and implications for evaluation policy. *Educational Evaluation and Policy Analysis, 38*(4), 738–756. https://doi.org/10.3102%2F0162373716666166

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the instructional quality assessment. CSE Technical Report 681*. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Measures of Effective Teaching. (2012). *Learning about Teaching: Initial findings from the Measures of Effective Teaching Project.*

http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. P*eabody Journal of Education*, *79*(4), 33–53.

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). Authentic intellectual work and standardized tests: Conflict or coexistence? *Improving Chicago's Schools*.

No Child Left Behind Act of 2001. Pub. L. No. 107-110, 20 U.S.C. § 6319.

Nordhaug, O., & Gronhaug, K. (1992). Strategy and competence in firms. *European Management Journal*, *10*(4), 438–442.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Ouellett, M. L. (2007). Your teaching portfolio: Strategies for initiating and documenting growth and development. *Journal of Management Education*, *31*(3), 421–433.

Painter, B. (2001). Using teaching portfolios. *Educational Leadership*, *58*(5), 31–34.

Peterson, K. D., Stevens, D., & Ponzio, R. C. (1998). Variable data sources in teacher evaluation. *Journal of Research and Development in Education*, *31*(3), 123–132.

Peterson, K. D., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, *14*(2), 135–153.

Polgampa, A. S. V., & Huang, F. (2017). Teachers are the propellers in bolstering students outcomes: Review of efficacy of science teachers. *European Journal of Education Studies*, *3*(6), 701–716.

Polikoff, M. S. (2015). Does the test matter: Evaluating teachers when tests differ in their

    sensitivity to instruction. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing*

    teacher evaluation systems: New guidance from the measures of effective teaching

    project (pp. 278–302). Jossey-Bass.

Ravitch, D. (Ed.). (2010). *Debating the future of American education: Do we meet national*

    standards and assessments? Brookings Institution Press.

Reid, D. B. (2017). US principals' interpretation and implementation of teacher evaluation

    policies. *The Qualitative Report, 22*(5), 1457. http://nsuworks.nova.edu/tqr/vol22/iss5/19

Renaud, R. D., & Murray, H. G. (2005). Factorial validity of student ratings of instruction.

    Research in Higher Education, *46*(8), 929–953.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student

    achievement. *The Quarterly Journal of Economics*, *125*(1), 175–214.

Rothstein, J. (2016). Can value-added models identify teachers' impacts? *UC Berkeley: Institute*

    for Research on Labor and Employment. https://escholarship.org/uc/item/3xt316zw

Rousmaniere, K. (2007). Presidential address: Go to the principal's office: Toward a social

    history of the school principal in North America. *History of Education Quarterly*, *47*(1),

    1–22.

Saldana, J. (2009). *The Coding Manual for Qualitative Researchers* (3rd ed.). Sage.

Sanders, W. L., Ashton, J. J., & Wright, S. P. (2005). *Comparison of the effects of NBPTS*

    certified teachers with other teachers on the rate of student academic progress. Final

    Report. National Board for Professional Teaching Standards.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* University of Tennessee Value-Added Research and Assessment Center.

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

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation. Research Report*. Consortium on Chicago School Research.

Schulz, J., Sud, G., & Crowe, B. (2014). *Lessons from the field: The role of student surveys in teacher evaluation and development*. Bellwether Education Partners.

Schwab, R. L., & Iwanicki, E. F. (1988). Merit pay: Stressor or motivator? https://files.eric.ed.gov/fulltext/ED303848.pdf

Simkin, L., Charner, I., & Suss, L. (2010). *Emerging education issues: Findings from the Wallace Foundation survey.* The Wallace Foundation.

Stake, R. E. (1995). *The art of case study research*. Sage.

Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching project. *Education Finance and Policy*, *10*(4), 535-572.

Steinberg, M. P., & Sartain, L. (2021). What explains the race gap in teacher performance ratings? Evidence from Chicago Public Schools. *Educational Evaluation and Policy Analysis*, *43*(1), 60-82.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques and procedures for developing grounded theory*. Sage.

Strauss, V. (2013). *Why the 'learning pyramid' is wrong*. The Washington Post. https://www.washingtonpost.com/news/answer-sheet/wp/2013/03/06/why-the-learning-pyramid-is-wrong/

Stronge, J. H. (2018). *Qualities of effective teachers*. ASCD.

