Declining Morale, Diminishing Autonomy, and Decreasing Value: Principal Reflections on a High-Stakes Teacher Evaluation System

Noelle A. Paufler, *University of North Texas*

Abstract

Since the adoption of teacher evaluation systems that rely, at least in part, on controversial student achievement measures, little research has been conducted that focuses on stakeholders' perceptions of systems in practice, specifically school principals. This study was conducted in a large urban school district to better understand principals' perceptions of evaluating teachers based on professional practice and student achievement. Principals in this study strongly expressed concerns regarding: a) the negative impact of the teacher evaluation system on morale; b) their lack of autonomy in evaluating teachers and making staffing decisions; and c) their perceived lack of value as professionals. Examining the implications of teacher evaluation systems is increasingly important to better understand the intended and unintended consequences of these systems in practice.

Keywords: Teacher evaluation; Accountability; Educational policy; Value-added models; School leadership; Principal voice

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

Introduction

In recent years, growing public demand for quantifiable measures of school and teacher effectiveness has dominated the policy debates surrounding educational accountability in the United States and other countries. In response to the widespread criticism of traditional evaluation systems using teacher education and credentials (Tucker & Stronge, 2005), many states have adopted teacher evaluation systems that rely, at least in part, upon complex statistical measures of student achievement, such as value-added models (VAMs), in addition to measures of teachers' professional practice (Amrein-Beardsley, 2008, 2014; Braun, 2005; Lavigne, 2014; Lavigne & Good, 2014). Despite the recent passage of the Every Student Succeeds Act (ESSA, 2016), which eliminates the federal mandate in the United States that teacher evaluations are linked to students' test scores, some states are still moving forward with stronger accountability reforms (Felton, 2016; Loewus, 2017; Will, 2016).

Regardless of the intended and unintended consequences associated with recruiting, hiring, promoting, developing, and retaining teachers based on such evaluation outcomes, what is grossly missing from the research literature about these measures are the school administrator and teacher perceptions of them in use (Harris & Herrington, 2015; Hopkins, 2016; Kraft & Gilmour, 2016). In other words, how principals and teachers experience and perceive these systems and related high-stakes consequences remains largely unexamined, and consequently ignored, at multiple policy levels.

This article discusses the findings and implications of a survey research study conducted in a large urban school district to examine the perceptions of principals toward their district's teacher evaluation system. At the time of this study, the school district, one of the largest in its state and among the largest urban districts in the United States, evaluated teachers of core content areas (e.g., reading, writing, mathematics, and science) based on their professional practice (25%), instructional practice (25%), and value-added scores (50%) (e.g., the SAS® Education Value-Added Assessment System [EVAAS®]). While VAM-based outcomes are still wrought with criticisms, mainly about the validity, reliability, potential bias, and fairness of measure outputs, evaluation results have been and are increasingly being used to make high-stakes employment decisions in this district and across the nation (Amrein-Beardsley, 2014; Baker, Oluwole, & Green, 2013; Berliner, 2014; Corcoran, 2010; Herlihy, Karger, Pollard, Hill, Kraft, Williams, & Howard, 2014; Lavigne, 2014; Lavigne & Good, 2014).

Framework

The Measure and Punish (M&P) Theory of Change provides the framework for understanding the policy context in this study. Although generally conceived of as a colloquial expression, this M&P Theory of Change, as operationalized by Audrey Amrein-Beardsley (2014) and applied here, suggests that "by holding districts, schools, teachers, and students accountable for performance on the states' large-scale standardized achievement tests, administrators will supervise the schools better, teachers will teach better, and students will learn more, particularly in the nation's lowest performing schools" (p. 72). Based on this logic, change in performance can be evoked by a series of rewards and punishments linked to measured outcomes. In

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

essence, students will learn more if they are better taught, but teachers will not teach better unless they are subject to high-stakes consequences for poor outcomes (Smith, 2004). Accordingly, federal and state policy that holds educators accountable, in this case through a high-stakes teacher evaluation system, will ensure that students meet higher standards of learning, as measured by their performance on high-stakes tests (Amrein-Beardsley, 2014; Smith, 2004).

In short, this M&P Theory of Change suggests that students, especially those in the highest-need schools and districts, will learn more as a result of a stronger accountability policy. As such, perceptions of stakeholders as the primary recipients of, and actors within, larger, complex evaluation systems driven by this theory of change, specifically with regards to system implementation, must be better understood given the myriad intended and unintended consequences associated with evaluation outcomes. The findings in this study, based on the perceptions of principals in one large urban school district regarding their district's teacher evaluation system, give voice to stakeholder concerns about the rationality of this M&P Theory of Change in practice.

