Teacher Perceptions of Teacher Evaluations in the Fort Zumwalt School District

This manuscript has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of school administration and K-12 education.

Alex J. Tripamer
Alison G. Reeves
Elizabeth J. Meinz
Southern Illinois University Edwardsville

This study examines the perceptions of teachers in one Midwest school district relative to the current teacher evaluation system used in the district as well as a new model of teacher evaluation connected to Missouri's educator standards. To fully understand the perspective of the teachers, this study incorporates a mixed-methods approach which provides a variety of quantitative and qualitative data for analysis. Using survey data and focus group interview data, this study revealed that teachers perceive limitations associated with the current, more traditional evaluation process. Teachers understand benefits to the new Missouri model, but have reservations about its impending implementation. This study provides implications for the district as it moves forward with changing its teacher evaluation process.

Introduction

One of the most important responsibilities of a school administrator is conducting teacher evaluations that foster each teacher's ability to ensure student achievement (Darling-Hammond, 2000). Through effective evaluation, teachers have the opportunity to build their capacity in areas related to subject matter, affective behavior, and didactic processes (de la Rosa, 2005). However, previous research has provided evidence that the predominant practices in teacher evaluation do not always result in increased student achievement. Medley and Coker (1987) found a low rate of accuracy in school principals' judgments of the performance of teachers under their supervision. Principals are able to identify the best and

ISSN: 1532-0723

worst teachers, but are not able to distinguish teachers who occupy the middle range (Jacob & Lefgren, 2008). Other studies have demonstrated that traditional teacher evaluation methods are unable to provide meaningful understanding of teaching practices beyond sorting out those teachers lacking the most basic skills (Darling-Hammond, 1986).

Until recently, little research has been conducted to determine which types of teaching evaluations can lead to increased student achievement. The Gates Foundation (2010) has brought considerable attention to this issue within the last few years through the publication of their Measures of Effective Teaching project (the MET project). Additionally, some researchers have begun to craft teacher evaluation instruments that are specifically designed to assist teachers in improving student achievement (Danielson, 2009; Marzano, Frontier, & Livingston, 2011). In this study, we seek to add to the literature on current teacher evaluation models by studying the implementation of a new teacher evaluation system in a large, suburban, Midwestern school district.

Our first purpose of this mixed-methods study is to gain an understanding of how well the current teacher evaluation process is working for teachers in the Fort Zumwalt School District, to determine items in an evaluation process that would benefit the district's teachers, and to understand teachers' thoughts regarding a new teacher evaluation model proposed by the state of Missouri. Our second purpose is to understand teacher perceptions of the current evaluation process used in the Fort Zumwalt School District, as well as gain an understanding of how this same group perceives future changes to the teacher evaluation process. We sought to answer several research questions:

Quantitative Questions

1. To what extent do the teachers in the Fort Zumwalt School District feel the current evaluation system is beneficial to them and to their students? We predict that most teachers will perceive the current teacher evaluation process presently used by the Fort Zumwalt School District as not beneficial to them or their students.
2. To what extent do teachers in the Fort Zumwalt School District feel that characteristics associated with the *Missouri Model Teacher Standards* will lead to better teacher evaluations? We predict that teachers will view the content in new Missouri model as better criteria for evaluative feedback than the current system used in the district.

Qualitative Questions

3. What are teachers' perceptions of the current teacher evaluation process used in the Fort Zumwalt School District?
4. What are teachers' thoughts regarding a new teacher evaluation system aligned with the *Missouri Model Teacher Standards*?

Review of Literature

Historical Perspective

A formal method of describing teachers' traits and attributes emerged in the 1920s as Charters and Waples (1929) identified desirable traits and actions of teachers by generating characteristics to be used in the evaluation process. Beecher (1949) provided examples of teacher evaluations based on teacher characteristics that were perceived as valuable during this time frame. During the 1940s and 1950s, teachers were evaluated by "presage variables" or specific teacher traits (Danielson & McGreal, 2000).

Clinical supervision took root in the 1960s and 1970s with teacher evaluation models that included a pre-observation conference, classroom observation, analysis, post-observation conference, and analysis of the analysis (Marzano et al., 2011). These models led to more formal examinations of teaching and the evaluation of teachers using structured models such as Madeline Hunter's model of teaching as a standard for excellence. In the 1970s, Hunter's model gained recognition as a research-based methodology of instruction and teacher evaluation (Danielson & McGreal, 2000).

Current Trends in Teacher Evaluation Practice

The trend in teacher evaluation systems is moving from a teacher-centered set of behaviors, such as Hunter's teaching model, to a sharper focus on student learning (Danielson, 2009). There has been a great deal of research in the last 20 years examining the efficacy of various instructional strategies and their effects on student achievement (Hunter, 1982; Marzano, 2003). Much of the research on effective learning strategies has been included in recent teacher evaluation models (Danielson, 2009; Marzano et al., 2011; The New Teacher Project, 2010; Tucker & Stronge, 2005). Enough collective wisdom and technical savvy exists in the field of education to measure teacher quality and assess teachers both validly and reliably (Ingvarson & Rowe, 2008).

