Educational Evaluation and Policy Analysis
March 2020, Vol. 42, No. 1, pp. 134–160
DOI: 10.3102/0162373719893338

Policy Implementation, Principal Agency, and Strategic Action: Improving Teaching Effectiveness in New York City Middle Schools

Julie Cohen
University of Virginia
Susanna Loeb
Brown University
Luke C. Miller
James H. Wyckoff
University of Virginia

Ten years ago, the reform of teacher evaluation was touted as a mechanism to improve teacher effectiveness. In response, virtually every state redesigned its teacher evaluation system. Recently, a growing narrative suggests these reforms failed and should be abandoned. This response may be overly simplistic. We explore the variability of New York City principals' implementation of policies intended to promote teaching effectiveness. Drawing on survey, interview, and administrative data, we analyze whether principals believe they can use teacher evaluation and tenure policies to improve teaching effectiveness and how such perceptions influence policy implementation. We find that principals with greater perceived agency are more likely to strategically use tenure and evaluation policies. Results have important implications for principal training and policy implementation.

Keywords: *educational policy, leadership, evaluation, principals, mixed-methods, regression analyses*

Researchers have rightly paid attention to the role that teachers play in a variety of student outcomes, from academic achievement (e.g., Chetty et al., 2014) to school engagement (e.g., Liu & Loeb, 2018) to social and emotional skills (e.g., Blazar & Kraft, 2017). Policymakers and educators have explored multiple approaches to human capital reform—from professional development and coaching programs to financial incentives, along with teacher evaluation and rigorous tenure standards—to shift teaching practice and improve student outcomes. Although some of these approaches have been successful in pilots or targeted applications, rarely have they produced sustained success at scale.

Teacher evaluation provides a prominent recent example. Over the last decade, most states have implemented redesigned teacher evaluation, following a confluence of research (e.g., Kane & Staiger, 2012) and substantial federal policy incentives (e.g., Race to the Top, Teacher Incentive Fund, No Child Left Behind Act [NCLB] waivers). Studies show strong positive effects of evaluation policies in some settings, especially when the policies provide regular feedback to teachers (Dee & Wyckoff, 2015;

Papay et al., 2016; Taylor & Tyler, 2012). However, large scale studies of teacher evaluation and performance pay in New York, Tennessee, and Texas (Fryer, 2013; Marsh et al., 2011; Springer et al., 2010, 2012, 2016) show little benefit for students. Moreover, systematic studies of revised teacher evaluation systems demonstrate that in most states nearly all teachers are rated as effective or better (Kraft & Gilmour, 2017). This result mirrors teacher evaluation ratings prior to evaluation reform (Weisberg et al., 2009). Taken together, these results have increasingly led pundits and the popular press to conclude these systems have failed to improve teaching effectiveness and student outcomes when implemented at scale and, given their cost, should be eliminated (Corbin & Strauss, 2015; Dynarski, 2016; Gates & Gates, 2018; Iasevoli, 2018; National Council on Teacher Quality, 2017).

Dismissing policies as ineffective because of inconsistent results may be premature. A rich literature on policy implementation provide evidence that well-designed policies, successful in smaller pilots, often disappoint when implemented at scale (Fixsen et al., 2005). Numerous studies have found that principals do not implement teacher evaluation systems in the ways consistent with the policy's design (Donaldson & Mavrogordato, 2018; Donaldson & Woulfin, 2018; Marsh et al., 2017; Stecher et al., 2018; Youngs, 2007; Youngs & King, 2002). The reasons underlying failed implementation are varied. A policy that fails to achieve its intended outcome because its design is overly complicated is quite different from one that fails because school personnel have insufficient resources to implement it reliably or one where policymakers failed to insure the engagement of school leadership to embrace the approach. Understanding more about the factors that facilitate or hinder successful implementation can inform the design of policies that are more likely to achieve desired outcomes.

Policies intended to improve teaching effectiveness are usually designed by states or districts but implemented by school leaders. However, these school leaders have been insufficiently featured in the teaching effectiveness literature, though this has shifted somewhat in the last decade (Donaldson & Woulfin, 2018; Grissom, 2011; Harris et al., 2010; Marsh et al., 2017). The studies that do foreground the role of principals suggest they are critical actors in policies targeting teacher evaluation and development (Burch & Spillane, 2005; Donaldson, 2013; Kardos et al., 2001; Smylie & Hart, 1999; Youngs, 2007). Taken together, these studies suggest that policies targeting teaching effectiveness are unlikely to realize their objectives unless principals strategically implement policies in service of such goals.