Stosich, E. L., Bae, S., & Snyder, J. (2018). Redesigning assessment and accountability: An introduction. *Education Policy Analysis Archives, 26,* 7. http://dx.doi.org/10.14507/epaa.26.2906

Sung, Y. T., Chang, K. E., Yu, W. C., & Chang, T. H. (2009). Enhancing teachers' learning and reflection through structured digital portfolios. *Journal of Computer Assisted Learning*, *25*, 375–385.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, *102*(7), 3628–3651.

Temam, S., Billaudeau, N., & Vercambre, M. N. (2019). Overall and work-related well-being of teachers in socially disadvantaged schools: A population-based study of French teachers. *BMJ Open*, *9*(9), e030171.

Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education. Education Sector Reports*. Education Sector.

Tucker, P. D., Stronge, J. H., Gareis, C. R., & Beers, C. S. (2003). The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference? *Educational Administration Quarterly*, *39*(5), 572–602.

United States Department of Education. (2009). *Race to the top program: Executive summary*.

    https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

van der Lans, R. M., van de Grift, W. J., van Veen, K., & Fokkens-Bruinsma, M. (2016). Once is

    not enough: Establishing reliability criteria for feedback and evaluation decisions based

    on classroom observations. *Studies in Educational Evaluation*, *50*, 88–95.

van der Lans, R., van de Grift, W., & van Veen, K. (2018). Developing an Instrument for

    Teacher Feedback: Using the Rasch Model to Explore Teachers' Development of

    Effective Teaching Strategies and Behaviors. Journal of experimental education, 86(2),

    247-264. https://doi.org/10.1080/00220973.2016.1268086

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample

    size sufficiency in interview-based studies: systematic analysis of qualitative health

    research over a 15-year period. BMC medical research methodology, 18(1), 148.

    https://doi.org/10.1186/s12874-018-0594-7

Wallace, T. L., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us

    about teaching? Empirically testing the underlying structure of the tripod student

    perception survey. *American Educational Research Journal*, *53*(6), 1834–1868.

Wechsler, M., Tiffany-Morales, J., Campbell, A., Humphrey, D., Kim, D., Shields, P., & Wang,

    H. (2007). The *status of the teaching profession 2007*. The Center for the Future of

    Teaching and Learning.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national*

    failure to acknowledge and act on differences in teacher effectiveness. Executive

    Summary. Second Edition. New Teacher Project.

Wessel-Powell, C., Buchholz, B. A., & Brownell, C. J. (2019). Polic(y)ing time and curriculum: how teachers critically negotiate restrictive policies. *English Teaching: Practice & Critique*, *18*(2), 170–187. https://doi.org/10.1108/ETPC-12-2018-0116

Whitehurst, G. J. (2002). *Raising student achievement: The evidence of high quality teaching* [Conference presentation]. Standards-based Teacher Education Project 2002 Summer Conference.

Wilkerson, D. J., Manatt, R. P., Rogers, M. A., & Maughan, R. (2000). Validation of student, principal, and self-ratings in 360 feedback® for teacher evaluation. *Journal of Personnel*

Wright, P. M., & McMahan, G. C. (1992). Theoretical perspectives for strategic human resource management. *Journal of Management*, *18*(2), 295–320.

Yin, R. K. (1984). *Case study research: Design and Methods*. Sage Publications.

Appendix A

Research Timeline

The following timeline will guide the conduct of the research but is designed to be fluid. While

all efforts will be made to proceed based on the timeline, it may be adjusted as the research

unfolds.

Month 1	Participant Recruitment Conduct 1 or 2 Initial Participant Interviews Begin Transcription of the Initial Interviews Observe Any Relevant District Evaluation Meetings/Generate Field Notes Begin Document Collection
Month 2	Complete Remaining Initial Participant Interviews Continue Transcription of the Initial Interviews Observe Any Relevant District Evaluation Meetings/Generate Field Notes Begin Document Collection
Month 3	Continue Transcription of the Interviews Begin Coding of the Initial Interviews/Field Notes/Documents
Month 4-5	Complete Coding of the Initial Interviews Begin Coding of the Initial Interviews/Field Notes/Documents Develop Follow-up Interview Questions
Month 5-6	Conduct 3 to 5 Follow-up Interviews Begin Transcription of the Follow-up Interviews Begin Coding of the Final Interviews
Month 7	Complete the Follow-up Interviews Complete the Transcription of the Follow-up Interviews Complete the Coding of the Final Interviews
Month 8-10	Analysis and Conclusions