Two recent models of teacher evaluations include as a primary focus, effective learning strategies that lead to student growth. Danielson's Framework for Teaching and the Marzano Teacher Evaluation Model divide teaching attributes into various domains containing related behaviors. Both of these evaluation models emphasize quality conversations between the principal and teacher focused on teaching and learning (Danielson, 2007; Marzano et al., 2011).

Wiesberg, Sexton, Mulhern, and Keeling (2009) from the New Teacher Project wrote *The Widget Effect* to document their research regarding schools' lack of ability in assessing teacher effectiveness. They have defined the Widget Effect as "the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher" (p. 4). Their concerns led to the development of the New Teacher Project's teacher evaluation process standards known as *Teacher Evaluation 2.0* (2010). This model provides a framework that stresses the importance of incorporating student learning in teacher assessment.

Value-added Teacher Evaluation

In addition to requiring that states adopt rigorous standards for defining excellent teaching, new legislation contains specific requirements for the inclusion of student assessment data in the teacher evaluation process (American Recovery and Reinvestment Act, 2009). Tucker and Stronge (2005) detailed several school districts' attempts to assess teacher quality using student achievement and learning as a primary factor. They concluded that the use of evidence of student learning should be present in quality teacher evaluation systems.

Tekwe et al. (2004) labeled this method of incorporating student achievement in teacher evaluation as value-added stating,

> Value-added is a term used to label methods of assessment of school/teacher performance that measure the knowledge gained by individual students from one year to the next and then use that measure as the basis for a performance assessment system. (p.12)

Value-added models have emerged in the research as one way to statistically determine a teacher's effectiveness using various types of assessments as the primary data source (Gordon, Kane, & Staiger, 2006; Hill, Kapitula, & Umland, 2011; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Milanowski, 2011; Tekwe et al., 2004; & Toch & Rothman, 2008). Wright, Horn, and Sanders (1997) suggested that teacher evaluation processes should have a component detailing the teacher's effect on student growth. Despite the popularity of value-added models in the literature, some researchers have raised concerns with the use of these evaluation models (Amrein-Beardsley, 2008; Ishill & Rivkin, 2009). Specifically, they discovered issues related to insufficient validity of value-added research, inability for districts to implement the model, student background issues, inability to apply the model to all teachers, and classroom assignments as potential difficulties in using value-added measures for teacher evaluation.

Theoretical Framework

Two theories provide the theoretical foundation for this study: (a) the halo effect (Thorndike, 1920) and (b) Theory X and Y (McGregor, 1960). The halo effect was developed by E.L. Thorndike (1920) and was used to study military officers' evaluations of soldiers. The halo effect refers to the idea that that officers relied mainly on general perceptions of certain characteristics when evaluating soldiers. Lachman and Bass (1985) found a number of personality traits that might contribute to the halo effect of teachers by administrators (e.g., courtesy, fairness, intelligence, sincerity, tact, etc…). This study did not identify many of the instructional pieces so important to producing high student achievement, such as those contained in the *Missouri Model Teacher Standards*. A halo error can result from a principal's general impression of a teacher, thus influencing his/her evaluation of the teacher (Feeley, 2002).

Considering the halo effect in the context of this study, we would expect teachers to appreciate an evaluation based on clear standards, such as those outlined in the Missouri standards instead of an evaluation based on the perceptions of the principal. The task for principals is to separate any personal connections they may have with teachers and focus on

clearly defined strategies when providing feedback to teachers. In addition, a well-defined evaluation system that is grounded in research may not be well received by marginal teachers who have historically received positive evaluations due to their close personal relationship with their administrators.

Another theory useful to examine the motivation of teachers' willingness to improve their performance is McGregor's Theory X and Y (1960). Maslow's hierarchy of needs served as the basis for McGregor as he established this theory (Mattaliano, 1982). Theory X assumes that employees dislike work and managers consider employees to be inherently lazy. Theory Y assumes employees are ambitious, self-motivated, and exercise self-control.

Teachers generally view principals as holding Theory X assumptions when it comes to relations with staff members (Sabanci, 2008). The traditional model of teacher evaluations can be viewed from the standpoint of Theory X, while newer models of teacher evaluation which encourage a stronger professional relationship between the teacher and principal are consistent with Theory Y. As we conducted interviews with teachers regarding the district's current evaluation process and their perceptions of changes to the evaluation process, we noted how McGregor's theory applies, if at all, to teacher perceptions.

