EFFECTS OF COLLABORATIVE TEACHING EVALUATION PROCESS ON

TEACHER/ADMINISTRATOR COLLEGIALITY

By

Gary Alan Storts Jr.

A Thesis Presented to

The Faculty of Humboldt State University

In Partial Fulfillment of the Requirements for the Degree

Master of Arts in Education

Committee Membership

Dr. Eric Van Duzer, Committee Chair

Dr. Kenny Richards, Committee Member

Dr. Eric Van Duzer, Graduate Coordinator

December 2018

ABSTRACT

EFFECTS OF COLLABORATIVE TEACHING EVALUATION PROCESS ON
TEACHER/ADMINISTRATOR COLLEGIALITY

Gary Storts

The job of the professional educator is becoming ever more complex. Research has shown that effective teacher evaluation processes and collegial relationships between the teacher and the principal both positively impact student achievement. This study investigates how a collaborative teacher evaluation process effects collegiality between the teacher and the site administrator. Four teachers who had experienced both traditional and collaborative reviews in a small rural school district in Northern California were interviewed to assess the impact that a collaborative evaluation had on their professional relationship with their principal.

The results of this research confirmed that teaching evaluations, however they are structured, produce varying levels of anxiety. However, the anxiety was short lived and temporary, with three of the four teachers finding value in the collaborative evaluation process once completed. The results identified technology related obstacles to the collaborative approach that generated frustrations. Two participants reported that they preferred components of the more traditional teaching evaluation.

TABLE OF CONTENTS

INTRODUCTION

Perhaps more than at any other time in the history of education in the United States of America, educators are beings asked to embrace collaboration – collaboration is taught and emphasized as a preferred evidenced-based instructional strategy, and collaboration with other teachers is planned and structured in professional learning communities. Research has also demonstrated the positive impacts that a healthy teacher-principal relationship can have on student achievement. We also have an abundance of research that points to the importance of developing teaching evaluation systems that effectively measure and develop our professional educators. The evidence is clear that these two factors are key ingredients for a successful school site. As such, the next reasonable question to ask is why collaborative components are not part of the teaching evaluations process. Collaborative teacher evaluations involve an evaluator observing a lesson, then discussing that lesson's strengths and weaknesses with the classroom teacher, before jointly working together to improve classroom instruction. Given the evidence, when a teacher and a site administrator are able to collaborate by sitting down to have a professional conversation reflecting on a recently taught lesson, the experience should support higher levels of professional collegiality.

This study will be used to determine whether or not a collaborative teaching evaluation has a measurable impact on teacher-administrator collegiality, as observed through teacher interviews. For the purposes of this study, collegiality is defined as "sharing responsibility in a group endeavor."

   This thesis will begin with a review of the extant research, before explaining the methods used in selecting and conducting the interviews to examine whether or not collaboration embedded into a teacher evaluation effects the level of collegiality between the teacher and the administrator. A summary of results will follow, before a discussion about the evidence and limitations of the study are presented.

LITERATURE REVIEW

"Alone we can do so little; together we can do so much."

- Helen Keller

Introduction

To fully analyze the benefits of utilizing a collaborative approach in the teaching evaluation process, it is first helpful to review the history and philosophies behind past teacher evaluation reform efforts, as well as examine current practices in the United States. It is for these reasons that this literature review begins with a historical overview of the teaching evaluation process, before examining research findings that detail limitations to the traditional approach of teacher evaluation. Next, this literature review will focus on successful research-based and research-supported components to the evaluation process. After exploring what works and why in regards to effective teacher evaluation, this review will examine the importance of collegiality between teachers and administrators, as well as share findings on common characteristics of successful schools, and the benefits of collegiality to teachers, students and stakeholders. Finally, this review will conclude with recommendations for further research, and a proposal for this study.

Purpose of Teacher Evaluations

Efforts to improve student achievement have led states, districts, and schools across the country to develop, or implement, new teacher evaluation systems (U.S. Department of Education, 2009). Considerable amounts time and energy have been spent to bring improvements to the teacher evaluation process. Two significant failings of past teacher evaluation efforts include: (1) Teacher evaluation systems have not accurately measured teacher quality because they've failed to do a good job of discriminating between effective and ineffective teachers, and (2) teacher evaluation systems have not aided in developing a highly skilled teacher workforce (Bill & Melinda Gates Foundation, 2011; Toch & Rothman, 2008; U.S. Department of Education, 2009; Weisberg, Sexton, Mulhern, & Keeling, 2009).

A teacher evaluation process that includes both "measurement" and "development" goals is crucial to a successful teacher evaluation process (Marzano, 2012). In a 2012 study, Marzano surveyed 3,000 educators on teacher evaluations and whether teacher measurement or teacher development was the end-goal of the evaluative process. Seventy-six percent of the respondents believed that both teacher measurement and development should be by-products of the teacher evaluation, but more emphasis should be placed on development. To understand the objective of teacher measurement and development, a historical review of teacher evaluations will identify past successes and shortcomings.

Evolution of Teacher Evaluations

In the beginning of the 18th century, education was not yet recognized as a professional discipline as local governments and clergy were used to hire and evaluate teachers (Tracy, 1995). By the mid-1800s, as industrialization spread across the United States, the need emerged for more comprehensive school systems and specialized instruction in specific disciplines (Marzano, Frontier, & Livingston, 2011). Larger student populations also led the way for "principal" teachers who could assume increasingly complex leadership roles, which included teacher evaluations. Supervisors were required to have subject area knowledge as well as an understanding of teaching skills, (Bolin & Panaritis, 1992). Pedagogical skills became viewed as a necessary component of effective teaching (Tracy, 1995).

The last half of the 19th century and early part of the 20th century saw a conflict of opposing educational views (Marzano, Frontier, & Livingston, 2011). John Dewey, one of the most prolific educational writers of his era, saw democracy as the central link to successful human progress. Dewey further promoted the notion of schools fostering opportunities to practice citizenship as a way to further nurture democratic ideals (Dewey, 1938, 1981).

In contrast to Dewey's position of democracy as the hallmark of a successful schooling experience, Fredrick Taylor viewed scientific management, and the measurement of specific behaviors, as a powerful means of improving schools (Marzano, Frontier, & Livingston, 2011). Taylor argued that if there were 100 ways to shovel coal, one best method could be determined and applied for greater efficiency and increased results (Taylor, 1911). Taking Taylor's ideas into account, both Cubberly (1929) and

Wertzel (1929) began proposing the use of student learning data to determine the effectiveness of a teacher or a school. This approach gained educational acceptance through World War II.