Characteristics of the Currently Available Vendor-developed SPSs

Reprinted from Geiger and Amrein-Beardsley (2019).

	iKnow MyClass Survey	K12 Insight Engage Survey	My Student Survey STeP Survey	Panorama Survey	Tripod Survey	YouthTruth Student Survey
	Characteristics of the Currently Available Vendor-developed SPSs					
Parent Company	Corwin (SAGE)	K12 Insight	My Student Survey	Panorama Education	Tripod Survey	YouthTruth
Development	Dr. Russell Quaglia (Quaglia Institute for Student Aspirations)	Unknown	Dr. Ryan Balch (Vanderbilt University for Doctoral Dissertation)	Dr. Hunter Gelbach and Research team (Harvard Graduate School of Education)	Dr. Ronald Ferguson (Harvard University Lecturer on Public Policy)	Phil Buchanan, Ellie Buteau, Threlfall (Center for Elective Philanthropy)
Highest Degree of Lead Developer	Dr. Russell Quaglia: Ed.D., Educational Administration, Columbia University	Unknown	Dr. Ryan Balch: Ph.D., Leadership and Policy Studies, Vanderbilt University	Dr. Hunter Gelbach: Ph.D., Educational Psychology, Stanford	Dr. Ronald Ferguson: Ph.D., Economics, Massachusetts Institute of Technology	Phil Buchanan: MBA, Harvard University; Ellie Buteau: Ph.D., Social-Personality Psychology Valerie Threlfall: MBA, MPP, Northwestern University, Harvard University

Cost	$250/school plus $2.50 per student	Unknown	Unknown	$2.50 per student	$3/classroom plus $50/school ($2000 minimum)	Unknown
Constructs/Domains	Class Efficacy, Cooperative Learning, Critical Thinking, Discipline Problems, Engagement, Positive Pedagogy, Relationships, and Relevance	Unknown	Coach, Content Expert, Counselor, Manager, Motivator, and Presenter	Classroom Climate, Classroom Engagement, Classroom Learning Strategies, Classroom Mindset (Dispositional Behavior), Pedagogical Effectiveness, Classroom Rigorous Expectations, Classroom Belonging, Classroom Teacher-Student Relationships, and Valuing of the Subject	Captivate, Care, Challenge, Clarify, Classroom Management, Confer, and Consolidate	Academic Rigor and Expectations, Classroom Culture, Instructional Methods, Personal Relationships, Relevance, and Student Engagement
Grade level(s)	Grades 3–5; Grades 6–12 (short and long versions)	Unknown	Grades 6–12	Grades 3–5; Grades 6–12	Grades K-2; Grades 3–5; Grades 6–12	Grades 3–5; Grades 6–12
Number of	Grades 3–5	Unknown	Unknown	Grades 3–5	Grades K-2: 36;	Grades 3–5: 25;

Items				survey: 43; Grades 6–12 survey: 50	Grades 3–5: 36; Grades 6–12: 36	Grades 6–12: 30
survey: 27; Grades 6–12 survey: 20 (short), 50 (long)						
Response Options	Five-point Likert scale	Unknown	Five-point Likert scale	Five-point Likert scale; Scale adjusted for each content item	Five-point Likert scale	Grades 3–5: Three-point Likert scale; Grades 6–12: Five-point Likert scale
Online Administration	Yes	Unknown	Yes	Yes	Yes	Yes
Paper Administration	No	Unknown	Yes	Yes	Yes	Per Request
Main website	iKnowMyClass Survey	K12 Insight Engage Survey	My Student Survey STeP Survey	Panorama Survey	Tripod Survey	YouthTruth Student Survey
Response When Contacted	Provided survey items and additional documents	Refused to provide survey or additional information; cited its proprietary nature	No response	N/A; Survey items and additional documents publicly shared online	Required approved IRB proposal and use of data for student's dissertation to share survey items	Provided survey items and additional documents

Appendix C

Interview Protocol Form

Introduction

My name is Matthew Kaiser, a doctoral candidate student from Indiana University who is studying student perception surveys as a component of teacher evaluations. In this interview, I am going to ask you about your perceptions of the strengths, weaknesses, and potential opportunities of student perceptions as part of teacher evaluations. As we talk, I would like you to consider how the integration of student perspectives in teacher evaluations will affect an administrative approach to instructional leadership.