Method

Sample/Participants

Quantitative. We distributed an electronic survey to every teacher (n = 1300) in the school district in August 2012. Three hundred and fifty-eight completed surveys were returned resulting in a final return rate of 28%. Responses were received from teachers from every one of the district's 24 schools. We also collected information about the participants' total years of teaching experience and the grade(s) each participant teaches. The majority of the responses came from teachers with 18 years or less of teaching experience. High school teachers had a higher participation rate than middle and elementary school teachers.

Qualitative. We conducted focus group interviews at five schools in the Fort Zumwalt School District in September 2012. We selected two secondary schools and three elementary schools to ensure an equitable representation of K-12 staff within the district. Using the mean score from the survey for teacher perceptions of the *Missouri Model Teacher Standards*, we selected schools that appeared both eager and reticent about implementation of the new evaluation process. Membership of each focus group was comprised of teachers in various grades and subjects and varied from 6 to 11 participants. Focus group sessions were conducted at each school and lasted from 25 to 35 min in length.

Research Design/Overall Approach

This study utilized a mixed-methods approach. We collected survey data from teachers in the Fort Zumwalt School District regarding their perceptions of content contained in the district's current teacher evaluation process and content contained in the state of *Missouri's Model Teacher and Leader Standards*. Additionally, we collected qualitative data through the use of focus groups comprised of teachers from five district schools to gain their insights regarding the teacher evaluation process.

Data Sources/Instruments

Quantitative. For the quantitative piece of this study, we developed a 21-item survey. Items on the survey reflected specific criteria currently used by the Fort Zumwalt School District as well as criteria included in the Missouri model. Teachers rated each item on a seven-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree) regarding whether the particular item should be included in a teacher evaluation.

Qualitative. For the qualitative portion of this study, we conducted five semi-structured focus group interviews. The goal of the focus group sessions was to expand on the quantitative data and to gain the insiders' perspective on what the teacher evaluation process truly means to them. The questions used to guide the focus group discussions were designed to allow participants to share more specific information about their impression of the evaluation system currently used in the school district and allowed participants to share their thoughts and concerns regarding the implementation of the Missouri model.

Procedures

Quantitative. Permission to administer the survey was obtained from the school district. The survey was sent to teachers via email. Informed consent was obtained from each respondent. Teachers did not provide unique identifying information, thus ensuring confidentiality in their responses. Teachers had two weeks to respond to the survey. Survey data were exported the into a Microsoft Excel spreadsheet to be analyzed and processed for use in SPSS.

Qualitative. We invited teachers from the various schools to participate in focus group interview sessions. We met with each focus group at their school before the start of the school day, so as not to interfere with contractual obligations with the teachers. We obtained informed consent from each participant. We audio recorded each focus group session, and transcribed the recordings into Microsoft Word documents to be used in analysis.

Data Analysis

Quantitative. We used descriptive statistics to determine measures of central tendency and variability on all individual survey items and to calculate two summary variables. The summary variables provide an overall representation of teachers' perceptions for Fort Zumwalt-related evaluation content and Missouri model-related content. Each of the survey items were considered interval scales of measurement. We created two tables detailing descriptive statistics including the mean, standard deviation, minimum value, maximum value, and skew for each survey item.

To determine whether all items on the survey measured a common construct, we calculated Cronbach's alpha to determine the reliability of the overall survey. Additionally, we determined the reliability for two different groups of questions in the survey: (a) the items addressing teacher perceptions of the district's current evaluation system and (b) the items addressing teacher perceptions of the Missouri model. We compared these two groups of questions to determine if a relationship exists between the ratings of the two models. Finally, we conducted a *t*-test to determine whether there was a significant difference in the ratings of the models.

Qualitative. We transcribed each focus group interview and analyzed the transcripts, using open and focused coding techniques. To assist in organizing the transcript data, we also wrote five memos. Three overall themes emerged from the analysis of the transcripts.

Findings

Quantitative

To determine if the survey items measured a common construct, several Cronbach's alphas (measures of internal consistency reliability) were computed. Cronbach's alphas for the entire survey, survey items specific to the Fort Zumwalt teacher evaluation, and survey items specific to the Missouri model were .98, .96, and .96 respectively. Each of the surveys were found to be highly reliable.

Measures of central tendency were computed on all individual items and on two summary variables representing the aggregate of both Fort Zumwalt-related survey items and Missouri model-related survey models (see Tables 1 and 2). By looking at the mean scores from both tables, it can be inferred that there is overall high, positive regard for the content of both teacher evaluation instruments.