In this article, we assess the variability of New York City (NYC) principals' implementation of state and district policies intended to promote teaching effectiveness with a particular focus on principals' belief in their ability to improve their teacher workforce, which we term principal perceived agency. We survey and interview middle school principals to understand whether principals believe they can use teacher evaluation and teacher tenure review policies to improve the effectiveness of their teachers. Linking our survey and interview data to rich administrative data, we are able to examine how differences in perceived agency influence proximal outcomes intended by the policies, including counseling out of teachers perceived to be ineffective and/or extension of pretenure teachers' probationary periods. Specifically, we focus on three research questions:

Research Question 1 (RQ1): To what extent do principals perceive they have agency to influence the teaching effectiveness in their schools? How does perceived agency vary by the attributes of teachers?

Research Question 2 (RQ2): Does principal perceived agency vary systematically with the attributes of principals and their schools?

Research Question 3 (RQ3): Do principals with different levels of perceived agency use different policy implementation strategies?

We find that principals vary in their belief that they can improve teaching in their schools. Principals with greater perceived agency are more likely to strategically use district policies concerning tenure review and evaluation, with the articulated goal of improving the teacher

workforce. The results of this study highlight the central role of principals in the implementation of policies targeting teachers and foreground the importance of principals' belief in their own abilities to achieve state and district policy goals.

The goal of our analyses is to develop hypotheses about how principal perceived agency drives policy implementation and improvements in teaching effectiveness, which can be rigorously tested in future analyses.

In setting the foundation for that future work, this study makes three important contributions to the growing body of studies that focus on principal beliefs and strategic action in policy implementation (Donaldson & Woulfin, 2018; Marsh et al., 2017; Youngs, 2007). First, we examine these relationships across hundreds of principals working in numerous school contexts, unlike the prior research which has focused on small numbers of school leaders in smaller districts (Donaldson & Mavrogordato, 2018; Sinnema & Robinson, 2007; Youngs, 2007; Youngs & King, 2002; for an exception, see Goldring et al., 2015). Second, given our large sample, we can examine differences in behavior between principals expressing quite different beliefs, instead of trying to extrapolate from small differences in perceived agency across principals. If differences are not evident between principals expressing particularly high and low agency, then they are unlikely to exist across principals with smaller differences. Third, our study is the first to our knowledge that recognizes and assesses principals' differential perceived agency for different populations of teachers, and then analyzes how these differences are associated with variation in policy implementation. Some prior research has focused on principals' beliefs about particular populations—new teachers (Youngs, 2007) or low-performing teachers (Donaldson & Mavrogordato, 2018)—but the literature is lacking work that examines how the same principals perceive and execute their work with different populations of teachers. Overall, the study provides new empirical evidence of how principals make sense of and implement teacher evaluation and tenure policies.

Background and Framework

Principal leadership is associated with a range of positive teacher outcomes, including increased teacher satisfaction (Grissom & Loeb, 2011), lower teacher turnover rates (Boyd et al., 2011; Grissom, 2011), and teachers' commitment to school reform (Yu et al., 2002). Exit surveys of teachers find that the single most important factor in teacher retention is the leadership of principals (Boyd et al., 2011; Johnson et al., 2012; Ladd, 2011). What is less clear is how principals influence a school's teaching force, though research suggests this involves a focus on the composition of the teacher workforce, as well as on the opportunities for capacity building and instructional improvement for teachers in the schools (Cohen-Vogel et al., 2013; Grissom et al., 2013).

Principals and Policy Implementation

Among their many responsibilities, principals are acknowledged as the instructional leaders of their schools. Through their instructional leadership, principals can influence the teaching effectiveness in their building through a focus on the composition of the teacher workforce as well as on the opportunities for teachers' capacity building and instructional improvement (Cohen-Vogel et al., 2013; Grissom et al., 2013). Research has focused on how principals use different strategies to support novice teachers, focusing on induction processes (Kardos et al., 2001; Youngs & King, 2002) and mentoring (Feiman-Nemser, 2001; Johnson & Birkeland, 2003; Youngs, 2007). Principals provide differential scaffolding for novices that is then associated with teachers' perceptions of their work and retention decisions (Youngs, 2007).