Immediately following World War II, the pendulum swung the opposite direction as education began to move away from scientific management (Marzano, Frontier, & Livingston, 2011). Teacher evaluation literature of this time period began focusing on the teacher, with an emphasis placed on developing teacher skills (Marzano, Frontier, & Livingston, 2011). The role of the evaluator shifted during this time with school management taking on a greater number of roles and responsibilities in teacher evaluation (Coleman, 1945). Whitehead (1952) noted the importance of six areas of teacher supervision including classroom visitation, demonstration teaching, faculty meetings, orienting new teachers, lesson planning and in-service training. Whitehead also surveyed teachers about their perceptions of teacher supervision. A consensus amongst the surveyed teachers revealed the need for follow-up conferences after classroom observations. The recognition that effective classroom observation was vital to student success set the tone for an influential shift in supervisory practices across the United States (Whitehead, 1952).

The importance of ensuring effective teaching as one of administrators' main priorities was solidified by the late 1960s and quickly became known *as clinical supervision* (Goldhammer, 1969). By 1980, one study found that 90% of school administrators were using some form of *clinical supervision* (Bruce & Hoehn, 1980). Today, many teaching evaluation models still resemble the overall structure of the

clinical evaluation cycles, including the following phases: pre-observation conferences, classroom observations, analysis, supervision conference, and an analysis of the evaluative process (Marzano, Frontier, & Livingston, 2011). The aim of this evaluative model was to use the five phases of the clinical supervision process to diagnose successful classroom practices. As well intentioned as the clinical supervisory model was, over time, the five phases were reduced to a series of mandatory tasks. The thoughtful, collegial, data-driven approach which was once thought to greatly improve student success faded from its desired purpose (Marzano, Frontier & Livingston, 2011).

While a subscribed five-step process for a teacher evaluation has lingered since the 1980s, the Hunter model, a seven-step model for classroom lesson development, provided the next major influence in teacher supervision and evaluation, referred to as mastery teaching (Hunter, 1984). The pre-conference, observation, and post-conference all centered on the concept of mastery teaching. Evaluators used script tapping, or observational shorthand, to find elements of mastery teaching embedded in each lesson. After the observation, the lesson's various details would be categorized and a rating would be assigned based on the tenants of mastery teaching following a conference with the teacher (Fehr, 2001). Mastery teaching was widely used, but was often described as being a prescriptive supervisory practice, and as a result, fell out of favor during the early-1990s (Marzano, Frontier, & Livingston, 2011).

While the Hunter model focused primarily on teaching pedagogy and the components of a successful lesson, the next wave of supervisory reform placed an increased emphasis on considering, then developing, a teacher's career goals (Marzano,

Frontier, & Livingston, 2011). This differentiated approach to teacher evaluation placed importance on the notion that teachers have some input and control over their professional development (Glatthorn, 1984). Differentiated opportunities would be made available to teachers based on specific and individual needs. Glickman (1998) was another champion of the differentiated approach to teacher supervision and believed that teacher development should be the main aim of the evaluation process. As one could imagine, the era of developmental teacher evaluation reform was in sharp contrast, and drew substantial criticism, from the proponents of clinical supervision and mastery teaching. The pendulum was set to swing again and the next school supervisory reform era would more closely inspect the intricacy of the teaching profession, and place a greater emphasis on teacher evaluation (Marzano, Frontier, & Livingston, 2011).

The Danielson model, first introduced in the mid-1990s, sought to capture the full complexity of the classroom teaching process (Danielson, 2007). Danielson's model included four domains: planning and preparation, the classroom environment, instruction, and professional responsibilities. The goal of the Danielson model was to accomplish three tasks; first, honor the complexity of teaching, second, provide the language for professional conversations, and third, provide structure for teacher self-assessment and reflection. One of the powerful contributions of the Danielson model was the alignment of 76 standards of quality teaching, and breaking those individual components into four performance levels, e.g., unsatisfactory, basic, proficient and distinguished. The specificity of the teaching standards and their union with the detailed performance levels makes that the Danielson model the most detailed and comprehensive supervisory model

yet developed, and serves as the initial reference point for current teacher evaluation reform efforts (Marzano, Frontier, & Livingston, 2011).

Twenty-first century efforts to improve the teacher evaluation process swung yet again, from supervision to evaluation, and teacher behavior to student achievement (Tucker & Stronge, 2005). Tucker and Stronge (2005) insisted that both instructional practices and evidence of student learning should be valued concurrently, concluding that there is a direct correlation between teacher effectiveness and student learning.

The Widget Effect (Weisberg, Sexton, Mulhern, & Keeling, 2009) notes the failure of teacher evaluation systems in the United States to produce accurate information regarding individual teachers' instructional practices, citing 73% of surveyed teachers stating that their most recent evaluation did not identify any teaching areas that needed improvement and those teachers who did have identified areas of needed improvement, only 45% found the suggestions of their supervisors helpful (Weisberg, Sexton, Mulhern, & Keeling, 2009). *The Widget Effect* research suggest a complete overhaul of the teacher evaluation system, but before embarking on a comprehensive revision of teacher evaluation systems, we must first identify the limitations of the current teacher evaluation processes.

<div align="center">Limitations of Current Teacher Evaluation Practices</div>

Why is it so difficult to improve the teacher evaluation process? Kraft and Gilmour (2016) asked this questions after revisiting *The Widget Effect (2009)*. First the researchers, reviewed some of the data *The Widget Effect* (2009) produced, including, the

fact that less than 1% of teachers were rated in the lowest performance category, despite 83% of administrators and 57% of teachers stating they could identify teachers on campus who are ineffective. Kraft and Gilmour (2016) investigated whether this perception of teacher ineffectiveness has changed in the years since *The Widget Effect* was published.   In reviewing data from 19 states and an intensive analysis of one urban district, the researchers found that less than 3% of teachers were rated below proficient/exemplary on a four or five point performance scale (Kraft & Gilmour, 2016). The researchers also surveyed principals, who on average estimated that 28% of teachers performed below proficient, a sharp contrast between the professional evaluation ratings and perceived teacher ratings. A natural question to ask is, why do so few teachers receive below proficient ratings, despite the fact that school supervisors estimate that greater than one-quarter of their teaching staffs are not performing adequately inside their classroom?

Marshall (2017) believes one flaw of current supervisory practice is simply the time-consuming nature of annual performance evaluations – approximately four hours per evaluation. Marshall notes that the frequency of evaluations, generally only 1-2 times per year, leaves hardly enough time to have meaningful conversations with teachers to impact performance.   While interviewing principals from across the country, Kraft and Gilmour (2016) discovered that the daunting workload involved in assigning low performance ratings is a likely reason for the discrepancy of actual ratings given to teachers and the perception of daily effectiveness. The researchers cited the time consuming practices of gathering the needed observational evidence required to justify a

below adequate performance rating, the mandatory corrective action plans, and the extensive support needed to assist underperforming teachers, coupled with the likely legal dispute involved with dismissing an ineffective teacher as barriers to an effective evaluative process (Kraft & Gilmour, 2016). Marshall (2017) also noted that most post observation and teacher evaluation conferences finalizing ratings occur in April or May, when teachers are more than 75% of the way through the current school year and about to embark on a two-and-a-half month summer vacation. Complicating the poor timing is the fact that many teachers view the feedback and developmental suggestions given to them by their principal as not beneficial to their teaching practice.