Literature review
Historical background

Despite ongoing efforts to improve the quality of schools and increase student learning in the United States by evaluating and developing the skills of teachers, the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983) purported that low student achievement was placing the nation "at risk" of continued economic decline. Among several recommendations outlined in the report, the National Commission on Excellence in Education (1983) suggested that comprehensive evaluations and increased salaries would foster higher expectations of professional competence among teachers. In *Action for Excellence*, the Task Force on Education for Economic Growth from the Education Commission of the States (ECS, 1983) reiterated the importance of teacher competency cited by the national commission and called for the development and implementation of "systems for fairly and objectively measuring the effectiveness of teachers and rewarding outstanding performance" (p. 39). The task force's emphasis on evaluation with an ancillary focus on the need to professionalize teaching reaffirmed the belief that "better teachers and better teaching" (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985, p. 62) are the primary factors in educational improvement.

As part of an ongoing effort to improve student learning outcomes, the goals for standards-based educational reform, supported by former president George H.W. Bush and subsequent administrations, were first outlined in Goals 2000 and later incorporated into the updated Elementary and Secondary Education Act (ESEA), under the No Child Left Behind (NCLB) Act (2002). By mandating that all students demonstrate proficiency on state-determined standards in reading and mathematics by 2014, the NCLB Act spurred the development of large-scale standardized tests in every state for the purposes of measuring student learning and, ultimately, school and teacher quality (David & Cuban, 2010).

Following the passage of the NCLB Act, states were prompted to develop and implement accountability systems to measure school and teacher quality in order to

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

be eligible for Race to the Top (RttT) and Teacher Incentive Fund (TIP) grants from the United States Department of Education (2009, 2010). Accordingly, states were required to provide evidence of compliance with regards to their accountability system, which had to rely, at least in part, on student performance on large-scale state-level standardized tests (Amrein-Beardsley 2008, 2014; Braun, 2005; Corcoran, 2010; United States Department of Education, 2009). Subsequently, states across the nation developed and implemented such systems based on quantitative measures of teacher and school effectiveness, often associated with high-stakes consequences (Amrein-Beardsley & Collins, 2012; Baker, Oluwole, & Green, 2013; Berliner, 2014; Corcoran, 2010; Weisberg, Sexton, Mulhern, & Keeling, 2009).

Current policy context

As part of a new wave of accountability policy, ESSA (2016) recently eliminated the federal mandate in the Unites States that teacher evaluation systems must be linked to student test scores; however, more than one year after the act's passage, most states are still mired in debate, and in some states litigation, surrounding the use of measures of student achievement to evaluate teachers (Hazi, 2017; Loewus, 2017; see also Will, 2016). ESSA provides states with greater flexibility in terms of how to evaluate teachers, and as a result, states are "all over the map" (Will, 2016, p. 31) with regards to how they intend to proceed. As of late 2017, some states (e.g., Indiana, Louisiana) are reexamining the use of student test scores in teacher evaluations, while others (e.g., Alaska, Arkansas, Kansas, Kentucky, North Carolina, Oklahoma) have discontinued or moved away from this practice, at least to some extent (Loewus, 2017). For example, teachers in Kentucky may still receive student growth data, but its inclusion in their evaluation is at the discretion of their districts (Loewus, 2017). Teachers in other states are still required to provide evidence of student learning (e.g., Connecticut, Nevada, Utah), but they cannot use students' scores on state standardized tests as evidence. Legislatures in Arizona, Maine, and New Mexico approved bills to discontinue or reduce the weight of student achievement, but the governor of each state vetoed the respective legislation (Loewus, 2017). Perhaps most concerning, litigation is or was recently underway (at various stages in the judicial process) in six states (Colorado, Florida, New Mexico, New York, Tennessee, and Texas) that challenges the use of student growth measures in teacher evaluations for high-stakes decision-making (Hazi, 2017).