Evaluating teachers with classroom observations has taken an increasingly prominent role in principals' work (Goldring et al., 2015). Although many evaluation systems, including NYC's, also incorporate student learning outcomes, we focus here only on the observation and feedback components of evaluation systems for a number of reasons. Whereas measures of student learning (MOSL) outcomes such as value-added metrics can only be computed for the relatively small percentage of teachers in tested grades and subjects, observation and feedback can be used with all teachers in a district. Moreover, we theorized that observational components of teacher evaluation are more subject to principal influence and

our construct of interest, principal perceived agency. Through observations and feedback, principals can recognize teachers' strengths and address weaknesses, monitor and influence teacher development, work to retain strong teachers, and counsel out weaker teachers (Kimball & Milanowski, 2009; Smylie & Hart, 1999).

The promise of teacher evaluation systems has led to their wide adoption, but the implementation and impact of these policies has been uneven (Burch & Spillane, 2005; Coburn, 2016; Donaldson & Woulfin, 2018; Kraft & Gilmour, 2016; Marsh et al., 2017). Three recent studies highlight the variation in policy implementation. Marsh and colleagues (2017) find that schools in New Orleans use the same teacher evaluation system in quite different ways. The implementation at some schools was "reflective," embracing the process of teacher evaluation and enhancing it; others were compliant, while still others were resistant. Donaldson and Woulfin (2018) examine the implementation of Connecticut's policy and highlight how principals varied in their framing of evaluation policies—as tools for either accountability or development—and engage in a range of "discretionary activities" in implementing these policies. They suggest that discretionary activities, in turn, either enhance or mitigate Connecticut's policy's likelihood of achieving its intended goals. Examining teacher evaluation across several school districts and charter management organizations, Stecher and colleagues (2018) found that principals often do not implement evaluation policies as intended, possibly limiting associated improvements in teaching effectiveness or student outcomes. These studies highlight variability in principals' actions in policy implementation, but they provide limited insight into the beliefs that might be associated with such actions and outcomes. These beliefs are the focus of this article.

District personnel and principal preparation programs have important roles to play here. Researchers have noted the scant training principals receive in leveraging evaluations and associated tenure processes to improve the quality of the teacher workforce (Donaldson & Mavrogordato, 2018; Halverson et al., 2004). Unfortunately, we know little about whether or how principals learn to implement district policies and whether programs or in-service supports cultivate the knowledge, skills, and dispositions that facilitate workforce development. Once on the job, principals do report concerns over whether district personnel will support the ways in which they evaluate teachers (Ingle et al., 2011; Van Sciver, 1990) and their decisions to remove low-performing teachers (Donaldson & Mavrogordato, 2018; Donaldson, 2013; Youngs & King, 2002). Training can help standardize implementation, but only if it targets the sources of variation in implementation, including principals' beliefs about different teachers they are charged with assessing and developing.

Teacher observations embedded in evaluation systems depend on the interactions between principals and teachers. Principal beliefs in their abilities to influence teachers, their comfort with providing negative but constructive feedback, and their perceptions of teacher capabilities all feed into how they implement teacher evaluation policies. Many principals struggle with the controversial nature of evaluation systems and use evaluation as a vehicle only for providing teachers with praise (Firestone et al., 2013; Kimball & Milanowski, 2009; Kraft & Gilmour, 2016) instead of targeted, constructive feedback (Donaldson, 2013; Halverson et al., 2004; Halverson & Clifford, 2006). Principals are also likely to differ in their use of teacher evaluation for counseling out ineffective teachers, though little research has focused on principal beliefs around strategic retention decisions (Balu et al., 2010; Donaldson & Mavrogordato, 2018; Grissom & Loeb, 2011; Yariv, 2006).

Principals' Beliefs About Different Teachers

Principals' beliefs about how they can best improve the teaching at their school factors into how they interact with low-performing teachers (Donaldson, 2013; Kardos et al., 2001; Youngs, 2007). Youngs and King (2002) underscore that principals' beliefs about teacher capacity for development and corresponding actions around supporting teacher improvement play a crucial role in school culture and teaching practices. Some principals who encourage less-effective teachers to leave use district evaluation measures when making these personnel decisions (Grissom et al., 2014; Jacob, 2011). Others focus more on improvement and supporting teachers to develop

necessary skills, seeing counseling out or removal as a last resort. Still others frame low-performance as contextual, for example, seeing the teacher as being in the "wrong grade" or teaching a "difficult group" of students (Donaldson & Mavrogordato, 2018).