Additional challenges include the common practice of advanced warning of an upcoming observation (Marshall, 2017). If the teacher being evaluated has received advanced notice regarding the precise date and time of the evaluation, the principal is likely to see the optimal lesson, which may not be the typical lesson, allowing a potentially ineffective or marginal educator an opportunity to go undetected and not have their deficiencies addressed (Marshall, 2017).

Evaluators themselves often face personal challenges when evaluating teachers (Kraft & Gilmore, 2016). Some principals interviewed were hesitant to give beginning teachers a below average rating out of a desire to not discourage those new to the profession. A few principals interviewed by the researchers mentioned that they did give constructive feedback to beginning teachers, just outside of the professional evaluation process (Kraft & Gilmour, 2016). Principals who gave new teachers an evaluative break rationalized their decision by believing that assigning a low rating would come at the loss

of relational trust, and that trust was essential for mobilizing the collective efforts of the school.

Kraft and Gilmore (2016) also found evaluator discomfort as an obstacle in assigning a teacher a subpar evaluative rating. The researchers interviewed a principal who noted that the most difficult part of their job was communicating a poor performance rating and that, "not everyone is capable of that" (Kraft & Gilmour, 2016, p.15). The human aspect of knowing that a probationary teacher could lose their job as a result of a low evaluative rating weighed heavily on many of the principals interviewed. "The last thing I think I wanna do as a human being is to watch another human being walk out with their head down, dejected, because they lost their job…" (Kraft & Gilmour, 2016, p.16).

In a Fall 2017 commentary in the *AASA Journal of Scholarship and Practice*, Geirger and Amrein-Beardsley (2017) referenced additional factors evaluators confront that make the traditional approach to teacher evaluations challenging. Given these challenges, Geirger and Amrein-Beardsley (2017) discuss three specific types of observational data manipulation that occur as a result of evaluators feeling pressure to correlate their observational rating with a numerical measurement of student growth over time. In the education profession, these examples of data manipulation are commonly referred to as Campbell's Law (1976). "Campbell's Law states that, in essence, the higher the consequences or stakes surrounding almost any quantifiable event (e.g., one that is based on numerical scores or outcomes), the more likely the scores or outcomes are subject to pressures of corruption and distortion, as directly related to the relative importance or weight of the consequences attached" (p. 46). Mainly as a result of high-

stakes testing, Campbell's Law has been prominent in education related discussions for decades, and the last two federal education policies, No Child Left Behind (2001) and Race to the Top (2011), place an increased emphasis on student achievement. Race to the Top (2011) incentivized states with 4.35 billion dollars in federal funds to adopt a new component to teacher evaluations that numerically measures student academic achievement over time using value-added models (VAMS). As a result, school administrators felt the pressure to manipulate observational data by (1) artificially inflating observational ratings to protect against potential termination or loss of tenure, (2) artificially deflating observational ratings - which occurs much less frequently than inflating observational ratings but, when it does occur is often the result of evaluators trying to ensure that the ratings, when looked at as a whole, fit a normal bell curve, and (3) artificially conflating observational ratings to directly correlate with student growth, or VAM scores (Geirger & Amrein-Beardsley, 2017). Greiger and Amrein-Beadsley (2017) caution school administrators and evaluators that artificially manipulating observational ratings is poor practice and diminishes the combined validity of the observation and VAM, as well as any inferences that might be drawn to persuade future federal education policy.

There is general consensus that the relationship between teacher contribution and student growth and achievement should be part of the teacher evaluation, however a growing body of evidence suggests that VAMs are not consistent from year to year and may not be the most accurate indicator of teacher effectiveness despite its growing popularity (Darling-Hammond, Amrein-Beadsley, Haertel, and Rothstein (2012).

Darling-Hammond, Amrein-Beadsley, Haertel, and Rothstein (2012) referenced a study examining VAM data from five school districts. The study found that of those teachers who scored in the bottom 20% of rankings the first year of the study, approximately one-quarter remained in a similar ranking the next year. Interestingly, approximately one-third of those same teachers moved into the one of the highest categories the following year. Similar inconsistencies were observed for teachers who scored in the top rankings. The researchers were also wary of hidden risk factors associated with VAM, such as teachers teaching to the test if student test scores are tied to evaluating teacher effectiveness. As one Houston Independent School District representative was quoted "I do what I do every year. I teach the way I teach every year. My first year got me pats on the back; my second year got me kicked in the backside. And for year three, my scores were off the charts. I got a huge bonus, and know I am in the top quartile of English teachers. What did I do differently? I have no clue" (Amrein-Beardsley & Collins, 2012, p. 15).

Marshall (2017), former 15 year Boston, MA elementary principal and founder of the Marshall Memo, believes that teacher evaluation, if done right, can be a powerful device for improving teaching and student learning. Marshall also recommends that school boards and superintendents come face-to-face with four hard truths: First, students learn much more from some teachers than from other teachers, which is a result of day-in, day-out specific practices that successful teachers implement. Second, each school building has a range of teachers from dynamic to less than effective. Third, students with barriers (special needs, family troubles, bullying, etc.) have a more immediate need for

dynamic teaching then their more fortunate peers. And fourth, the traditional teacher evaluation process makes it difficult for the evaluator to implement change for teachers who require immediate assistance.

There are numerous limitations to the traditional teacher evaluation system, and the frustrating aspect for all stakeholders is the time and energy spent on a generally inaccurate and ineffective process. The next question is simply, what works and why?

Teacher Evaluations: What Currently Works and Why?

Previously in this literature review I noted that the goal of the teacher evaluation process is for both measurement and development of teacher effectiveness (Marzano, 2012) in order to provide an optimal learning environment for our students. Fortunately, there is sufficient research to inform us on successful current teacher evaluation practices (Danielson, 2010).

Teacher self-assessment, in concert with mutually agree upon teaching standards and teaching frameworks, is becoming a common occurrence in many state and district teacher evaluation systems (Danielson, 2010). The Colorado Department of Education (2017) encourages educators to be thoughtful and reflective about their day-to-day teaching practice and believes that self-assessment is a staple component of teacher evaluation best practice, citing evidence that references adults' engagement in self-reflection as a factor for teachers to more likely retain learning compared with learning that is enacted upon them. Additionally, the Colorado Department of Education notes a significant opportunity afforded to teachers who then engage in goal setting with their

evaluator, mentioning an important opportunity to discuss performance planning, as well as build trust and discuss instructional supports that are available to assist in successful classroom instructions. Researchers Ross and Bruce (2007) support the practice of self-assessment as a valuable tool and exercise for enhancing teacher effectiveness so long as self-assessment is partnered with additional professional development efforts.