To facilitate notetaking, I would like to digitally record our conversation today. For your information, only researchers on the project will be privy to the recordings, which will be eventually destroyed after they are transcribed. All information will be held confidential, and your participation is voluntary. You may withdraw your participation at any time if you feel uncomfortable. Thank you for agreeing to participate.

This interview should take no more than 60 minutes. With your permission, I will record the interview. Do I have your permission to begin the recording?

You have been selected to speak with us today because you have been identified as someone who has a lot of knowledge to share about teaching, learning, and evaluation in this district. My research project focuses on the inclusion of student perception surveys into teacher evaluation instruments and administrator perceptions of that practice. My study does not aim to evaluate your techniques or experiences. Rather, I am trying to learn more about teaching and evaluation and hopefully learn about administrator practices that evaluate teacher quality.

1. Interviewee Background

    We will begin with a few questions about you.

    a. How long have you been in your present position?
    b. How long have you been in this school district?
    c. What is your highest degree?
    d. What did you teach in the classroom?
2. Evaluation Information
    a. Briefly describe your role in developing the current MSDWT evaluation system.

        Probes: How are you currently involved in the assessment in the MSDWT?

    b. Approximately how many staff do you currently evaluate as a primary or secondary evaluator?
    c. What motivates you to use the teacher evaluation system?
3. Teacher Evaluation Background
    a. How do you use the teacher evaluation system as a formative tool for teachers?

        Probes: Is it working? Why or why not?

b. What resources are available to administrators when applying the teacher evaluation system?
c. What rewards do administrators receive from the district for engaging in unique evaluative practices?
d. What resistance have you or your colleagues encountered to the current teacher evaluation system?

4. Teacher Evaluation Methods
a. What personal characteristics do you feel exceptional teachers share?
b. What professional characteristics do you feel exceptional teachers share?
c. To what extent does the current teacher evaluation system in MSDWT allow you to evaluate these characteristics?
d. How has the observational protocol in the district changed over the course of your time as an administrator in the district?

   Probe: Have the changes been positive?

e. What portions of the rubric tell you the most about the quality of a teacher?
f. What measures are taken in your building to ensure that observations are consistent across raters? In the district?

   Probe: How effective are these measures?

g. How do artifacts posted by teachers inform their evaluations?
h. Are teacher scores generally higher or lower than your observational scores for them?
i. To what extent do you feel administrators can trust artifacts to tell them the quality of their teachers.

5. Primary and Secondary Student Measures
a. Which primary and secondary teacher measures best differentiate quality teachers from average teachers in the district?
b. Are teacher primary and secondary measure scores generally higher or lower than your observational rubric scores?
c. In what ways could primary and secondary measures in the district be improved?

6. Student Perception Surveys in Evaluation
a. Have you ever worked in a school or district that incorporated student surveys as part of teacher evaluations?
b. Describe any conversations that you have had with others in education about including student perception surveys in evaluation.
c. Could student perception surveys provide insight into teachers as instructional planners or their implementation of professional development?
d. What, if anything, do you think students at your level could tell you about teacher quality?
e. Probe: At other levels in the district?
f. Would students rate teachers they like higher than teachers they do not like?
g. If you conducted student perception surveys in your building, do you think the information gathered would align with your classroom observations of teachers?
h. If you conducted student perception surveys in your building, do you think the information gathered would align with your classroom observations of teachers?

   i. If you conducted student perception surveys in your building, do you think the information gathered would align with the teacher primary and secondary measure scores?

   j. What would be the appropriate weight for student perception surveys if they were added to our current evaluation system?

7. Post-interview Comments and/or Observations:

Appendix D

Indiana University Informed Consent Statement for Social Behavioral Research

Student Perception Surveys as a Component of a Comprehensive Teacher Evaluation System: A Case Study of Elementary Principals

IRB Protocol Number:

ABOUT THIS RESEARCH

You are being asked to participate in a research study. Scientists do research to answer important questions that might help change or improve the way we do things in the future.

This consent form will give you information about the study to help you decide whether you want to participate. Please read this form, and ask any questions you have before agreeing to participate in the study.

PARTICIPATION IN THIS STUDY IS VOLUNTARY

You may choose not to participate in the study or to leave the study at any time. Deciding not to participate or deciding to leave the study later will not result in any penalty or loss of benefits to which you are entitled and will not affect your relationship with your college or university.

WHY IS THIS STUDY BEING DONE?

The purpose of this research study is to understand what elementary school principals in one Midwestern school district believe are the relative strengths, weaknesses, and potential affordances of using student perception within the context of the teacher evaluation system in their district.