The available research provides few insights into how principals' beliefs about teachers are associated with different strategies for supporting, and if necessary, exiting them (Donaldson & Mavrogordato, 2018). Donaldson and Woulfin (2018) argue that more research is needed to understand principals' decision-making processes around teacher-focused policies, and they foreground the importance of attending to both principal perceived agency and contextual constraints in analyzing policy implementation. No study to our knowledge has examined whether these kinds of perceptions of teachers vary across school contexts or across populations of teachers. In this article, we focus on this issue: how principals use different strategies to improve distinct populations of teachers at their schools.

The desire to improve teaching effectiveness, coupled with the central role of principals to achieving that goal, underscores the importance of a systematic understanding of how principals approach the implementation of teacher policies. How does perceived agency influence principals' use of information provided through these policies to shape their decisions on teacher professional development and teacher retention? To what extent do principals see evaluation systems as actionable formative assessments for the teachers with whom they work? Understanding principals' perspectives on the reform approach can shed light on the extent to which reforms might be more effective with additional supports for principals or whether the approach has more fundamental flaws.

Perceived Agency and Strategic Policy Implementation

Schools are complex organizations and many factors influence principals' efforts to improve teaching. At the risk of oversimplification, we delineate a conceptual model of the connections among principals, our focal policies (teacher observation and feedback for consequential evaluation purposes—and tenure review), and teaching effectiveness (see Figure 1). We focus on the extent to which principals believe they *can* improve teaching effectiveness, which we term "principal perceived agency" and which we hypothesize is crucial to how they engage with policy. Agency is the capacity to intentionally take the appropriate action in pursuit of achieving a specific goal (Bandura, 2006; Coburn, 2016; Donaldson & Woulfin, 2018); in our study, improving teaching effectiveness through teacher development or shifting the composition of specific segments of the teacher workforce. Although we cannot directly observe this capacity in principals, we can, through our survey, measure principals' perceptions of their agency. Our measure of perceived agency is akin to self-efficacy, or the belief in one's ability to influence various processes and effect change (Bandura, 1982); and, in fact, some may prefer to label our measure as self-efficacy rather than perceived agency. We choose not to use the term self-efficacy because our survey questions are meaningfully different and more policy-specific than many of those commonly used to measure self-efficacy (e.g., Federici & Skaalvik, 2012).

Prior research suggests that principals with lower levels of self-efficacy struggle to strategize about methods for improving their schools (Tschannen-Moran & Gareis, 2004). We theorize that perceived agency concerning teacher tenure review and annual teacher evaluation policies may be associated with more effective implementation. Without a belief that they *can* improve teaching, principals are unlikely to use the policies in systematic ways to either shift the composition of their teacher workforce or promote the development of their existing teachers (Donaldson & Mavrogordato, 2018). Strategic policy actions include the ways in which principals report engaging with these policies, including the frequency of their observations of teachers, the provision of feedback from evaluation and tenure reviews, and their observed strategic retention decisions, including tenure determinations.

A large body of literature suggests that schools' structural and relational features influence teaching and learning (e.g., Bryk & Schneider, 2002) and that principals' own characteristics and the attributes of their school contexts contribute to their perceived agency and to