In order to consistently and reliably measure teacher effectiveness, professional teaching standards and teaching frameworks have been developed and widely implemented in teacher evaluations across the country (Danielson, 2010). Danielson (2011) describes the need for a consistent definition of good teaching, "It is not sufficient to say, 'I can't define good teaching, but I know it when I see it'" (p. 35). California's professional standards and frameworks provide a common description of successful classroom teaching practices, as well as other professional teaching responsibilities essential to student learning and teacher growth, such as creating and maintaining effective environments for student learning, planning instruction and designing learning experiences for all students, and developing as a professional educator (California Standards for the Teaching Profession, 2009). Once all members who comprise the evaluation team – teachers, administrators, instructional coaches, etc. – share a common understanding of what is expected of a professional teacher, the conversations that occur as a result of the evaluation become much more meaningful (Danielson, 2011).

At local levels, once professional standards have been established, they can then be translated into teacher evaluation instruments, (Darling-Hammond, Amrein-Beadsley, Haertel, & Rothstein (2012). This approach has proven successful in several Cincinnati

public schools by producing increased teacher effectiveness ratings with an emphasis on student learning.   The ratings were also used to identify areas for future teacher growth, (Milanowski, 2004).

Researchers Reinhorn, Johnson, and Simon (2017) studied Massachusetts schools that received their states top accountability rating after implementing a uniformed Massachusetts teacher evaluation process. During interviews with teachers at these distinguished schools, it was noted that administrators observed each classroom at least twice a month and each teacher received detailed standard aligned feedback on only one or two standards per visit, and usually within 24 hours of the classroom observation. Marshall (2017) supports the principle that multiple unannounced visits into teachers' classrooms paint a more accurate picture of day-to-day classroom life compared to the traditional 1-2 time per year observation model with supervisors being allowed to focus on one teacher's need at a time, making difficult conversations less uncomfortable and reducing teacher defensiveness to feedback.

In addition to multiple observations and unannounced visits on behalf of the evaluator, teachers self-rating their own performance using a classroom video recording has been shown effective in recent research (Marzano, Frontier & Livingston, 2011). Although not widely used in the teacher evaluation process, Rosaen, Lundeberg, Cooper, Fritzen, and Terpstra (2008) discovered that when beginning teachers viewed themselves on camera, they paid more attention to their instructional practices, and more attention to their students, especially as it relates listening to the students' questions and answering teacher provided prompts. To further illustrate the value of video Calandra, Brantly-

Diaz, Lee, and Fox (2009) divided beginning teachers into two groups for guided reflection after an observation. The first group immediately engaged in a post-observation conference with their evaluator and were later asked to write a reflection about their lesson. The second group did not participate in a post-observation conference and instead was asked to capture the lesson on digital video before completing the same writing reflection exercise as group one. The researchers discovered that group two "tended to write longer and more pedagogically more connected responsive pieces than their non-video recorded (NVR) counterparts…" (Calandra, Brantly-Diaz, Lee, & Fox, 2009, p. 81) "We also found that the video recorded (VR) group described transformations in their thinking about teaching, which was less evident in the more technical NVR group writing" (p. 81). The traditional approach to evaluation, and the accompanying post-observation conference, had the teacher sitting and listening and not actively engaged in a reflective process (Calandra, Brantly-Diaz, Lee, & Fox, 2009). In support of video observations, Sewall (2009) found video-based observations to have many positive benefits, but perhaps most notably was that the teacher actively delivered most of the commentary on their lesson and enthusiastically engaged in deep levels analysis and self-reflection during the evaluative post-conference.

An aspect of the teacher evaluation process that must not be overlooked is the value and merit of collaboration. For the six distinguished Massachusetts schools recognized for their top accountability rating, researchers Reinhorn, Johnson, and Simon (2017) noted that the collaborative component embedded in the formal teacher evaluation system was part of a broader strategy of campus-wide improvement that included peer

observation, instructional coaching, teacher team collaboration, and whole staff professional development. The researchers noted that while the evaluation process is primarily an individual experience, many teachers at these schools would take their formal rating to their peers to gather further feedback. Collaboration was a part of their campus culture (Reinhorn, Johnson, and Simon (2017). Danielson (2016) believes that collaboration can be embedded into a teacher evaluation system in two distinct ways: (1) novice teachers could be collaborating with more experienced mentor teachers during initial evaluation years, and (2) experienced teachers in good standing should be eligible to apply for teacher leadership positions; these positions would come with training and support, as well as extra compensation or release time during the regular school day.

Similar to Danielson's notion of novice teachers receiving mentorship from accomplished teachers in good standing, the National Commission on Teaching and America's Future (NCTAF) (1996) suggests the use of Peer Assistance and Review (PAR) programs, where highly effective veteran teachers provide guidance to beginning as well as struggling veteran teachers. Key collaborative features of the PAR program include mentor teachers consulting with their peers needing assistance in identified evaluative domains, as well as the system of due process where a panel of teachers and administrators make recommendations about personnel decisions based on evidence from formal teacher evaluations.   NCTAF research (1996) found that beginning teachers who receive strong support are more likely to have classes that achieve greater results in reading than their less fortunate beginning teaching peers.

Along with support from their more experienced teaching colleagues, there is the growing belief in some states and school districts that student perception surveys can provide a valid and reliable way to measure a teachers overall classroom performance (Bill and Melinda Gates Foundation, 2011). In the state of Georgia, 75% of teachers who participated in their state's pilot student feedback project found student responses helpful in adopting better classroom practices, (Bufford, 2015). According to Schulz, Sud, and Crowe (2014) "The jury is still out, however, on whether student surveys will join classroom observations and student achievement data as a third common measure in newly redesigned teacher evaluation systems, or if adoption will remain limited to a small number of progressive districts and CMOs (charter management organizations)" (p.16). These same researchers believe that student perception surveys can be a useful addition to teaching evaluations, concluding that student survey data can provide feedback to teachers and administrators which can be used to target specific areas for improvement, professional development, and recognition.

The current trends in the ever swinging pendulum of teacher evaluation systems clearly indicate a desire for greater collaboration, which leads to the question: whether or not a collaborative teacher evaluation process leads to improved teacher and administrator collegiality.

Collegiality: Why It Matters?

Collegial: adj. 1. Collegiate. 2. (of colleagues) sharing responsibility in a group endeavor (Webster's University College Dictionary, 2001).

Ensuring student success takes a unified and collegial effort beyond just the classroom walls (Success for All Foundation, 2012). Success for All (SFA) is a validated research-based whole-school reform strategy that centers on shared campus-wide goals, targets, and implementation foci crafted in a collegial manner. SFA's leadership for the continuous improvement model is a "collaborative leadership system that brings school staff together to focus everyone's efforts on success for every child." (p.7).   Instructional teams, as well as school-wide solutions teams develop quarterly plans, meet on a regular basis to implement the plan, and report on progress. Together, this collegial approach creates a campus climate that provides top-notch instruction and student supports that ensure student success (SFA, 2012).