HOW MANY PEOPLE WILL PARTICIPATE?

If you agree to participate, you will be one of up to 10 participants in this study.

WHAT WILL HAPPEN DURING THE STUDY?

If you agree to be in the study, you will do the following:
- Participate in one semi-structured interview of up to 60 minutes, which will be conducted via Zoom and recorded with your permission.
- Provide copies of documents that are relevant to the teacher evaluation process in your school district. The documents may include informal and formal observations of teachers, evaluation guidebooks, teacher evaluation rubrics, and other documents that may prove useful and relevant to the study's topic.
- Review drafts of the completed research to provide feedback, guidance, and/or verification that the researcher has correctly represented your perspectives in the data.

WHAT ARE THE RISKS OF PARTICIPATION IN THE STUDY?

While participating in the study, you are not expected to experience any significant risks. However, in rare cases, a participant in a qualitative research study may experience nervousness, agitation, or anxiety while being interviewed by the researcher or observed in their professional work setting. If this should occur, please tell me that you feel uncomfortable or that you do not want to answer a question. You are not required to participate in any aspect of this research that makes you feel uncomfortable.

WHAT ARE THE POTENTIAL BENEFITS OF PARTICIPATION IN THE STUDY?

We do not expect you to receive any direct benefit from participating in this study, but we hope to learn things that will help scientists in the future. Your feedback may help future administrators create more effective measures for educational evaluation.

WILL I RECEIVE MY RESULTS?

The results from this research study may appear in published research articles, book chapters, books, and other scholarly outlets. In addition, you may request a copy of your completed interview transcript.

HOW WILL MY INFORMATION BE PROTECTED?

Efforts will be made to keep your personal information confidential. We cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. No information that could identify you will be shared in publications about this study.

Organizations that may inspect and/or copy your research records for quality assurance and data analysis include groups such as the study investigator and their research associates, the Indiana University Institutional Review Board or its designees, and state or federal agencies that may need to access the research records (as allowed by law).

WILL MY INFORMATION BE USED FOR RESEARCH IN THE FUTURE?

Information collected from you for this study may be used for future research studies or shared with other researchers for future research. If this happens, information that could identify you will be removed before any information is shared. Since identifying information will be removed, we will not ask for your additional consent.

WILL I BE PAID FOR PARTICIPATION?

You will not be paid for participating in this study.

WILL IT COST ME ANYTHING TO PARTICIPATE?

There is no cost to you for participating in this study.

WHO WILL PAY FOR MY TREATMENT IF I AM INJURED?

In the event of physical injury resulting from your participation in this study, necessary medical treatment will be provided to you and billed as part of your medical expenses. Costs not covered by your health-care insurer will be your responsibility. In addition, it is your responsibility to determine the extent of your health-care coverage. There is no program in place for other monetary compensation for such injuries. However, you are not giving up any legal rights or benefits to which you are otherwise entitled. If you are participating in research that is not conducted at a medical facility, you will be responsible for seeking medical care and for the expenses associated with any care received.

WHO SHOULD I CALL WITH QUESTIONS OR PROBLEMS?

For questions about the study, contact the researcher, Matthew Kaiser, at 317-340-0749, or Dr. Chad Lochmiller at (812) 856-8235. If you cannot reach the researcher during regular business hours (i.e., 8:00 a.m. to 5:00 p.m.), please contact the IU Human Subjects Office at 800-696-2949 or at irb@iu.edu.

For questions about your rights as a research participant; to discuss problems, complaints, or concerns about a research study; or to obtain information or to offer input, please contact the IU Human Subjects Office at 800-696-2949 or at irb@iu.edu.

CAN I WITHDRAW FROM THE STUDY?

If you decide to participate in this study, you can change your mind and decide to leave the study at any time in the future. The study team will help you withdraw from the study safely. If you decide to withdraw, please contact Matthew Kaiser at 317-340-0749.

PARTICIPANT'S CONSENT

In consideration of all of the above, I give my consent to participate in this research study. I will be given a copy of this informed consent document to keep for my records. I agree to participate in this study.

Participant's Printed Name: _____

Participant's Signature:_____**Date**:_____

Printed Name of the Person Obtaining Consent:_____

Signature of the Person Obtaining Consent:_____**Date**:_____

Appendix E

Interview Questions

Qualitative Interview Study

1. Describe the responsibilities you have regarding teacher evaluation?

   Probe: Have these responsibilities changed at all?

   Probe: If so, how?