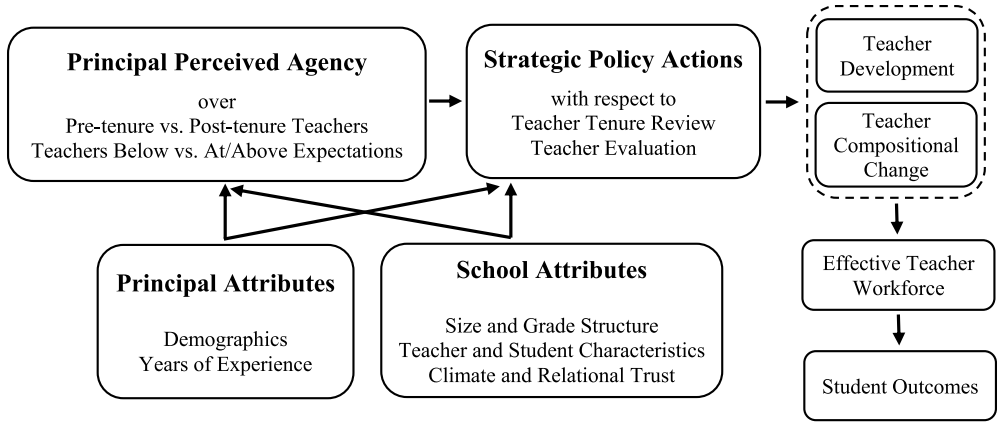

FIGURE 1. *Conceptual framework.*

the strategic policy actions they use (Ladson-Billings, 2009; Pacheco, 2009). A more experienced principal working in a smaller school where teachers regularly collaborate might well feel more agency over improving teaching effectiveness. In contrast, a novice principal working in a large school with a history of animosity between teachers and school leadership may well feel less agency over teachers. Similarly, a principal's own skills and experiences likely influence their differential perceived agency across contexts. In this way, we conceive of perceived agency as an interaction between a principal's belief that he or she can improve the effectiveness of a particular subset of the teacher workforce at his or her school (i.e., pre- or posttenure teachers; teachers below or exceeding performance expectations) with the particulars of his or her school's context that may or may not reinforce those beliefs. Our goal is to better understand the variation in principals' perceived agency, how it corresponds to contextual and individual differences, and how it predicts strategic policy action.

NYC Policies Around Teachers

Situating this research in NYC has several advantages. First, NYC is the largest school district in the country, with more than 1,500 schools. It includes some of the most academically rigorous schools in the country, as well as some of the lowest performing schools. Although NYC is unique in some ways, its diversity provides a rare opportunity to explore principal decision making across a variety of contexts. Second, we are able to link the district's rich administrative data on principals, teachers, and students to a survey of NYC middle school principals and detailed interview data with a subset of principals. The NYC context affords a rare opportunity to connect nuanced reports of principal decision making and strategic actions to an array of administrative variables about principals, teachers, students, and schools.

We ground our exploration of principal perceived agency and strategic action in two district policies that rely heavily on principal discretion and resource management: teacher tenure review and annual teacher evaluation. Beginning in 2009–2010, NYC changed the tenure review process, infusing more information (e.g., information on student progress) and increasing the responsibility and accountability of principals to ensure that teachers met challenging performance standards (NYC Department of Education [NYCDOE], 2009). The district also encouraged principals to recommend more teachers have their probationary period extended an additional year to allow the teachers more time to demonstrate that they met the performance standards appropriate for tenure. In fact, the approval rate decreased from 94% in 2009 to 58% in 2011 (Loeb et al., 2015). Those not receiving tenure typically had their probationary periods extended an additional year (increasing from 4% in 2009 to almost 40% in 2011), and "extended" teachers were much more likely to leave their schools.

Since 2012–2013, principals in NYC schools have used a system called *Advance* to annually evaluate all teachers on a four-category effectiveness rating scale (Highly Effective, Effective, Developing, and Ineffective) based on classroom observations and MOSL. In the 2014–2015 school year, the annual evaluation system shifted to a heavier emphasis on formative, ongoing feedback that teachers could use to improve their practice, rather than the summative measures. Although no research of which we are aware has analyzed variation in the implementation of *Advance*, we theorize that principals' perceived agency is associated with their assessment of the evaluation system's usefulness and the strategies they use to implement it. In particular, we focus on strategic actions around classroom observations and feedback, what *Advance* terms "measures of teaching practice" (MOTP), rather than the "MOSL" which includes measures based on value-added or growth models and goal setting processes. Principals have less discretion around the MOSL than the MOTP, and literature suggests school leaders increasingly emphasize observational measures as a tool for both formative and summative evaluations (e.g., Goldring et al., 2015).

NYC's teacher tenure review process and annual teacher evaluation system aim to infuse more and higher quality information into principals' assessments of teacher performance and associated decision-making processes. These policies also provide mechanisms by which teachers receive guidance on their weaknesses and benchmark their progress addressing those weaknesses. In this study, we aim to understand the variation in principals' beliefs in their abilities to influence teaching effectiveness, and how this variation predicts their differential use of the policies.