Shah (2012) also points to the strong correlation between collegial relationships and school improvements and success. Shah highlights a study of less advantaged public schools in Chicago where campuses that were identified as having a strong professional learning communities improved at a rate four times faster than schools without these collegial systems in place. Further acknowledging the importance of meaningful professional staff relationships, Noddings (2014) noted, "A school is not just a center for the production of learning. It is a place which people become attached" (p.18). "It is a place where educators break down boundaries to work collaboratively, planning and teaching with creativity, and with the steady purpose of producing better adults" (p.18).

"The level of collaboration demanded by modern teaching is unprecedented," (Barnwell, 2015). The key to unlocking scholastic success is more than just reading, writing, and arithmetic; professional collegial relationships matter (SFA, 2012). Bell

(2001) noted that a sense of family was a common characteristic among the High-Performing, High-Poverty (HP2) schools in California. "Collegiality, collaboration, inclusion and a sense of community were an integral part of how the schools conducted business. Staff was trusted with responsibility to accomplish the school's academic and nonacademic goals" (p.10). HP2 schools also demonstrated shared leadership among administrators, faculty and parents, as well as collaboration on school goals (Bell, 2001). Collegiality matters at all levels and once trust and transparency are campus climate norms, schools have a foundation upon which goals, monitoring of student success, and professional conversations about evidence are practiced regularly (Thiers, 2017).

Hiebert and Stigler (2017) researched the drastically different instructional practices found in other countries, particularly Japan, with the traditional instructional model of teaching in the United States. They discovered that the biggest problem behind the performance of the US in international comparisons was that the U.S. had no consistent approach for improving instruction. Hiebert and Stigler (2017) conclude that Americans have focused too much on improving teachers by attempting to recruit superior new teachers, increasing the rigor of the teaching standards, increasing accountability, professional development, value-added measures, and making it easier to dismiss unsuccessful teachers. Meanwhile, in Japan, educators have placed their attention on continuously improving teaching through the collegial campus efforts of lesson study. As a result, instructional practices in Japan have improved dramatically, while U.S. teaching efforts have remained the same for the last half-century (Hiebert & Stigler, 2017).

The desire of a highly collaborative, collegial workplace is not isolated to the educational realm. Studies have identified three common factors in hospitals that ensure better patient health care for a reduced cost (Anrig, 2013). These factors include: (1) a highly collaborative culture built on teamwork, (2) an usually sophisticated attentiveness to test data to monitor patient progress and respond to problems, and (3) an orientation toward ongoing adaptation rather than rigid adherence to established routines (Anrig, 2013).

A 2010 study conducted by the University of Chicago of 400 beat-the-odds Chicago elementary schools revealed similar factors leading to in their successes. Those factors included: (1) clear curriculum choices, materials, and assessments aligned across all grade-levels with the input of teachers, (2) peer assistance that included an open-door policy for fellow colleagues and outside consultants, (3) strong parent and community ties with integrated interventions for students, and (4) leadership that distributes responsibility for student achievement (Anrig, 2013). Anrig (2013) references a final study of effective schools that share closely related characteristics of Chicago's high performing schools, including: (1) teacher/administrator collaboration time for developing and selecting instructional materials, (2) setting aside time each week for teacher collaboration to focus on instruction, (3) teachers being open to coaching and observation, (4) teachers and administrators closely monitoring assessment data to identify areas of student struggles.

Collegial collaboration directly benefits teachers (Shah, 2012) and is considered a critical component to teacher professional development. "Collegial communities create

such a cooperative climate that heightens the level of innovation and enthusiasm among teachers and provides a continuous support for staff and professional development" (Shah, 2012, p.1243). An effective use of in-year assessments would be allowing teaching teams to review data that could be used to enhance instructional practices, while focusing on a collegial practice as the main emphasis of professional development (Hiebert & Stigler, 2017). A collegial professional development emphasis is also a tremendous benefit to novice or struggling teachers who would be able to utilize a Peer Assistance and Review (PAR) type program, where more successful and more experienced teachers serve as mentors and coaches for their struggling colleagues (Anrig, 2013).

Perhaps the most fundamental reason for pursuing a collegial campus environment is the direct benefit to students. Increased levels of collaboration among teachers correlates with improved student achievement (Schmoker, 1999). Effective collaborative teaching teams are able to increase the rigor and expectations for their students, allowing for more critical thinking opportunities (Shafer, 2016). A study of high-performing Tennessee elementary schools discovered a shared best practice of reserving time for horizontal collaboration (Shah, 2012). Quintero (2017) agrees student performance increases dramatically when teacher have frequent and instructionally focused conversations with their colleagues.

Conclusion

"The teacher principal relationship in schools is an important component in achieving the goals of a school and more importantly to increasing students' overall achievement level" (McFarlane, 2012). Given the wealth of knowledge and data that supports the importance of teacher evaluations, as well as the success of collegiality on the school campus, it seems as though we may be doing our students, teachers, and school stakeholders a disservice by not marrying these two vital components of scholastic success. If collegiality is key to success at school site, then why is it that given all of the teaching evaluation shifts throughout U.S. history, that teacher and administrator collaboration has not become a more consistent component to the teacher evaluation process? Charlotte Danielson may have summarized traditional evaluative practices over the years best: it's done to them, not with them (Marshall, 2015).

Can we continue to implement the collaborative nature of pre-and-post observation conversations, keeping these positive components in place, but infuse other aspects of the evaluative process with greater collaboration? Can teachers self-assess using common professional standards, then have both the teacher and the administrator agree on one or two specific goals related to those standards? Can video be utilized to record the observation, allowing the teacher and administrator an opportunity to view the lesson together – engaging in a collegial lesson study conversation? Can the teacher and administrator collaboratively measure the success of the lesson using a rubric based on professional teaching standards? And will these efforts lead to improved collegiality between the school leader and the teaching staff? Goddard, Goddard, Kim and Miller (2015) note, "Leaders have the tremendous potential to influence the collective work and

beliefs of teachers in their schools" (p.527). Connecting school leadership, teacher practice, and school district belief systems is critical for student achievement (Goddard, Goddard, Kim, & Miller, 2015). Although leadership, teacher evaluations, and collaboration have individually been broadly studied, there appears to be an absence of research combining these key school components.

METHODS

The overall goal of formal teaching evaluation process is to measure and develop effective practices. However, the overwhelming majority of research on this topic concludes that traditional teacher evaluations do neither. This study will use a locally developed semi-structured interview, consisting of 8 questions, (appendix A) designed in an effort to measure the collaborative evaluations effectiveness/impact on teacher and administrative collegiality, as well as gauge the degree to which participants' value the collaborative evaluation process. Four tenured teachers at a rural elementary school district who have, in recent years, participated in a traditional teacher evaluation, and who, this school year, participated in the newly crafted collaborative evaluation, chose to participate in the study.