Local context

As mentioned previously, the school district in this study is one of the largest in its state. It is among the largest urban districts in the United States, currently consisting of almost 300 schools, enrolling more than 200,000 students, and employing more than 10,000 teachers. The district serves a diverse student population in terms of racial/ethnic background (more than 60% Hispanic, more than 20% African American, and less than 15% White, Asian, American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, and two or more races). Many of the students in the district are also considered high-needs based on their qualification for Title I funds (more than 90%), eligibility for free or reduced lunch (more than 70%), at-risk status

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

(more than 70%), limited English language proficiency (more than 30%), or eligibility for special education services (more than 5%).

For more than five years, students in this state, depending upon their grade level, have been assessed in core content areas using the state standardized tests. Annual assessments include reading and mathematics (grades 3–8), writing (grades 4 and 7), science (grades 5 and 8), social studies (grade 8), and high school end-of-course assessments (in English I and II, Algebra I, Biology, and United States History). At the time of the study, value-added scores (i.e., generated via the SAS® EVAAS®) for teachers in this district who taught core content areas (e.g., reading, writing, mathematics, science) comprised 50 percent of their evaluation score.

While traditional measures of achievement are based on the growth observed on an individual student's or cohorts of students' large-scaled standardized test scores at one or two points in time, VAMs attempt to isolate the effects of individual teachers on their students' learning from one year to the next (Amrein-Beardsley, 2008, 2014; Braun, 2005; Harris, 2011; Papay, 2010; Scherrer, 2011) by predicting the student's performance on a test based on prior achievement and (often) background characteristics (e.g., racial or ethnic background, special education needs, English language proficiency, socioeconomic status) (Amrein-Beardsley, 2008; 2014; Braun, 2005; Harris, 2011; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Papay, 2010; Scherrer, 2011).

Although VAM-based outcomes remain controversial in terms of their validity, reliability, potential bias, and fairness, this district has been widely recognized for its historically high-stakes use of EVAAS® scores for making personnel decisions. Teachers have been subject to penalties and sanctions (e.g., related to teacher merit pay, retention) based on their evaluation scores. In fact, the district has made adverse employment decisions for teachers based on their low VAM scores. As such, perceptions of stakeholders, principals in this case, must be better understood given the intended and unintended consequences associated with evaluation outcomes in this district in practice.

Methods
Research design
In this survey research study, the researcher worked with a third party to design and administer an online survey examining principals' perceptions of their district's teacher evaluation system (see Appendix), specifically to better understand their perceptions of evaluating teachers based on professional and instructional practices as well as student achievement (i.e., value-added scores). The single-strand mixed-methods design used in this study included an online survey instrument with both closed- and open-ended items, which was administered to all principals in the district.

Data collection
Survey instrument
In terms of instrument design, the survey included 12 closed-ended questions and one open-ended question. The survey questions were broadly aligned to the following constructs: a) teacher/principal morale and district/school culture (questions 1, 2,

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

and 9); b) principal autonomy (questions 3, 4, 5, 6, 7, and 8); and c) principals' plans for future employment (questions 10, 11, and 12). The open-ended survey item invited respondents to provide additional comments, if they were so inclined. It is important to note that the structure of some of the closed-ended survey items, and the alignment of the overarching research question (i.e., principals' perceptions of their teacher evaluation system) to those items, reflects the interests of the third-party survey administrators. As such, it is important to acknowledge the role of the third party and the researcher's positionality in the context of this study (Creswell, 2014).

Researcher positionality

The researcher is/was not directly affiliated with either the district or the third party in this study. Rather, the third party was a professional organization indirectly affiliated with the district (i.e., it had the direct knowledge of the teacher evaluation system development and implementation necessary for survey design, and it had permission to use available contact information to disseminate the survey to prospective participants). The researcher's primary responsibility included analyzing the quantitative and qualitative survey data collected to triangulate findings and draw conclusions, independently from both the third party and the district.

The researcher's external role in data collection and analysis merits some specific considerations. First, serving in an external role in data collection helped to mitigate the potential dilemmas of conducting "backyard research" (Glesne & Peshkin, 1992, pp. 22–23). Whereas direct affiliation with either the district (e.g., as an employee) or the third party (e.g., through membership in the organization or receiving compensation for conducting the analysis) might have introduced bias, the researcher's positionality lends credibility to data analysis in terms of triangulating the quantitative and qualitative data (Creswell, 2014; Lincoln & Guba, 1985). However, the researcher also acknowledges potential limitations in data interpretation as a result of the language used in some items (e.g., yes/no, Likert-type response options). With this in mind, readers are encouraged to consider the voice of participants as reflected in both closed- and open-ended responses and evaluate the study findings for their potential contribution to the gap in existing literature.