2. How would you describe an exceptional teacher?

   Probe: How does this teacher's practice differ from those of others?

3. How do you use the evaluation system to help teachers grow and improve their practices?

4. What do you see as benefits and drawbacks of the current teacher evaluation systems?

5. What potential improvements could be made?

   Probe: How would you recommend that we introduce these types of improvements?

6. If we were to ask students to reflect on their teachers practice what would they say?

   Probe: How do you think their responses would vary?

   Probe: How would you explain the variance?

7. If you were to ask a student about their teacher, what do you think would matter the most to them?

   Probe: What would they say about instruction?

   Probe: How would a student describe good instruction?

   Probe: How would students judge the difference between good and bad instructions?

   Probe: Does the age of the student impact your thinking on this subject?

8. Would students rate teachers they like higher than teachers they do not like?

9. Thinking from the perspective of a teacher, what do you think they would see as potential benefits or drawbacks to including student perceptions in their evaluations?

    Probe: How might this look different from an effective teacher to an ineffective teacher?

10. What do you see as the potential benefits of including student perceptions in teacher evaluations?

    Probe: How do you think your staff would react to the change?

Appendix F

Documentation for Data Collection and Analysis

Document Collected	Research Question 1 How should school districts incorporate student perception surveys as a component of their comprehensive teacher evaluation system?	Research Question 2 To what extent do principals believe that student perception surveys will produce accurate and actionable information to improve instruction?	Research Question 3 To what extent do principals believe student perception surveys align with their instructional observation ratings of teachers?	Research Question 4 What implications for the use of student perception surveys in teacher evaluation are uncovered by exploring administrator perceptions?
Informal Observation Scripting, Ratings, and Comments	X		X	X
District Teacher Evaluation Guidebook	X			
Teacher Evaluation Rubric	X	X		
Evaluator Briefs	X			
HR Correspondence to Teachers Regarding Evaluation	X			
Yearly Teacher Evaluation PowerPoint	X	X	X	X

Appendix G

Document Collection Protocol

To: Research Participants

From: Matthew Kaiser
         Doctoral Candidate Student, K-12 Educational Leadership and Policy
         Indiana University

Subj: Documents Requested for Research Study

The purpose of this research study is to understand what elementary school principals in one Midwestern school district believe are the relative strengths, weaknesses, and potential affordances of using student perception within the context of the teacher evaluation system in their district. I am requesting that you provide copies of the following documents to assist me with my research. The documentation you provide should not contain identifiable information. The information will be stored on a secure server hosted by Indiana University.

The types of documents to be collected are derived from mathematics and science education courses and include:

- Informal observations
- Formal observations or observation forms
- Evaluation guides or guidebooks
- Teacher evaluation rubrics
- Other documents, forms, or guides that are relevant to teacher evaluations

The information can be sent to Matthew Kaiser via email as a PDF document. My contact information is provided below:

     Matthew Kaiser
     matkaise@iu.edu

We sincerely appreciate your assistance with this research study. If you have questions or concerns, please contact Matthew Kaiser.

Appendix H

Observation Guide

Description of the Observational Approach

The research team will conduct classroom observations. Observations will occur during class meetings and other public activities. The observations will seek to corroborate what participants describe in the context of the semi-structured interviews and what we identify in the documents. During observations, members of the research team will record their observations on a laptop computer. The notes will be saved electronically and uploaded to a computer-assisted qualitative data analysis software package. Files will be stored without reference to individual participants. Participant names will be replaced with generic identifiers (e.g., faculty member, undergraduate student, and graduate student).

Classroom Observations

The purpose of classroom observations is to document what measures are used throughout the teacher observation by the administration.

Specific data points may include:

- discussions of lessons between administrator and teacher
- observations of administrative actions during an observation

Appendix I

Recruitment Email

Recruitment Email Message

Dear _____,

I am a doctoral candidate student at Indiana University, conducting a qualitative research study that seeks to understand what elementary school principals in one Midwestern school district believe are the relative strengths, weaknesses, and potential affordances of using student perception in the context of the teacher evaluation system in their district. To complete this study, I am engaging with elementary administrators with experience in teacher evaluations.

I identified you as a potential participant because of your administrative role in the identified geographical location. I would like to conduct an interview with you for up to 60 minutes. I may request an additional follow-up conversation with you for further clarification if needed. In addition, I may ask you to provide copies of artifacts from your observations (e.g., observation forms, informal observations, and evaluation rubrics) and/or to allow me to observe your teacher evaluation process.