During the last week of March, 2018, eight teachers were invited to participate by using a simple random sample of tenured teachers that have experienced both traditional and collaborative reviews from a list maintained by the district. Four teachers agreed to participate, (2 male and 2 female) all of whom are tenured elementary educators (13-29 years of teaching experience). Participants were invited by email to participate in a semi-structured interview that took approximately a half-hour. The interviews were scheduled to take place during the month of April, 2018 in a quiet, confidential location of the participant's choosing and required written consent prior to the interview. Both prepared prompts and ones that emerge as part of the conversation were used to fully investigate

perceptions, attitudes and opinions of the teachers. Interviews were recorded, transcribed and analyzed using a qualitative coding process.

Confidentiality was afforded to all interview participants by providing the following assurances; (1) data was summarized and aggregated so that no individual can be identified, (2) direct quotes were not be used unless permission was granted by the participant, (3) participants were given the opportunity to select a pseudonym, (4) and the code sheet used to link pseudonyms to participants was hand written and kept in a locked filing cabinet separate from interview data.

An on-site School Counselor was available to assist any participant experiencing emotional stress that may arise from participation in the study.

The consent form included a statement: "The goal of this research is to determine what works. No information shared during the interview will be used in any supervisory or evaluative manner and will be held confidential."

RESULTS

Introduction

This chapter will draw upon the main themes and present the findings which arose out of

the interview process and subsequent data analysis. First, a brief profile of each of the

participants will be presented. The key themes that emerged following the data analysis

on the collaborative teaching evaluation and the resulting impact on teacher and

administrative collegiality were; (1) Psychological Impact of Being Evaluated, (2)

Adding Technology Created Both Positives and Negatives, and (3) Consistent Desire for

Increased Teacher-Principal Presence. All themes are interconnected.

Participants

- Participant one is a male general education teacher with 23 years of experience,
  who has been employed at 4 schools during the course of his career, and is
  currently teaching in his third year at his current placement.

- Participant two is a female general education teacher with 13 years of experience,
  who has been employed at 3 schools during her career, and is currently teaching
  in her tenth year at her current placement.

- Participant three is a male single subject teacher with 29 years of experience, who has been employed with the same school district, serving two elementary school campuses during his career, and is currently in his twenty-ninth year at his current placement. Participant three is retiring at the end of the school year.

- Participant four is a female multiple subject teacher with 14 years of experience, who has been employed at 6 schools in her career, and is currently in her ninth year at her current placement. Participant four also serves as the Chair Bargaining Representative for the school district's teacher's union.

Psychological Impact of Being Evaluated

All of the participants felt that being evaluated by their site administrator created heightened levels of anxiety and stress. Participant two described her pre-evaluative experience as feeling, "Super-nervous and anxious – evening crying a little bit." Another participant added:

> *The evaluation is my site administrator's judgement. Your evaluation is like your grade as a teacher. It's like you're an A, you're a B, you're a failure. What we do is very personal and it takes so much from you – from your gut and from your viscera. It's such an identity thing to be a teacher. It's not just a job that you get to just get up and go do. When you are going to be judged on something that is so personal and essential to your identity, it's really hard. For someone to look at everything you are and say, 'that's bad', there is fear involved."*

Another participant spoke at length about the perceptions of how teacher evaluations have been handled with past administrations:

> *I don't think that there's ever not going to be complete and total fear that your administrator is not looking for a way to prove that you're not doing a good enough job. Many teachers feel that how their evaluation is going to go is based on their relationship with their administrator. There's a real fear, because again, teaching is such a personal thing."*

(Participant Four)

For all four participants, their initial feelings of nervousness and anxiousness that the teaching evaluation provoked were greatly, if not completely, reduced once the experience was completed. When one participant was asked whether or not his anxiousness wore off, he commented:

> *I now I look at evaluations the same way that I look at Christmas Eve. Christmas Eve has all the anticipation and it's usually better that Christmas morning itself. Usually, when it comes to nervousness, the anticipation of the nervousness is not as bad as what you're nervous about.*

(Participant One)

Participant two added, "I was actually surprised at how it was alright. I thought this isn't as bad as I thought it was going to be." This feeling of relief was echoed by participant four, "The process was not as bad as I thought."

Multiple teachers interviewed commented on how watching the video portion of the collaborative evaluation with their administrator was the most beneficial and meaningful part of the entire evaluative process.

> *I remembered to just remind myself that there are some things that I am pretty good at, which is hard for me. I'm much more practical and more apt to look at the glass half-empty. I'm more apt to look at the negative and identify the problems. I'm a nitpicker. But reflecting, I'm okay. I'm doing all right. There are some things that I am better than just okay at.*
>
> *(Participant four)*

Participant two agreed, "I really feel like watching yourself on camera is valid. We can always learn from ourselves."

The most veteran teacher interviewed provided a glimpse into how a new teacher fresh out of the credential program might perceive the collaborative evaluation process:

> *For me it was not too stressful, but I think for a first year person, it would probably be very stressful. There's the technology component, and the extensive rubric. And seeing yourself on video tape, you know, I would say is a bit a hurdle for a young teacher.*
>
> *(Participant 3)*

Technology: Barriers and Benefits

While all the participants felt varying degrees of relief once the evaluation process was completed, all the participants felt that the addition of video recording the lesson provided certain barriers and benefits. The most notable was the barrier that teachers experienced with the recording of the lesson, particularly with the audio.

The technology piece that didn't work was not my fault and made me super frustrated. I followed the instructions like I was told to do and then it didn't work. I was so angry and frustrated because I thought I had a really good lesson and I can't really go back and teach that lesson.

(Participant 2)

Participant one agreed, "… if the technical troubles were eliminated from the picture, I think people would have a much more positive outlook on the evaluation as a whole. It was the technical parts that were throwing a lot of people off, and stressing them out."

Another participant felt that the authenticity of the lesson was not captured by simply video and audio recordings. The same teacher stated that she taught multiple lessons to try and capture a typical lesson. She stated:

The lesson that was evaluated was my seventh attempt and it was a long time coming. I'm comfortable with videotaping any of my lessons, and just letting it be as it may. I still think that the human aspect is absent. The subtle nuances were missed by my principal.

(Participant 4)

Most teachers who participated ultimately felt that utilizing video and audio recordings provided more positives than negatives. When comparing the traditional evaluations that participant one had been experienced in his career with the most recent collaborative evaluation, he commented, "I liked it more because I was not looking over my shoulder wondering what my principal was writing down." He went on to add that he

felt that the video recorded lesson provided a more true and typical lesson than when an evaluator sits down in the back of the classroom.