Participants

In total, 273 principals were invited to participate in the survey, and of these, 70 responded (25.6%), including principals from high (n = 9/70, 12.9%), middle (n = 13/70, 18.6%), elementary (n = 40/70, 57.1%), and other (e.g., charter, combined; n = 8/70, 11.4%) schools. Relative to the total number of principals invited to participate, the response rate for principals by type of school is as follows: high (n = 9/44, 20.5%), middle (n = 13/40, 32.5%), elementary (n = 40/147, 27.2%), and other (n = 8/42, 19.0%). Middle schools had the highest principal response rate (32.5%), and other (e.g., charter, combined) schools had the lowest response rate (19.0%).

Given this response rate (25.6%), it is conceivable that responses may reflect the sentiments of principals who are most troubled by the evaluation system in their district; thus, results might represent the extremes in the population, and they may not generalize to the larger population of district administrators. Additionally, the survey

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

instrument did not include demographic questions to help mitigate anticipated concerns about anonymity. Accordingly, the researcher was not able to determine sample representativeness based on demographic characteristics. It is important to note that this study was conducted so that readers might make more naturalistic generalizations from the findings (Stake & Trumbull, 1982), versus statistically significant inferences from the sample to the population. Accordingly, study results do merit consideration given the high-stakes consequences attached to value-added output in this district, in practice. It is this set of perceptions, as stated, that are still missing from many of the conversations surrounding value-added uses and consequences.

Data analysis

The researcher analyzed respondent survey data by calculating descriptive statistics based on participants' responses to the 12 closed-ended questions and by coding qualitative data using three rounds of "constant comparison" to develop codes and categories. The researcher then constructed a list of common themes based on the instances appearing in the raw data (Erickson, 1986; Glaser & Strauss, 1967; Smith, 1997). Specifically, the researcher read and pre-coded the entire corpus of qualitative survey data to highlight significant participant quotes (Layder, 1998), before manually coding each participant response using a structural coding method. The researcher selected structural coding for the first coding cycle as a means of "cod[ing] and initially categor[izing] the data corpus to examine comparable segments, commonalities, differences, and relationships" (Saldaña, 2013, p. 84), especially given the data set included multiple participants.

In the second coding cycle, the researcher used focused coding to identify the most frequent codes (e.g., morale, climate, pressure), develop categories (e.g., teacher recruitment and retention, the role of external decision-makers) (Charmaz, 2006), and generate themes (Erickson, 1986). These themes included: a) the impact of evaluation processes and outcomes on morale; b) the district-level pressures on principals with regards to teacher recruitment, hiring, promotion, and retention; and c) the (lack of) value placed on principals as professionals. After composing assertions and searching for confirming and disconfirming evidence in the data, the researcher extracted respondent quotes to warrant the assertions (Erickson, 1986; Smith, 1997; see also Saldaña, 2013).

Results

Impact on teacher and principal morale

Principals reportedly perceived the evaluation process and related outcomes as having a generally negative impact on morale. At least one out of three survey respondents ($n = 24/70$, 34.3%) indicated that they are not able to recruit and keep staff at their school, and only 5.7 percent of respondents ($n = 4/70$) agreed that the district's culture, reportedly influenced by the use of teacher evaluation data per principals' open-ended responses, contributed to high morale among their teaching staff. Relatedly, 88.6 percent ($n = 62/70$) agreed that, rather, there is a culture of intimidation in the district fostered by this particular evaluation system. See Table 1 for the proportions of responses to these survey questions.

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

Table 1. Principal respondents' perceptions on recruitment and retention, morale, and district culture

Statement	Yes	No	Not Applicable	*n*
Question 1: I am able to recruit and keep staff at my school.	46 (65.7)	24 (34.3)	0 (0.0)	70 (100.0)
Question 2: [Name of school district]'s culture contributes to high morale among my teaching staff.	4 (5.7)	65 (92.9)	1 (1.4)	70 (100.0)
Question 9: I feel there is a culture of intimidation in [name of school district].	62 (88.6)	8 (11.4)	0 (0.0)	70 (100.0)

Note: Counts are presented as raw numbers with respective valid proportions of the total participants in parentheses. Valid proportions may not total 100 due to rounding.