Principals' approach to the use of tenure reform and the teacher evaluation system may be informed by their sense of the market for teachers and their ability to recruit replacements for teachers who exit. On average, schools in NYC do not face teacher shortages (Dee & Goldhaber, 2017), but a robust literature documents that some schools have more difficulty recruiting teachers than other schools, and some subject areas are more challenging than others (cf. Boyd et al., 2005; Feng & Sass, 2017; Hanushek et al., 2004). Within a school district, teachers are attracted to schools where they perceive better working conditions, some of which are largely exogenous to principals, for example, the composition of students, but many working conditions are influenced by principals, for example, the working culture. To address this concern, as we explore principal perceived agency, we control for a variety of student attributes that have been associated with challenges in recruiting effective teachers.

Data, Measures, and Methods

Our goal is to understand how principals vary in their perceived agency to improve teaching effectiveness, how perceived agency differs across schools, and whether perceived agency is associated with different approaches to policy implementation. To address these questions, we augment rich administrative data on principals, teachers, students, and schools with two primary data sources: a survey of principals, which focuses on their perception of their agency over teaching effectiveness, and in-depth interviews of a subset of principals that explore these issues in more detail. Taken together, these measures provide different insights into principal perceived agency and policy implementation from self-reports and observed actions, affording a more complete analysis of our research questions.

Data

Principal Survey. Our principal survey had two goals. First, we sought to measure principals' sense of their ability to improve teaching effectiveness in their school through developing teachers and/or compositional change (retaining effective teachers and exiting ineffective teachers; Donaldson & Woulfin, 2018; Youngs, 2007). Second, we wanted to understand principal attitudes toward key teacher policies, and how principals were implementing these. We administered the survey online in the Spring and Summer of 2016 to principals in all NYC schools serving Grades 6, 7, or 8 ($n = 494$). A copy of the survey is found in Supplemental Appendix A (given in the online version of the journal). As an incentive for completing the survey, we gave a US$50 gift

card to each principal's school. A total of 258 completed surveys were returned for a 52% response rate. Table 1 presents characteristics of the middle schools in the survey sample and the full population; only one of these characteristics differs in statistically significant ways.

Principal Interviews. The surveys provide information about both principal perceived agency and the strategies used around policy implementation from a broad and representative group of middle school principals. To provide greater nuance about *how* and *why* principals made particular decisions around policies, we emailed all the principals who completed the survey and invited them to participate in an additional interview, with an incentive of US$100 gift card for their school. Our volunteer interview sample included 40 middle school principals, approximately 16% of the survey sample.

This interview sample was a convenience sample, and the group of principals we interviewed is not wholly representative of either the survey sample or the total population of NYC middle school principals (see Table 1 and Supplemental Table C3 in the online version of the journal). We conducted all interviews over videoconference, and each interview lasted between 1 hour and 2½ hours, depending on the level of detail provided by the participating principals. One of the authors, a postdoctoral fellow, and three doctoral students conducted all interviews using a semi-structured interview protocol focused on understanding why and how principals made decisions regarding policy implementation. At the conclusion of each interview, we member-checked notes with each interviewee to insure our interpretation matched the interviewee's interpretation (Creswell & Miller, 2000). A professional transcription service transcribed all recorded interviews.

Administrative Records. The administrative data files we obtain from the NYCDOE and the New York State Education Department allow us to place principal survey and interview responses in context. First, the NYCDOE employment records allow us to observe the work histories of all principals and teachers. Second, the Tenure Notification System files capture all NYCDOE tenure decisions made between 2008 and 2015. Third, the NYCDOE student demographic and assessment files, available from 1999 to 2016, provide us with information on all students in all NYCDOE schools. Fourth, the teacher–student linkage files allow us to match students to English language arts (ELA) and math teachers between 1999 and 2016. Fifth, NYCDOE's 2015–2016 school climate survey administered to teachers affords us some insight into how teachers view their principal's leadership. Finally, the State's annual School Report Card database and Institution Master Files together with the National Center for Education Statistics' Common Core of Data files provide characteristics of each school.

Measures

Perceived Agency. Our measures of principal perceived agency allow principals' perceptions of their ability to improve teacher effectiveness at their school to vary with characteristics of teachers themselves. We develop four measures that characterize perceived agency along two distinct dimensions: the tenure status of the teachers (pretenure vs. posttenure) and the performance of the teachers (performance below vs. meeting or exceeding expectations). (Each of the four measures relies on Survey Questions 3 and 4 shown in Supplemental Appendix A and the response distribution in Table C1 in Supplemental Appendix C in the online version of the journal.)