I have attached a study information sheet that provides additional information about your rights as a research participant. If you are interested in participating in this study, please let me know and provide a few schedules that might work for you to complete this interview.

Thank you for your consideration.

Matthew Kaiser
Indiana University

MATTHEW MICHAEL KAISER

EDUCATION

INDIANA UNIVERSITY EXECUTIVE ED DOCTORATE   MAY 2023

INDIANA UNIVERSITY, BLOOMINGTON INDIANA

MASTERS IN STRATEGIC MANAGEMENT   MAY 2015

INDIANA UNIVERSITY, BLOOMINGTON INDIANA

EDUCATIONAL SPECIALIST   MAY 2015

INDIANA UNIVERSITY, BLOOMINGTON INDIANA

MASTERS IN EDUCATIONAL LEADERSHIP   DECEMBER 2002

UNIVERSITY OF DAYTON, DAYTON, OHIO

BACHELOR OF SCIENCE IN ELEMENTARY EDUCATION AND HISTORY

MINOR   MAY 1998

MIAMI UNIVERSITY, OXFORD OHIO

LICENSE

INDIANA SUPERINTENDENT LICENSE (P-12)

INDIANA BUILDING LEVEL ADMINISTRATORS LICENSE FOR ALL SCHOOLS

WORK EXPERIENCE

DIRECTOR OF HUMAN RESOURCES

METROPOLITAN SCHOOL DISTRICT OF WASHINGTON TOWNSHIP   2019- Present

- **Hiring:** Attracted and hired talented teaching staff that have become critical building assets
- **Staff Development:** Developed clear and manageable systems to measure the use and effectiveness of the Marzano Vocabulary strategies building-wide
- **Teacher Evaluation:** Assisted in the development of a teacher evaluation tool to be used as an alternative to the RISE framework

PRINCIPAL
METROPOLITAN SCHOOL DISTRICT OF WASHINGTON TOWNSHIP
NORTHVIEW MIDDLE SCHOOL   2014-2019
EASTWOOD MIDDLE SCHOOL   2009- 2014

- **Supervision:** Supervise all certified and non-certified staff in a building comprised of 60+ staff and 800+ students
- **Hiring:** Attracted and hired talented teaching staff that have become critical building assets
- **Professional Learning Communities:** Developed a systematic and documented method of intervention utilized by all staff building-wide
- **Classroom Walkthroughs:** Implemented a system of classroom walkthroughs utilizing technology adopted in other district schools
- **Staff Development:** Developed clear and manageable systems to measure the use and effectiveness of Marzano Vocabulary strategies building wide
- **Teacher Evaluation:** Assisted in the development of a teacher evaluation tool to be used as an alternative to the RISE framework
- **School Calendar:** Represented building-level administrators in a committee to develop an alternative school calendar
- **IB Accreditation:** Led the building to achieve accredation as an IB World School
- **Teacher/ Student Hours:** Worked with teaching and administrative staff to develop new teacher and student hours
- **Schools to Watch:** Eastwood was named a School to Watch by the National Forum in 2011
- **Building Morale:** Created a culture of collaboration and support leading to the most positive climate audit in building history

CENTER GROVE COMMUNITY SCHOOL CORPORATION

CENTER GROVE MIDDLE SCHOOL NORTH   2007–2009

- **Supervision:** Supervise all certified and non-certified staff in a building comprised of 70+ staff and 850+ students
- **Reading Strategies:** Trained staff in the implementation of reading strategies and developed methods of tracking reading strategy usage
- **North P.R.I.D.E:** Created a program to recognize and reward students displaying exceptional character

- **Thursday Night/ Saturday School Parent Initiative:** Developed and maintained a tutoring program for unmotivated students and their parents
- **Classroom Walkthroughs:** Extensively used classroom walkthrough experiences to guide building instruction
- **Curriculum Mapping:** Assisted staff in the development, maintenance, and use of diaries, collaborative, and consensus curriculum maps
- **NCA Accreditation:** Served as a building committee leader and the corporation site visit contact
- **Professional Learning Communities:** Led staff in the development and implementation of multiple common assessments and facilitated the implementation of PLC activities
- **Attendance:** Implemented attendance procedures that led to the highest attendance rates in school history
- **Adequate Yearly Progress:** Led building staff to achieve AYP for the first time in school history