Concerns about low teacher morale were commonly expressed in open-ended responses as well. One respondent explained that there is a "culture of 'mean' in this district … [that] is competitive and has caused many teachers to think about 'my' students rather than 'our' students as a whole." Another respondent described morale as "the lowest in the history of our district," adding that "the constant bashing of teachers and administrators is reaching a boiling point." A third principal noted that "all that matters [in this district] are the test results. Everything is about numbers." Another noted that "teachers with high professional standards are not valued. There is not a sense of teamwork." Several principals cited the overwhelming emphasis on students' test scores rather than teaching and learning as an exceptionally negative consequence of the current evaluation system.

As per the principals' responses to open-ended questions, their own morale has also deteriorated in the wake of system implementation. One principal explained that "we are tired of being told how 'bad' and 'ineffective' we are as administrators and teachers. You can only tell someone this so long before it becomes a reality. [It is a] self-fulfilling prophecy!" Although a few principals cited their important role in shaping school culture and morale and noted few problems in these areas, they also acknowledged that "many principals are pressured and scared to come forward" or "have not been so fortunate" at their own schools.

Another principal described the generally negative impact of the district culture on staffing, noting that it "does not assist with recruitment and retention of teachers." Although denying any knowledge of a quota for file review and staff problems, another principal cited the need "to reassure [school staff] because they feel intimidated and almost paranoid by such gossip." Others reported similar "horror stories from colleagues" and described little "sense of [job] security" as well as rapid evaluation system changes with "inadequate communication" as detrimental to staff morale or retention.

Principals' (lack of) autonomy as evaluators

Many principals also described significant district-level pressures perceivably placed on them to ensure or guarantee specific evaluation outcomes. In response to the survey, almost half of principals (*n* = 33/70, 47.1%) reported always or frequently feeling pressured by their superiors to make decisions with which they wholly disagree. More

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

specifically, 42.8 percent (n = 30/70) and 45.7 percent (n = 32/70) of principals, respectively, reported always or frequently being told which teachers to place on a plan or that they must bring more teachers forward for file review. Accordingly, four out of ten principals (n = 28/70, 40.0%) reported that they are always or frequently pressured to give lower scores on teacher observations and evaluations than they think are deserved. Nearly four out of ten principals (n = 27/70, 38.6%) reported having been told what proportion of teachers should be placed on a plan of improvement. Accordingly, 40.0 percent (n = 28/70) indicated that they have been told they must recommend a minimum proportion of teachers for termination or nonrenewal each year. See Tables 2 and 3 for the proportions of responses for each survey question.

Table 2. Principal respondents' perceptions on district-level pressures in evaluating teachers

Statement	Always	Frequently	Occasionally	Never	n
Question 3: I am pressured to make decisions regarding teacher evaluation and retention by my supervisors that I disagree with.	8 (11.4)	25 (35.7)	26 (37.1)	11 (15.7)	70 (100.0)
Question 4: I am told which teachers to place on a plan of improvement.	8 (11.4)	22 (31.4)	27 (38.6)	13 (18.6)	70 (100.0)
Question 5: I am told that I am bringing too few teachers forward for file review.	11 (15.7)	21 (30.0)	21 (30.0)	17 (24.0)	70 (100.0)
Question 6: I am pressured to give lower scores on observations and evaluations than I think are deserved.	8 (11.4)	20 (28.6)	24 (34.3)	18 (25.7)	70 (100.0)

Note: Counts are presented as raw numbers with respective valid proportions of the total participants in parentheses. Valid proportions may not total 100 due to rounding.

Table 3. Principal respondents' perceptions on teacher plans of improvement and termination or nonrenewal

Statement	Yes	No	n
Question 7: I have been told I must have a certain percentage of teachers on plans of improvement.	27 (38.6)	43 (61.4)	70 (100.0)
Question 8: I have been told I must recommend the termination or nonrenewal of a certain percentage of teachers each year.	28 (40.0)	42 (60.0)	70 (100.0)

Note: Counts are presented as raw numbers with respective valid proportions of the total participants in parentheses. Valid proportions may not total 100 due to rounding.