To develop hypotheses about the relationships between perceived agency and policy implementation, we examine perceived agency nonparametrically, dividing principals into three groups of perceived agency (low, medium, and high). We exploit the variability in perceived agency to detect relationships which might be lost by only examining linear relationships. We calculated two statistics: (a) the percent of questions with a low-agency response ("Not at All" or "Some") to the relevant questions and (b) the percent of questions with a high-agency response ("A Lot").[1] We label a principal as "low perceived agency" with respect to a specific group of teachers if he or she provided a low-agency response to at least 75% of the relevant perceived agency survey questions. Similarly, we label a principal "high perceived agency" with respect to a specific group of teachers if he or she provided a high-agency response to at least 75% of the relevant questions. The remaining principals are assigned to the medium perceived agency category.

TABLE 1

Characteristics of New York City Middle Schools by Data Source

	All middle schools			Surveys			Interviews		
Characteristic	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
School characteristics									
Serve a grade below sixth grade	494	28.7	45.3	258	28.7	45.3	40	20.0	40.5
Serve a grade below eighth grade	494	17.4	38.0	258	12.4†	33.0	40	22.5†	42.3
% Teachers who applied for a transfer (2013–2014)	483	11.1	14.1	253	10.4	14.2	40	17.6**	25.0
Principal characteristics									
Age	491	46.5	8.9	258	47.8*	8.5	40	47.1	8.4
Hispanic (%)	491	16.7	37.3	258	17.1	37.7	40	10.0	30.4
White (%)	491	50.1	50.0	258	52.3	50.0	40	70.0*	46.4
Black (%)	491	29.3	45.6	258	27.1	44.6	40	15.0	36.2
Female (%)	491	59.9	49.0	258	58.1	49.4	40	60.0	49.6
Years as principal at the school	494	5.4	4.4	258	5.4	4.5	40	6.6	4.05
Principal taught at the school (%)	493	20.3	42.2	258	20.5	40.5	40	17.5	38.4
Teacher characteristics									
Average teacher experience	494	6.2	3.1	258	6.5	3.1	40	5.4*	3.1
% Teachers on probationary status	494	29.7	18.1	258	27.9	17.8	40	35.0*	20.4
% Teachers below the 25th percentile in math[a]	365	25.4	22.9	206	23.3	20.6	25	19.4	13.7
% Teachers below the 25th percentile in ELA[a]	361	23.6	21.6	204	23.2	21.7	25	19.9	19.2
Student characteristics									
% Black	493	32.1	27.6	257	29.0	27.2	40	25.5	24.6
% Hispanic	493	41.9	26.2	257	43.0	26.5	40	50.1	25.7
% Free/reduced-price lunch	493	73.6	19.4	257	74.4	19.0	40	78.0	15.6
School enrollment (100s)	489	5.9	4.0	258	6.2	4.3	40	6.7	4.3
% Students proficient in ELA[a]	492	11.0	5.1	257	10.7	4.7	40	10.8	5.8
% Students proficient in math[a]	492	7.0	4.9	257	7.2	5.0	40	6.9	4.8
% of Students proficient in neither math or ELA[a]	370	52.4	26.5	206	52.4	26.9	26	45.1	23.4

Note. Difference-in-means tests compare survey sample with all middle schools and compare interview sample with the survey sample. ELA = English language arts.

[a]Measured the year before the principal arrived at the school.

$^{\dagger}p < .1$. $*p < .05$. $**p < .01$.

Strategic Actions. We examine six measures of principal strategic actions for the tenure review process and four measures for the *Advance* teacher development and evaluation system. Prior literature on principals' implementation of evaluation and tenure policies informed the selection of measures of strategic actions. For example, numerous studies indicate principals implement evaluation systems in distinct ways, observing teachers more or less frequently (Kraft & Gilmour, 2017; Marsh et al., 2017; Youngs, 2007) and providing distinct types of feedback (Donaldson & Mavrogordato, 2018; Donaldson, 2013). Less work has focused on principals' implementation of tenure policy, but our prior work (Loeb et al., 2015) and the options available to NYC principals informed the selection of corresponding strategic actions around tenure. All these measures are taken from the principal survey with the exception of information on the number of tenure decisions resulting in a teacher's probationary period being extended, which