Table 1
Descriptive Statistics on Fort Zumwalt Evaluation Items

Item	*M*	*SD*	Min	Max	Skew
1	6.39	1.32	1	7	-2.92
2	6.54	1.20	1	7	-3.68
3	6.33	1.24	1	7	-2.82
4	6.16	1.31	1	7	-2.40
5	6.25	1.26	1	7	-2.65
6	6.02	1.48	1	7	-2.15
7	5.58	1.48	1	7	-1.25
8	6.30	1.38	1	7	-2.72
9	6.36	1.27	1	7	-2.99
10	5.60	1.58	1	7	-1.39
Average	6.15	1.14	1	7	-3.03

Table 2
Descriptive Statistics on Missouri Model Evaluation Items

Item	*M*	*SD*	Min	Max	Skew
1	6.63	1.15	1	7	-4.19
2	6.45	1.24	1	7	-3.25
3	6.23	1.25	1	7	-2.58
4	6.38	1.22	1	7	-3.13
5	6.58	1.20	1	7	-3.87
6	6.09	1.30	1	7	-2.29
7	6.07	1.39	1	7	-2.19
8	6.33	1.27	1	7	-2.90
9	5.89	1.49	1	7	-1.77
Average	6.29	1.12	1	7	-3.48

The item with the lowest score from the Fort Zumwalt evaluation was completely unrelated to the lowest scoring item from the Missouri model. The item scoring the lowest on the Fort Zumwalt evaluation was "maintains instructional materials and equipment" — the lowest item from the Missouri model was "participates in professional collaboration to support student learning." However, the top-rated items from both sets were very similar in that both were focused on providing effective instruction. The item from the Fort Zumwalt evaluation receiving the highest rating was "presents effective instruction and uses a variety of instructional methods" — the item from the Missouri model that received the highest rating was "understands content knowledge and aligns accordingly with appropriate instruction." Thus, the data suggest that receiving feedback relative to the instructional act is valued most by teachers.

We conducted a Pearson's correlation between teachers' views on the Fort Zumwalt survey items and views on the Missouri Model items, after applying a log10 transformation to reduce the considerable skew in the variables. The correlation between the transformed variables, $r = .87$, indicates a strong positive relationship between the two types of survey items. In other words, the more positive ratings a teacher had for Fort Zumwalt evaluation items, the more positive ratings they had for the Missouri model.

Finally, we conducted a paired-samples *t*-test to compare ratings on the summary variables (averages) for the Fort Zumwalt items and the Missouri model items. There was a statistically significant difference between ratings, $t(357) = -6.66$, $p < .01$; overall, the effect size is large with Missouri model items rated more highly than the Fort Zumwalt items (see Tables 1 and 2). The survey results indicate that teachers in the district prefer the content associated with the Missouri model over the content contained in the current district evaluation instrument.

Qualitative

Three major themes emerged from the focus group interviews. First, perceived liabilities exist within the current teacher evaluation system. Second, there is apprehension regarding the implementation of the Missouri model. Third, there are perceived advantages to the Missouri model.

Perceived liabilities exist within the current evaluation system. Although focus group participants appeared to have a clear understanding of the evaluation process, they perceived there to be problems with the current system. Specifically, the formal classroom observation process, the subjective nature of the evaluation process, and the lack of professional growth associated with the evaluation process were three disadvantages noted by the focus groups. Each of these perceived disadvantages will be expounded in further detail.

Classroom observation. Teachers perceive the formal classroom observation process as one of the weaknesses associated with the current evaluation system. Specifically, teachers expressed concern that their yearly evaluations are based solely on two formal visits from the principal. One teacher likened the formal classroom observation process to a "dog and pony show." Although some teachers appreciated their individual conversations with principals regarding formal classroom observations and final evaluation conferences, the majority of teachers felt these conversations lacked sufficient substance to promote professional growth. One teacher stated, "My principal will give me vague or no feedback at all about why I'm in

this column." Another teacher stated, "They [principals] want to be too nice" instead of providing constructive feedback.

Subjective nature of evaluation process. Another area of concern expressed by teachers is the perceived subjective nature of the evaluation process. The inconsistent feedback teachers receive from different principals is one example of subjectivity. One teacher shared her experience with having multiple evaluators over the course of her career and how each one had "their things they looked for, which was different." Another teacher reported how two principals who evaluated her gave her two completely different evaluations. Some teachers also reported that newer principals tended to evaluate teachers "harder" in their eyes.

Consistent with the halo effect, teachers felt that the preexisting relationship between the teacher and principal had a strong influence on the result of the evaluation. As one teacher shared, "If you have a good relationship with your administrator, then you'll probably end up with a better evaluation." From the teachers' perspective, some teacher evaluations have more to do with the level of the teacher-administrator relationship than it does with being able to demonstrate effective instruction.

Evaluations are not conducive to professional growth. While most teachers seemed complimentary of their principals, many were not in favor of the school district's current evaluation process or the instruments used to evaluate teachers. The following quotes illustrate some of the candid opinions of the evaluation process held by teachers:

- "I think the evaluation forms are ridiculous."
- "They [evaluation forms] don't help me."
- "I often get frustrated when I get these evaluations back."
- "I am not real impressed at the way we evaluate teachers right now."
- "I don't think it's [evaluation form] a very good tool."