Principals frequently described perceived pressures in their open-ended responses as well. Regarding the need for alignment between teacher observation and VAM scores, one principal explicitly clarified feeling pressure that "teachers' [professional practice] ratings must reflect the EVAAS® scores," even if that means giving teachers a lower observation rating than they feel is deserved. Also describing pressure to terminate teachers or not renew their contract, one principal explained that "there is a general pressure that every principal feels to keep effective teachers and move out ineffective teachers, [but] I don't think there is any plan at the district level to help challenging schools." Another principal noted the impact of external decision-makers: "Personnel making big decisions on teacher [evaluation], [the] hiring process, building maintenance, etc. are not in the schools every day and don't un-

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

derstand school needs." In their open-ended responses, several principals explicitly expressed not only concerns about their own lack of autonomy as evaluators but also their frustration at the perceived lack of district-level support in improving teacher and student outcomes.

Lack of value for principals as professionals

Principal respondents' dissatisfaction with their current employment, as evidenced by their responses to closed-ended survey items about their future career plans, may be due, at least in part, to their perceived lack of professional value to the district, as reported in their open-ended responses. In terms of future career plans, four out of ten principals (n = 29/70, 41.4%) reported that they do not plan to or are unsure whether they will remain employed in the district for the next five years. In total, 77.1 percent (n = 54/70) would leave if they could find another position elsewhere, and 61.4 percent (n = 43/70) are currently making job inquiries in other districts. See Tables 4 and 5 for the proportions of responses for these survey questions.

Table 4. Principal respondents' five-year plan for employment

Statement	Yes	No	Not Sure	n
Question 10: I plan to stay in [name of school district] for five or more years.	41 (58.6)	22 (31.4)	7 (10.0)	70 (100.0)

Note: Counts are presented as raw numbers with respective valid proportions of the total participants in parentheses. Valid proportions may not total 100 due to rounding.

Table 5. Principal respondents' plans to leave the school district

Statement	Yes	No	Not Applicable	n
Question 11: I would leave [name of school district] if I could get another position.	54 (77.1)	13 (18.6)	3 (4.3)	70 (100.0)
Question 12: I am currently making job inquiries outside of [name of school district].	43 (61.4)	27 (38.6)	0 (0.0)	70 (100.0)

Note: Counts are presented as raw numbers with respective valid proportions of the total participants in parentheses. Valid proportions may not total 100 due to rounding.

In their open-ended responses, principals frequently described feeling undervalued as professionals. For example, one principal explained that "we do not get support of any kind but are expected to perform at high levels. We are constantly threatened and told we are replaceable." Another principal summarized these concerns, namely writing that the district:

> prompts a negative environment with all employees—administrators and teachers [and] promotes an unsustainable model of change that is only sustainable by employing a modernized version of the 1900s factory model of education. Administrators are cheap and unimportant raw materials that are chewed up and spit out without any care or concern for their personal well-being. [This district] has lost countless dedicated employees who admirably and effectively

served impoverished communities. Talent leaves when it is not val-
ued, but more importantly we have lost all [our] sense of humanity
in the name of ambiguous measures of student gains [as per
EVAAS®] that no one truly understands.

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

This principal added that "talent leaves but millions are spent recruiting the next
batch of unsuspecting recruits to replace administrators who have left." Based on
open-ended responses, the rapid pace of implementation and resulting instability of
the district's evaluation system, as explicated in this section, as well as other concerns
may provide context for respondents' dissatisfaction with their current employment.

Conclusions and implications

Despite the adoption of teacher evaluation systems that rely, at least in part, on com-
plex and controversial student achievement measures (e.g., VAMs), relatively little
attention has been given to stakeholders' perceptions of these systems in practice,
especially principals' perceptions. Principals in this study strongly expressed con-
cerns regarding: a) the negative impact of the teacher evaluation system on teacher
and principal morale; b) their lack of autonomy in evaluating teachers and making
staffing decisions; and c) their sense of their own limited professional value in the
district. While principals certainly eluded to or, in some instances, explicitly stated
that broader issues of control and accountability (and the resulting blame for per-
ceived failure) are at issue in this district, study findings suggest that issues surround-
ing perceived declining teacher and principal morale, diminishing principal
autonomy, and decreasing professional value are exacerbated by the evaluation sys-
tem, due to its design and/or implementation.

As noted prior, results likely do not generalize from this particular sample to the
population of principals throughout this district because those who responded might
have been the most distraught. These findings, however, are still key to understand-
ing principals' specific (not general) perceptions as subjects of, and actors within,
this particular teacher evaluation system. Regardless, principal respondents' overall
dissatisfaction with the district, reportedly augmented by the adverse effects of the
teacher evaluation system's implementation, is still troubling.