ASSISTANT PRINCIPAL
BROWNSBURG COMMUNITY SCHOOL CORPORATION
BROWNSBURG EAST MIDDLE SCHOOL   2005- 2007
- **Master Schedule:** Built, scheduled, and maintained a modified block master schedule for over 950 students
- **Supervision:** Supervised, observed, evaluated, and monitored the professional growth of certified and non-certified staff
- **Understanding by Design:** Utilized own training to lead the middle school math departments in rewriting the corporation curriculum in the UBD format
- **Mastery Learning:** Worked with groups of staff to incorporate mastery learning techniques into their planning and assessment
- **Service Learning Grant Recipient:** Located, applied for, and received funds through Learn and Serve Indiana to develop a corporation-wide service learning program
- **ENL:** Oversaw committees of teachers and parents that identified new ways to assist the growing ENL population both in and out of the classroom
- **Zeroes Aren't Permitted:** Developed and implemented a building-wide program to require every student to turn in 100% of their work in every class
- **Teaching with Love and Logic:** Guided staff through the process of moving from a standards-based school-wide disciplinary model to one based on guiding principles
- **FLAP Grant Recipient:** Personally wrote and was awarded a federal grant that made the creation of a corporation Chinese program possible

WORK EXPERIENCE CONTINUED

ATHLETIC DIRECTOR
BROWNSBURG COMMUNITY SCHOOL CORPORATION

BROWNSBURG WEST MIDDLE SCHOOL 2004- 2005
- **Scheduling:** Organized and scheduled the activities of 31 teams in 12 different sports
- **Budget:** Solely responsible for the control and review of the athletic budget and fiscal accountability
- **Philosophy:** Crafted an athletic department philosophy and communicated it effectively to staff, coaches, and parents

TEACHER EIGHTH GRADE AMERICAN HISTORY
BROWNSBURG COMMUNITY SCHOOL CORPORATION
BROWNSBURG WEST MIDDLE SCHOOL 2003- 2004
- Developed and implemented innovative lessons adapted to the needs of diverse learners

ATHLETIC DIRECTOR
HAMILTON TOWNSHIP LOCAL SCHOOLS
HAMILTON MIDDLE SCHOOL 2001- 2002
- Responsible for supervision and management of athletic functions

TEACHER EIGHTH GRADE GOVERNMENT
HAMILTON TOWNSHIP LOCAL SCHOOLS
HAMILTON MIDDLE SCHOOL 1998- 2003
- Responsible for government curriculum and preparing students for the ninth grade proficiency test

PRESENTATIONS

NATIONAL SCHOOLS TO WATCH CONFERENCE: Presented the Eastwood System of Professional Learning Communities

UNIVERSITY OF INDIANAPOLIS: Panel presentation to Masters in Education students

TAYLOR UNIVERSITY: Panel presentation at the Higher Education conference

BUTLER UNIVERSITY: Presentation to Masters in School Counseling students

ADDITIONAL EDUCATION

DEVELOPMENTAL DESIGN TRAINING

SUMMER 2017

INTERNSHIP: Worked with Dr. Jon Milleman in developing a tool for the improvement of teacher primary and secondary evaluation measures   FALL 2013

IB TRAINING: Trained in the IB application process, design cycle, and IB assessment criteria   SUMMER 2011 AND SUMMER 2013

AVID SUMMER INSTITUTE: Developed the AVID site team plan   SUMMER 2011

IB CONFERENCE OF THE AMERICAS: Trained to serve as IB head of school   SUMMER 2010

NATIONAL CONFERENCE ON DIFFERENTIATED INSTRUCTION: Trained in middle school instruction by Rick Wormeli   SUMMER 2008

ADDITIONAL ACTIVITIES

PRESIDENT   2017–2018
INDIANA CURRICULUM AND INSTRUCTION ASSOCIATION

CO-PRESIDENT   2013–2017
DISTRICT 7 OF THE INDIANA ASSOCIATION OF SCHOOL PRINCIPALS

PRESIDENT   2013–2014
MARION COUNTY PRINCIPALS ASSOCIATION

EVALUATOR
NORTH CENTRAL ACCREDITATION

AWARDS

PRINCIPAL OF THE YEAR   2019
INDIANA ASSOCIATION OF SCHOOL PRINCIPALS

ADMINISTRATOR OF THE YEAR   2016
WASHINGTON TOWNSHIP SCHOOLS

PAULA SILVER CASE AWARD   2014
UNIVERSITY COUNCIL FOR EDUCATIONAL ADMINISTRATION