There are items that principals must evaluate that are not of significant value for teachers, such as, "what does your classroom look like," "where are things stored," and "care of materials and supplies." The quantitative data from this study confirms these observations as the lowest-rated item on the survey was, "maintains instructional materials and equipment."

Apprehension regarding the implementation of the Missouri model.

Missouri's Department of Elementary and Secondary Education released the *Missouri Model Teacher Standards* during the summer of 2012. The Fort Zumwalt School District formally adopted the standards in September 2012. The timeline is important to the discussion of teacher apprehension because it provides context for the teachers' perspective. It was apparent, depending on the school, that there were varying degrees of teacher understanding about the state's new teacher evaluation model. Many teachers were not familiar with all of the nuances associated with the Missouri model.

Logistical concerns. There was some concern with the overall logistics associated with the actual process of the Missouri model. Teachers were curious to know whether or not there would be differences for tenured versus non-tenured teachers with the Missouri model. Another logistical concern expressed was the amount of time it will take principals to conduct evaluations on every teacher, every year. Teachers were concerned with the amount of time it will take principals to meet with teachers on a regular basis, conduct frequent walk-throughs, and review teacher documentation of evidence of student growth.

Measuring student impact. A significant piece of the Missouri model that lies in stark contrast with the current evaluation system is the emphasis placed on measuring the teacher's ability to effect student achievement, or simply stated "student impact." Focus group participants expressed concern with the student impact component. One teacher asked, "Are we going to gauge our effectiveness as an educator on that beginning to end-of-year growth?" Teachers were in favor of multiple assessment pieces such as portfolios, benchmark assessments, pre- and post-test measures, DRA reading scores, formative assessments, and summative assessments to document student growth throughout the school year.

Because of the emphasis on student impact, ethical considerations were raised. Teachers wanted to know who will have the gifted students assigned to their classrooms and who will have students with learning disabilities assigned to their classrooms. The concern was that if their performance is based on individual student performance, teachers might want students assigned to their classroom who perform better on assessments. They wanted assurance that students will be distributed equally among all classrooms. Despite the fact that our conversations never discussed employment decisions based on teacher evaluations, one teacher lamented that, "If raises are based on test scores and those evaluations, there will be cheating happening."

Professional considerations for implementing the Missouri model. To ease their apprehension regarding the implementation of the Missouri model, teachers were hopeful that the district administration will clearly communicate the roles and responsibilities associated with implementing the Missouri model for all parties. One teacher summed up her desire for this communication by stating that it would be beneficial to have, "a really good, comprehensive introduction to this with: this is what the administrators will do, this is what you [the teacher] will do, this is what we will do together."

Using the Missouri model, teachers will be evaluated on their performance for each item with one of four labels. The nomenclature used for these labels are "New Teacher," "Developing Teacher," "Proficient Teacher," and "Distinguished Teacher." Many experienced teachers were troubled by the idea that they may be evaluated as a "New Teacher" once the Missouri model is implemented. As one teacher said, "I think [the label] 'New' is insulting - insulting to anyone who has been doing it and working hard and trying for the last thirty-five years."

Special area teacher evaluations. Teachers who work in special areas such as music, physical education, art, and special education were concerned that the new evaluation instrument may not appropriately measure their performance. Because special area teachers and special education staff typically do not have the responsibility of administering state-wide standardized tests, the student impact component of the Missouri model surfaced as a concern for this population of teachers. One teacher mentioned, "How are we going to provide the right data to show? We're novices at this whole thing so far."

Advantages to implementing the Missouri model. In spite of their concerns regarding the implementation of the Missouri model, there are components contained in the model that appeal to teachers. Some of the advantages include professional growth for teachers, an increased focus on student achievement, and the emphasis on principal "walk-throughs." The higher level of emphasis the Missouri model places on professional growth is encouraging for teachers. Teachers perceive the current evaluation system as being weak in this particular area. As one teacher stated, "I don't think that it [the current evaluation system] allows for them to dig deep into their own professional growth."

The standards outlined in the Missouri model are more relevant to instruction and place an emphasis on student achievement. Thus, there is a connection between the teacher evaluation to district, building, and individual goals which makes sense to teachers. Teachers appeared motivated to personalize and develop their own goals in which they can demonstrate accountability towards student growth. Another connection from the quantitative side of this study corroborates the teachers' sentiments regarding the instructional piece. Teachers rated both the Fort Zumwalt evaluation item and the Missouri model item related to "providing effective instruction" highest on the survey.

One of the strategies consistently praised by the focus group participants was the use of walk-throughs by the principal. Walk-throughs can be simply defined as more frequent, less lengthy classroom visits by the principal. One teacher commented that this practice validates the importance of what teachers do and it also communicates to students what they are doing is important. One teacher said when the principal walks through her classroom, "that's when [they] really see the true reflection of your teaching." One teacher commented on the positive nature of this strategy, "I appreciate the principal being in the classroom so much more. Because when I get praise, and it's the same as the kids, I like to know if it's something I'm really doing well."