> *I really feel that the process of being recorded and watched is an overall positive experience. I'd say it's a positive because it is better to see yourself along with someone else to simultaneously catch the good, catch the bad, and be able to discuss it.*
>
> *(Participant four)*

Desire for Increased Teacher-Principal Presence

Two of the four participants desired to see their principal inside the classroom on a more regular basis, and preferred the traditional evaluator sitting-in-the-back model of evaluation. Participant two stated, "I just wanted my boss to come and sit in the back of the classroom and observe me."

Another participant went into greater detail as to why she preferred the traditional model:

> *I feel like a lot of the nuances of the classroom are left out. Although I like the fact that we can watch the video and reflect on it, it's not necessarily me that I think always needs to be seen in the video of my classroom, it's my students. So I feel that the engagement in the conversation at the student tables was missing and I feel like if a principal (is) in there and they interact with the kids, they can get a better understanding of the classroom as a whole. There's an authentic piece that interaction that was empty.*
>
> *(Participant four)*

Conclusion

The purpose of this chapter was to highlight the findings which emerged after conducting

the interviews that were carried out with the teaching staff. It is clear from the findings

of this research that evaluations produce a great deal of anxiety. However, as highlighted

in this chapter, anxiety is often short lived and temporary, and once completed the

majority of teachers found value in the collaborative evaluation process. Also, mentioned

in this chapter were the technology related obstacles that presented frustrations, as well as

the two participants reporting that they preferred components of the more traditional

teaching evaluation, namely the observer being present in the classrooms.

DISCUSSION

A collegial working relationship between the teaching staff and school site leadership has an evidenced-based history of elevating student achievement. From the interviews conducted in this study, one could conclude that there was not a substantial boost in collegiality after participating in collaborative evaluations. It was clear that teachers found value in being able to watch themselves teach a lesson on video, however, only two of the four teachers mentioned the benefit of being able to review their lesson with their evaluator. It was also quite clear that regardless of what type of evaluation is being conducted, it's stressful and a sense of relief is felt once the evaluation is completed. The interviews also demonstrated a clear set of barriers experienced with video recording the lesson which produced frustrations.

As I reflect on the results of the interviews, one could wonder whether or not, over time, many of these barriers would be lessened. Surely, the issues surrounding the video recording of the lesson would reduce as needed modifications are made. During the course of the interviews it was noted that the collaborative evaluation process is in its infancy at this school site. Did the fact that this was a new evaluation process cloud the openness of the participants? It was clear that two to the four participants wanted greater administrative presence in their classroom. Perhaps that simple suggestion could improve collegiality on campus. Perhaps the evaluator could be present in the classroom while also video-taping the lesson.

CONCLUSION

   There were limitations to this study. This was only one elementary school in rural northern California and only four teachers were interviewed.   All of the teachers interviewed had been teaching for more than 13 years. The opinion of teachers who went through the teaching credential program post common core are absent.   In this researcher's opinion, further research should expand outside of rural California, while also increasing the sample size of the participants and including a greater range in teaching experience.

REFERENCES

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS® EVAAS®) in the Houston independent school district (HISD): intended and unintended consequences. *Education policy analysis archives, 20*,(12). doi:http://dx.doi.org/10.14507/epaa.v20n12.2012

Anrig, G. (2013). From health-care reform, lessons for education policy. *Education Week, 32*(36), 36,40.

Barnwell, P. (2015). Keep an open door, and other ways to build in-school relationships. *Education Week, 34*(21), 8-9.

Bell, J. (2001). High-performing, high poverty schools. *Leadership, 31*(1), 8-11.

Bill and Melinda Gates Foundation. (2011). *Learning about Teaching: Initial finding from the Measures of Effective Teaching project*. Bellevue, WA: Author. Retrieved from www.gatesfoundation.org/college-ready-education/Documents/preliminary-findings-research-paper.pdf

Bolin, F. S., & Panaritis, P. (1992). Searching for a common purpose: A perspective on the history of supervision. In C. Glickman (Ed.), Supervision in transition (pp. 30-43). Arlington, VA: Association for Supervision and Curriculum Development.

Bouffard, S. (2015). Learning from Our Students: Surveys offer performance feedback to teachers. *Harvard Education Letter 31*(3), 1-3, 6.

Bruce, R. F., & Hoehn, L. (1980, December). *Supervisory practice in Georgia and Ohio*. Paper presented at Annual Meeting of the Council of Professors of Instructional Supervision, Hollywood, FL.

Calandra, B., Brantley-Diaz, L., Lee, J.K., & Fox, D.L., (2009, Fall). Using video editing to cultivate novice teachers' practice. *Journal of Research on Technology in Education, 42(1), 73-94.*

Campbell, D.T. (1976). *Assessing the impact of planned social change.* Hanover, NH: The Public Affairs Center, Dartmouth College. Retrieved from http://portals.wi.wur.nl/files/docs/ppme/Assessing_impact_of_planned_social_ch ange.pdf

California Department of Education (2009). Commission on Teaching Credentialing. *California Standards for the Teaching Profession (CSTP) (2009).* Retrieved September 14, 2017, from https://www.ctc.ca.gov/docs/default-source/educator-prep/standards/cstp-2009.pdf

Coleman, E., (1945). The "supervisory visit." *Educational Leadership*, 2(4), 164-167.

Collegial. (2001). Webster's University College Dictionary. Springfield, MA: MERRIAM-WEBSTER INC.

Collins, C. (2014). Houston, we have a problem: teachers find no value in the SAS education value-added assessment system (EVAAS®). *Education Policy Analysis Archives, 22*(98). doi:http://dx.doi.org/10.14507/epaa.v22.1594

Colorado Department of Education. (2017, August 17). Colorado Department of

Education Home Page | CDE. *BEST PRACTICES Teacher Self -Assessments*

Understanding the Colorado State Model Evaluation Process and Teacher

Self -Assessments. Retrieved September 14, 2017, from

http://www.cde.state.co.us/educatoreffectiveness/selfassessmentbestpractices

Cubberely, E.P. (1929). *Public school administration* (3rd ed.) Boston: Houghton

Mifflin.

Danielson, C. (2016). It's Time to Rethink Teacher Evaluation. *Education Week,*

35(28), 20, 24.

Danielson, C. (2010). Evaluations That Help Teachers Learn. *The Effective Educator,*

68(4), 35-39.

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

(2nd ed.). Alexandria: Association for Supervision and Curriculum Development.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012).

Evaluating teacher evaluation: popular modes of evaluating teacher are fraught

with in accuracies and inconsistencies, but the field has identified better

approaches. *Phi Delta Kappan, 93*(9), 8-15.

Dewey, J. (1938). *Experience and education*, New York: Macmillan

Dewey, J. (1981). *The philosophy of John Dewey* (J. McDermott, Ed.) Chicago:

University of Chicago Press

Fehr, S. (2001, August). The role of educational supervision in the United States public schools from 1970 to 2000 as reflected in the supervision literature. Unpublished dissertation, Pennsylvania State University, State College.