Given the passage of ESSA and the newfound flexibility afforded to states that
are no longer required to use student achievement data to evaluate teachers in high-
stakes ways, state and local policymakers and district leaders should carefully listen
to the voices of stakeholders who are charged with implementing policy-driven sys-
tems that directly and profoundly affect teachers and students in practice. Whether
through new legislation or in response to litigation, state and local policymakers
who are now considering or reconsidering whether or to what extent to use student
growth data, such as that derived from VAMs, to evaluate teachers should also
thoughtfully consider school leaders' concerns regarding the intended and unin-
tended consequences of evaluating teachers in high-stakes ways (e.g., the impact of
recruiting, hiring, promoting, developing, and retaining teachers based on contro-
versial measures of effectiveness). District leaders, regardless of whether their district
currently evaluates teachers using EVAAS® or another measure of student growth,
might seek to better understand the perspectives of their own stakeholders, again

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

giving particular consideration to principals, with regards to implementing evaluation systems in context. They might deliberately design or redesign evaluation systems that are found wanting. Ultimately, and most importantly, perceived declining morale, diminishing autonomy, and lack of professional value on the part of principals adversely impacts the students in the classroom.

Accordingly, additional research is needed if state and local policymakers, district leaders, and the general public are to better understand the implications of implementing new accountability systems in schools. The lived realities of practitioners matter, particularly given the high-stakes consequences that are often associated with poor evaluation outcomes, which in current policy contexts are arguably driven in large part by the M&P Theory of Change. Unfortunately, the consequences described by principal participants here have already dramatically impacted them as local administrators, and they have also impacted teachers and students. As such, it becomes even more imperative that the voices of the principals in this study inform impending policy changes in ways that help teachers and benefit students.

References

Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System. *Educational Researcher, 37*(2), 65–75. doi: 10.3102/0013189X083 16420

Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability.* New York, NY: Routledge.

Amrein-Beardsley, A., & 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*. Retrieved from http://epaa.asu.edu/ojs/article/view/1096 [July 24, 2017].

Baker, B.D., Oluwole, J.O., & Green, P.C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era. *Education Policy Analysis Archives, 21*(5). Retrieved from http://epaa.asu.edu /ojs/article/view/1298 [January 24, 2015].

Berliner, D.C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record, 116*, 1–31. Retrieved from http://www.tcrecord.org/Content.asp?Content Id=17293 [January 24, 2015].

Braun, H.I. (2005). *Using student progress to evaluate teachers: A primer on value-added models.* Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media /Research/pdf/PICVAM.pdf [January 24, 2015].

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* Thousand Oaks, CA: Sage.

Corcoran, S.P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice.* Providence, RI: Annenberg Institute for School Reform. Retrieved from http://www.annenberginstitute .org/publications/can-teachers-be-evaluated-their-students%E2%80%99-test-scores -should-they-be-use-value-added-me [July 24, 2017].

Creswell, J.W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

David, J.L., & Cuban, L. (2010). *Cutting through the hype: The essential guide to school reform* (Rev. ed.). Cambridge, MA: Harvard Education Press.

Education Commission of the States. (1983). Action for excellence. A comprehensive plan to improve our nation's schools. *Report of the Task Force on Education for Economic Growth.* Denver, CO: Author.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 119–161). New York, NY: Macmillan.

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

Every Student Succeeds Act (ESSA) of 2015, Public Law 114–195, § 129 Stat. 1802. (2016).

Felton, E. (2016, March 11). Southern lawmakers reconsidering role of test scores in teacher evaluations [Web log post]. Retrieved from http://blogs.edweek.org/edweek/teacher beat/2016/03/reconsidering_test_scores_in_teacher_evaluations.html?cmp=eml-enl -eu-news3 [April 16, 2016].

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago, IL: Aldine.

Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction.* White Plains, NY: Longman.

Harris, D.N. (2011). *Value-added measures in education: What every educator needs to know.* Cambridge, MA: Harvard Education Press.

Harris, D.N., & Herrington, C.D. (2015). Editors' introduction: The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher, 44*(2), 71–76. doi: 10.3102/0013189X15576142

Hazi, H.M. (2017). VAM under scrutiny: Teacher evaluation litigation in the states. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 90*(5-6), 184–190. doi: 10.1080/00098655.2017.1366803

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

Hopkins, P. (2016). Teacher voice: How teachers perceive evaluations and how leaders can use this knowledge to help teachers grow professionally. *NASSP Bulletin, 100*(1), 5–25. doi: 10.1177/0192636516670771

Kraft, M.A., & Gilmour, A.F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly, 52*(5), 711–753. doi: 10.1177/0013161X16653445

Lavigne, A.L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record, 116*, 1–29. Retrieved from http://www.tcrecord.org/library/abstract.asp?contentid=17294 [January 24, 2015].