Discussion

What was Learned

The first two research questions were designed to determine the extent to which teachers feel the content of the current teacher evaluation in Fort Zumwalt is working and to what extent they feel the content of the new Missouri model will lead to better teacher evaluations. The mean score for all of the survey content related to the Fort Zumwalt evaluation process was 6.15 out of a possible 7; the mean for the content in the Missouri model was 6.29. Comparing the summary variables becomes the key to answering the question of which evaluation system is preferred by the teachers in the Fort Zumwalt School District. In practical terms, the raw difference of 0.14 between the two ratings may not be considered substantial to district administrators; however, the results of the *t*-test indicate statistical significance towards the new Missouri model leading to better teacher evaluations than the current evaluation instrument.

It is also notable that in general, teachers who rate the current evaluation system positively also rate the Missouri model in a positive light (r between the two systems = .87). This result could lead to a variety of conclusions. First, there are more similarities than stark differences regarding the actual content of the two evaluation systems. Second, teachers have accepted that the evaluation process is a necessary component of their professional life. Third, teachers who desire meaningful feedback and embrace accountability will rate whatever process is used highly while those teachers who do not will rate the evaluation process poorly.

In addition to the quantitative data, there is qualitative data to support the relative weakness of the content outlined in the Fort Zumwalt evaluation instrument. Teachers seemed genuinely pleased that the Missouri model places more focus on student achievement as well as professional growth. The qualitative data also supports the survey data revealing that teachers feel the Missouri model will be a better evaluation tool than the current evaluation system.

Questions three and four were designed to allow teachers to provide personal insight into how they perceive both the current evaluation and the Missouri model. The qualitative data differ from the quantitative data with respect to how the teachers in the Fort Zumwalt School District perceive the current teacher evaluation system. While the quantitative data focused, by design, on the *content* of the two evaluation systems, the qualitative data revealed many perceived weaknesses associated with the *process* of the current evaluation system. It was evident that focus group participants had concerns regarding a number of evaluation processes contained within the current evaluation instrument such as formal classroom observations and the subjective nature of the evaluation process.

The evaluation system currently used within the Fort Zumwalt School is similar to the literature's description of clinical supervision from the 1960's and 1970's. Many teachers in the focus groups echoed findings from Marshall's (2005) research on this process. Among the limitations expressed by teachers include (a) ineffectiveness with current evaluation practices such as classroom observations, (b) receiving meager feedback from administrators, and (c) receiving evaluation documents that do not improve their practice.

Teachers do not perceive two annual formal, and often pre-scheduled, classroom observations provide sufficient data for principals to conduct a valid evaluation of their yearly performance. The alternative preferred method voiced by the teachers to counter this process is the implementation of regular walk-throughs by the principal. Teachers feel that if the principals are able to observe more frequent and natural classroom teaching acts, then they will have a better impression of what is actually going on in the classroom. Certainly, regular visits by the principal as opposed to infrequent visits that are scheduled in advance would eliminate the "dog and pony show," as one teacher described the current process. The benefit of frequent and unscheduled classroom visits by the principal is that over time, the principal obtains a better understanding of classroom activities throughout the entire school.

It is also important to recognize that observational data alone will not be sufficient for evaluating teachers using the Missouri model. The use of student achievement data is an equally important piece to this model. Teachers will need to provide artifacts demonstrating student growth. This process will require the teacher and principal to engage in meaningful discussions relative to all data collected regarding a teacher's performance. This collaboration allows the teacher and principal to mutually arrive at the teacher's level of performance based on both observational data and student achievement data.

Many teachers expressed concern with the distribution of students in their classrooms and its impact on an overall measurement of the teacher's effectiveness. It will be important to create classrooms that are equally representative of all academic levels. Random student assignment is not a sufficient method to accomplish this goal. An alternate strategy would be to assemble teams of teachers to collaborate on the process of creating classes so that each classroom has a balanced group of students.

The subjective nature of the current evaluation system was frequently discussed as a weakness by the teachers in this study. To counter the subjective nature of evaluations, the Missouri model offers two pieces that should be seen as a favorable change to teachers. The first item is a self-evaluation by the teacher on the three standards in which they will be evaluated. The second item contained in the Missouri model is the requirement by the teacher and principal to show evidence aligned with each level of performance. Both parties have access to a rubric that clearly defines each teacher rating along with examples of acceptable evidence.

Because student impact is a significant component in the Missouri model and because it marks a pointed departure from the current evaluation system, the concerns expressed by the teachers regarding what this piece will eventually encompass are reasonable. Teachers would like to see a multitude of assessment measures for gauging student impact as opposed to using just the state's end-of-year assessment. In particular, teachers of special area classes such as music, physical education, art will need to provide their own alternate assessments to demonstrate student impact.