Glatthorn, A. (1984). *Differentiated supervision*. Alexandria: Association for Supervision and Curriculum Development.

Glickman, C., Gordon, S. & Ross-Gordon, J. (1998*). Supervision of instructional: A developmental approach* (4th ed.) Boston: Allyn & Bacon.

Goddard, R., Goddard, Y., Kim, E.S., & Miller, R. (2015). A theoretical and empirical analysis of the roles of instructional leadership, teacher collaboration, and collective efficacy beliefs in support of student learning. *American Journal of Education, 121*(4), 501-530.

Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers.* New York: Holt, Rinehart & Winston.

Geiger, T., & Amrein-Beardsley, A. (2017). Administrators gaming test- and observation-based teacher evaluation methods: to confirm or confront the system. *American Association of School Administrators, 14*(3), 45-53.

Hiebert, J., & Stigler, J. (2017). Teaching versus teachers as a lever for change: comparing a Japanese and a U.S. perspective on improving instruction. *Educational Researcher, 46*(4), 169-176.

Hunter, M. (1984). Knowing, teaching, and supervising. In P. Hosford (Ed*). Using what we know about teaching* (pp. 169-192). Alexandria: Association for Supervision and Curriculum Development.

Kraft, M., & Gilmore, A. (2016). Revisiting the Widget Effect: Teacher Evaluation

    Reforms and the Distribution of Teacher Effectiveness. *Educational Researcher,*

    46(5), 234-249.

Marshall, K. (2017). "The Truth About Teacher Evaluation". *American School Board*

    Journal, 204(2), 76.

Marshall, K. (2015, October 24). The Marshall Memo: A Weekly Round-up of

    Important Ideas and Research in K-12 Education. *Effective Use of Teacher-*

    Evaluation Rubrics. Retrieved September 11, 2017, from

    https://www.marshallmemo.com/members/search2.php

Marzano, R. (2012). The Two Purposes of Teacher Evaluation. *Educational*

    Leadership, 70(3), 14-19.

Marzano, R., Frontier, T., & Livingston, D. (2011). *Effective Supervision*. Alexandria:

    Association for Supervision and Curriculum Development.

McFarlane, O. (2012). *The teacher principal relationship; the effect on student*

    outcome. Retrieved September 11, 2017, from

    https://www.academia.edu/9715632/Teacher_Principal_Relationship_the_effect_

    on_Student_Outcome_Introduction_to_Education_Administration

Milanowski, A. (2004). The relationship between teacher performance evaluation

    scores and student achievement: Evidence from Cincinnati. *Peabody Journal of*

    Education, 79(4), 33-55.

National Commission on Teaching and America's Future. (1996). What matters most:

    Teaching for America's future. New York, NY: Author

Noddings, N. (2014). High Morale in a Good Cause. *Association of Supervision for Curriculum Development, 71*(5), 14-18.

Quintero, E. (2017). Elevating Relationships: How Collaboration Shapes Teaching and Learning. *American Educator, 41*(2), 18-21.

Reinhorn, S., Moore, S., & Simaon, N. (2017). Investing in development: six high-performing, high-poverty schools implement the Massachusetts teacher evaluation policy. *Education Evaluation and Policy Analysis, 39*(3), 383-406.

Rosaen, C. L., Lundeberg, M., Cooper, M., Fritzen, A., & Terpstra, M. (2008). Noticing noticing: How does investigation of video records change how teachers reflect on their experiences? *Journal of Teacher Education, 59*(4), 347-360.

Ross, J., & Bruce, C. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching & Teacher Education, 23(2)*, 146-159. doi:10.1016/j.tate.2006.04.035

Schmoker, M. (1999). *Results: The key to continuous school improvement* (2nd ed.). Alexandria: Association for Supervision and Curriculum Development.

Schulz, J., Sud, G., & Crowe, B. (2014). Lessons from the Field: The Role of Student Surveys in Teacher Evaluations and Development. *Bellwether Education Partners,* Retrieved on September 11, 2017 from http://bellwethereducation.org/.

Sewall, M. (2009, Fall). Transforming supervision. Using video elicitation to support preservice teacher directed reflective conversations. *Issues in Teacher Education, 18*(2), 11-30.

Shafer, L. (2016). Teaching together for change: Five factors that make teacher teams

successful- and make schools stronger. *Useable Knowledge*. Retrieved on

September 11, 2017 from: https://www.gse.harvard.edu/news/uk/16/02/teaching-

together-change

Shah, M. (2012). The importance and benefits of teacher collegiality in schools - a

literature review. *Procedia - Social and Behavioral Sciences, 46*, 1242-1246.

doi:https://doi.org/10.1016/j.sbspro.2012.05.282

Slavin, R. & Madden, N. (2012). *Success for All*. Baltimore, MD: Success for All

Foundation.

Taylor, F.W. (1911). The principles of scientific management. Reprinted 2007, Sioux

Falls: NuVision

Thiers, N. (2017). Making progress possible: a conversation with Michael Fullan.

Educational Leadership, 74, 8-14.

Toth, T. & Rothman, R. (2008). *Rush to Judgement: Teacher evaluation in public*

education. Washington, DC: Education Sector.

Tracy, S. (1995, May/June). How historical concepts of supervision relate to

supervisory practices today. *The Clearing House, 68*(5), 320-324.

Tucker, P.D., & Stronge, J.H. (2005). *Linking teacher evaluation and student*

learning. Alexandra: Association for Supervision and Curriculum Development.

U.S. Department of Education. (2002) *No Child Left Behind (NCLB) Act of 2001,*

Pub. L. No. 107-110, § 115 Stat. 1425. Washington, DC: Author. Retrieved on

September 11, 2017 from http://www.ed.gov./legislation/ESEA02/

U.S. Department of Education. (2009). Race to the Top program executive summary.

    Washington, DC: Author. Retrieved on September 11, 2017 from

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

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

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

    Brooklyn: New Teacher Project. Retrieved September 4, 2017, from

    https://docs.gatesfoundation.org/Documents/the-widget-effect.pdf

Wertzel, W. (1929, February). Scientific supervision and curriculum building. *The*

    School Review, 37(2). 179-192.

Whitehead, M. (1952). Teachers look at supervision. *Educational Leadership*, 10(2),

    1011-1106

APPENDIX

Semi-Structured Collaborative Evaluation Interview Questions

1. Tell me about the most recent teacher evaluation that you have been a part of.

2. How did the two processes make you feel? Was it different than earlier evaluations?

3. If a new teacher was hired, how would you describe the collaborative evaluation

   process?

4. What do you remember most about your most recent evaluation?

5. How would you define collaboration?

6. What, if anything, surprised you about the evaluation?

7. Have you noticed any changes in student achievement as a result of what you learned
   in the evaluation process?

8. What would you expect to see in the future as a result of the collaborative evaluation?