Lavigne, A.L., & Good, T.L. (2014). *Teacher and student evaluation: Moving beyond the failure of school reform.* New York, NY: Routledge.

Layder, D. (1998). *Sociological practice: Linking theory and research.* London, UK: Sage.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* London, UK: Sage.

Loewus, L. (2017). Are states changing course on teacher evaluation? Test-score growth plays lesser role in six states. *Education Week, 37*(13), 1–17. Retrieved from https://www.edweek .org/ew/articles/2017/11/15/are-states-changing-course-on-teacher-evaluation.html [February 18, 2018].

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: United States Government Printing Office.

Newton, X.A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives, 18*(23). Retrieved from https://epaa.asu.edu/ojs/article/view/810 [January 24, 2015].

No Child Left Behind (NCLB) Act of 2001. Public Law 107–110, § 115 Stat. 1425. (2002).

Papay, J.P. (2010). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal, 48*(1), 163–193. doi: 10.3102/00002831210362589

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). London, UK: Sage.

Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *NASSP Bulletin, 95*(2), 122–140. doi: 10.1177/0192636511410052

Smith, M.L. (1997). Mixing and matching: Methods and models. *New Directions for Evaluation, 74*, 73–85. doi: 10.1002/ev.1073

Smith, M.L. (2004). *Political spectacle and the fate of American schools.* New York, NY: Routledge.

Stake, R.E., & Trumbull, D. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science, 7*(1 & 2), 1–12.

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

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

United States Department of Education. (2009). *Race to the top program: Executive summary.* Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf [July 6, 2013].

United States Department of Education. (2010). *Teacher incentive fund.* Retrieved from http://www2.ed.gov/programs/teacherincentive/index.html [July 6, 2013].

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our national failure to acknowledge and act of differences in teacher effectiveness* (2nd ed.). Brooklyn, NY: The New Teacher Project (TNTP). Retrieved from http://tntp.org/ideas-and-innovations /view/the-widget-effect [January 23, 2018].

Will, M. (2016). Assessing quality of teaching staff still complex despite ESSA's leeway. *Education Week, 36*(16), 31–32. Retrieved from http://www.edweek.org/ew/articles/2017/01/04 /assessing-quality-of-teaching-staff-still-complex.html?intc=EW-QC17-TOC &_ga=1.138540723.1051944855.1481128421 [April 9, 2017].

Wise, A.E., Darling-Hammond, L., McLaughlin, M.W., & Bernstein, H.T. (1985). Teacher evaluation: A study of effective practices [special issue]. *The Elementary School Journal, 86*(1), 60–121. doi: 10.1086/461437

IJEPL 13(8) 2018

Paufler

Principal Reflections
on a High-Stakes
Teacher Evaluation
System

Appendix
Survey instrument

1. I am able to recruit and keep staff at my school. (Yes/No)

2. [Name of school district removed]'s culture contributes to high morale among my teaching staff. (Yes/No/Not Applicable)

3. I am pressured to make decisions regarding teacher evaluation and retention by my superiors that I disagree with. (Always, Frequently, Occasionally, Never)

4. I am told which teachers to place on a plan of improvement. (Always, Frequently, Occasionally, Never)

5. I am told that I am bringing too few teachers forward for file review. (Always, Frequently, Occasionally, Never)

6. I am pressured to give lower scores on observations and evaluations than I think are deserved. (Always, Frequently, Occasionally, Never)

7. I have been told I must have a certain percentage of teachers on plans of improvement. (Yes/No)

8. I have been told I must recommend the termination or nonrenewal of a certain percentage of teachers each year. (Yes/No)

9. I feel there is a culture of intimidation in [name of school district removed]. (Yes/No)

10. I plan to stay in [name of school district removed] for five or more years. (Yes/No/Not Sure)

11. I would leave [name of school district removed] if I could get another position. (Yes/No/Not Applicable)

12. I am currently making job inquiries outside of [name of school district removed]. (Yes/No)

13. If there is anything else you would like to add, please do so here: (Open Ended)