Connections to Theoretical Framework

Consistent with the pitfalls of the halo effect, many teachers felt that some of their colleagues received positive evaluations due to their close friendship with the principal. By implementing the Missouri model that (a) requires the submission of artifacts demonstrating teacher effectiveness and (b) contains a clear set of evaluation criteria, the halo effect can be reduced. As indicated previously, the negative attributes of this phenomenon were clearly observed by focus group participants in their own evaluative experiences.

In the current evaluation system, teachers are not evaluated based upon student growth or performance. Measuring student impact becomes an important element of the Missouri model with respect to limiting the halo effect. For example, when administrators and teachers use student assessment data to measure student growth over the course of the year, the clarity of that data is likely to counteract a significant portion of the effect of the interpersonal relationship between the teacher and administrator. Of course, the importance of the numbers could become exaggerated and be used for inappropriate conclusions themselves, but the use of quantitative data is likely to reduce the impact of personal relationships on the evaluation outcomes.

The Missouri model also provides teachers and administrators with a rubric containing straightforward criteria describing each level of teacher performance for each indicator contained within the model. The principal and teacher collaboratively use these criteria to determine the teacher's rating. The current evaluation system lacks a set of thorough criteria — teachers are often evaluated on each item based on administrator opinion, which of course, makes them susceptible to the halo effect.

Principals' views of teachers will make the biggest difference towards the successful implementation of the Missouri model. Administrators have to believe this process will lead to teacher growth, and ultimately, student growth. It is important to note that merely adopting the Missouri model may not persuade principals holding Theory X beliefs to completely change their view towards teachers. However, the potential exists for principals to change their Theory X views with some teachers through additional classroom observations, data collection, and collaboration.

Although the Missouri model is naturally conducive to a Theory Y atmosphere, it will require an administrator who embraces this type of thinking and it is not a given that the Missouri model will automatically change principals' views. Teachers genuinely appeared eager for an evaluation system that focuses on student achievement and one that engages them professionally with their principals. Currently, there may be surface-level discussions between administrators and teachers based upon formal classroom observations. In order for the Missouri model to reflect the ideals of Theory Y, the conversations between

administrators and teachers must move to more meaningful discussions of classroom performance and they need to be more collaborative than what currently exists.

Implications

Missouri law mandates that all school districts alter their teacher evaluation systems by the 2014-2015 school year. At this point, the Fort Zumwalt School District has opted to use the Missouri model teacher evaluation as the district's new evaluation system. The teachers who participated in this study indicated several important considerations that could inform implementation of the Missouri model in their local district.

First, district administrators should consider providing in-depth professional development related to implementing the new evaluation system. As the focus group members indicated, district administrators need to clearly communicate roles and responsibilities for both teachers and building-level administrators. Next, the district can also assist teachers by defining which assessment data will be acceptable in terms of documenting student impact. However, teachers should also have the autonomy to select other pieces of evidence that demonstrate student growth. Allowing teachers the leeway to select their own evidence of student learning is consistent with a Theory Y framework and provides alternatives to documenting student impact required within the Missouri model. District administrators will also need to determine how special area teachers will be evaluated. Because student impact is a mandatory component of the Missouri model, the district will need to determine what kinds of assessment activities are appropriate for each discipline.

The district needs to recognize the substantial nature of change when it moves from the current evaluation system to the Missouri model. In order to address the issues previously mentioned, the district should use the 2013-2014 school year to learn about the new evaluation system. Finally, the discussion of school climate becomes paramount when discussing the effectiveness of implementing a new evaluation system. Because the Missouri model offers such a major departure from the current evaluation system, district and building administrators will need to create school climates that are conducive to positive change. Under the processes outlined in the Missouri model, school administrators will have to demonstrate collaborative leadership that may be uncomfortable for some and more natural for others.

Consideration for Further Research

Because this research is concentrated on one school district in Missouri, it is difficult to generalize the results to a broader audience. An interesting topic for future research would be to gather data from other Missouri school districts who have adopted the Missouri model to determine its effectiveness in terms of student achievement and teacher satisfaction. This research would allow the Department of Elementary and Secondary Education to determine whether or not the Missouri model evaluation system is making a positive difference in increasing student achievement.

Conclusion

The nature of teacher evaluations is transforming to adhere to current legislative standards and to validate recent research efforts in the area of effective evaluative practices. This study gives a voice to the teachers who will be evaluated with a process that differs greatly from traditional evaluation methods. It is evident that teachers desire an evaluation process that allows them to grow professionally and contemporaneously leads to greater student achievement.

It is encouraging to see the field of education moving to evaluative practices that focus on the teacher's impact on student learning. Specifically, the Missouri model will allow teachers to focus on a very limited number of goals each year that are tied to student growth. As one focus group participant shared, "Our [former] evaluations seemed to go a mile wide and an inch deep; now we might be able to go an inch wide, but a mile deep." Teachers appreciate the advantage the Missouri model will have relative to its impact on students.

School administrators who find themselves in the midst of implementing new teacher evaluation systems can use the information from this study to plan effectively. Of particular value are the findings that address thoughtful measures of student impact, conducting in-depth professional development, and providing insightful feedback to teachers. The teachers who participated in this study appear supportive of implementing a new evaluation model. Both the quantitative and qualitative data reveal teachers' frustration with the current, outdated evaluation process. The same data were used to indicate the teachers' support of an evaluation process that will ultimately lead to improved instructional practices. This final point is important when considering the essential goal of linking the teacher evaluation process to increased student achievement.

References

American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115.

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37(2),* 65-75.

Beecher, D. E. (1949). *The evaluation of teaching.* Syracuse, NY: Syracuse University Press.

Charters, W.W., & Waples, D. (1929). *The commonwealth teacher-training study*. Chicago, IL: The University of Chicago Press.

Danielson, C., & McGreal, T.L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2009). A framework for learning to teach. *Educational Leadership, 66*, 1-8.

Darling-Hammond, L. (1986). A proposal for evaluation in the teaching profession. *The Elementary School Journal, 86,* 530-551.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives, 8(1),* 1-44.

De la Rosa, P. (2005). Toward a more reflective teaching practice: Revisiting excellence in teaching. *Asia Pacific Education Review, 6(2).* 170-176.

Feeley, T. H. (2002). Comment on halo effects rating and evaluation research. *Human Communication Research, 28,* 578-586.
Gates Foundation (2010). Working with teachers to develop fair and reliable measures of effective teaching. Retrieved from http://www.metproject.org/downloads/met-framing-paper.pdf.
Gordon, R., Kane, T. J., & Staiger, D. O. (2006). Identifying Effective Teachers Using Performance on the Job. The Hamilton Project Policy Brief No. 2006-01. *Brookings Institution*, Retrieved from EBSCO*host*.
Hill, H.C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48(3),* 794-831.
Hunter, M. C. (1982). *Mastery Teaching.* El Segundo, CA: TIP Publications.
Ingvarson, L. & Rowe, K. (2008). Conceptualising and evaluating teacher quality: Substantive and methodological issues. *Australian Journal of Education, 52(1),* 5-35.
Ishill, J, & Rivkin, S. G. (2009). Impediments to the estimation of teacher value added. *Education Finance and Policy, 4(4),* 520-536.
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.
Lachman, S. J., & Bass, A. R. (1985). A direct study of the halo effect. *The Journal of Psychology, 119(6). 535-540.*
Marshall, K. (2005). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan, 86(10),* 727-735.
Marzano, R. J. (2003). *What works in schools: translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.
Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.
Mattaliano, A.P. (1982). Theory x or theory y – what is your style? *NASSP Bulletin, 66(37),* 37-40.
McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics. 29(1),* 67-101.
McGregor, D. (1960). *The human side of enterprise.* New York, NY: McGraw-Hill.
Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research. 80(4),* 242-247.
Milanowski, A. T. (2004). The relationship between teacher performance evaluation scores and student achievement: evidence from cincinnati. *Peabody Journal of Education, 79(4)*, 33-53.
Milanowski, A. T. (2011, March). *Validity Research on Teacher Evaluation Systems Based on the Framework for Teaching.* Paper presented at the American Education Research Association annual meeting, New Orleans, LA.
Missouri Department of Elementary and Secondary Education (2012). Executive Summary Missouri's Educator Evaluation System. Retrieved from http://dese.mo.gov/eq/documents/eq-ees-executive-summary.pdf.
Missouri Department of Elementary and Secondary Education (2012). *Research and proven practices*. Retrieved from http://dese.mo.gov/eq/documents/eq-ees-resources.pdf.

Missouri Department of Elementary and Secondary Education (2012). *Teacher evaluation*. Retrieved from http://dese.mo.gov/eq/documents/eq-ees-teacher-evaluation.pdf.

Sabanci, A. (2008). School principals' assumptions about human nature: implications for leadership in turkey. *Educational Management Administration & Leadership, 36,* 511-529.

The New Teacher Project (2010). *Teacher evaluation 2.0.* Retrieved from http://tntp.org/files/Teacher-Evaluation-Oct10F.pdf.

Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M. E., Roth, J., Ariet, M. Fisher, & Resnick, M.B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29(1),* 11-36.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology 4(1),* 25–29.

Toch, T. & Rothman, R. (2008). Rush to Judgment: Teacher Evaluation in Public Education. Education Sector Reports. *Education Sector*, Retrieved from EBSCO*host*.

Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

Weisberg, D., Sexton, S, Mulhern, J., & Keeling, D. (2009). *The Widget Effect*. Retrieved from http://widgeteffect.org/downloads/TheWidgetEffect.pdf.

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