TABLE 2

Measures of Principal Strategic Actions for Policy Implementation

Strategic action (source)	Values[a]	*M* (*SD*)
Teacher tenure review		
Percent of tenure decisions resulting in the extension of teacher's probationary period since 2010–2011 (administrative data)	0 to 100	34.6 (24.5)
Number of additional observations, above the required three, conducted of a teacher up for an initial tenure decision (Q10)	0, 1, 2, 3, or more	1.3 (1.2)
Number of additional observations, above the required three, conducted of a previously extended teacher up a follow-up tenure decision (Q13)	0, 1, 2, 3, or more	1.4 (1.2)
Number of teacher principal whose probationary period principal extends because the probationary period was insufficient to accurately assess the teacher (Q17e)	None, some, most, all did neither, did one, did both[b]	1.0 (1.0)
Provides additional supports (e.g., mentoring, coaching) to teachers having their probationary period extended and/or counsels these teachers to leave the school (Q21a, Q21b)		1.0 (0.5)
Number of teachers the principal counseled out of his or her school over the last three years (Q23)	0, 1–2, 3–4, 5 or more	1.6 (1.1)
Teacher evaluation		
Frequency of conversations (for at least 5 minutes) with pretenure teachers about their instructional practice (Q8a, Q8b)	Never or a few times a year, once a month, more than once a month[c]	2.2 (0.8)
Frequency of conversations (for at least 5 minutes) with posttenure teachers about their instructional practice (Q8c, Q8d)		2.1 (0.8)
Frequency of conversations (for at least 5 minutes) with teachers who you generally consider to be ineffective or developing about their instructional practice (Q8a, Q8c)		2.3 (0.8)
Frequency of conversations (for at least 5 minutes) with teachers who you generally consider to be effective or highly effective about their instructional practice (Q8b, Q8d)		2.0 (0.9)

[a]See Supplemental Table C2 (given in the online version of the journal) for more information on the distribution of these measures. [b]The "Don't Know" response was recoded as "No." [c]Each of these measures averaged together two items from survey Question 8 and rounded down to create the measured analyzed.

we calculate from administrative data. We list each of these measures (their source, values, and construction) in Table 2 and provide descriptive statistics for them in Table C2 in Supplemental Appendix C (given in the online version of the journal).

Principal Attributes and School Context. In linking the surveys to the administrative data, we create standard measures of the context in which principals work as well as their demographics and professional experience. We observe each principal's gender, race/ethnicity, age, years of experience as the principal at the current school, and whether the principal had previously been a teacher at the school. We characterize each principal's working context with a series of school-, teacher-, and student-level measures. Although all schools serve the sixth, seventh, or eighth grades, some schools also serve grades below sixth and/or grades above eighth. We characterize the teacher workforce with which the principal works with average years of teaching experience at the current school, the percent who are on probationary status (do not have tenure), and two value-added measures of teacher performance (the percent of teachers with an ELA value-added score in the bottom quarter of the district-wide distribution and the same for mathematics value-added score).[2] Finally, we capture the characteristics of students at each principal's school by variables

that measure the total student enrollment, the racial/ethnic composition of the student body, the percentage of students eligible for free/reduced-price lunch eligible, and their performance on the statewide assessments in mathematics and ELA. This standard set of school context measures captures both these observable characteristics of a principal's school but also are proxies for other important unobservable characteristics (such as community resources and preferences and the ease of hiring effective teachers) that may influence principals' perceived agency and their implementation of district policies.

Methods

To answer our three research questions, we use a variety of descriptive analytic techniques. We augment these analyses with insights gained from the principal interviews to further elucidate the constructs presented in our conceptual framework (Figure 1).

Principal Survey. We begin by developing an understanding of how principal perceived agency varies (RQ1) and to what extent contextual factors explain that variation (RQ2). We examine the distribution of the four perceived agency measures and assess the degree to which they are correlated. Drawing on survey data, we estimate a series of ordered logistic regression models to assess how principal and school characteristics are related to perceived agency (RQ2):

$$\mathrm{PA}_i = \beta + \alpha' P_i + \gamma' S_i + \theta' X_i + \lambda' T_i + \varepsilon_i. \qquad (1)$$

Equation 1 predicts the perceived agency of principal i as a function of vectors of principal (P_i), school (S_i), student (X_i), and teacher (T_i) characteristics. We estimate this model separately for each perceived agency measure.

Shifting to how principal perceived agency is correlated with their strategic actions to implement teacher policies (RQ3), we estimate regressions that predict a strategic action of principal i as a function of a perceived agency measure (low and high perceived agency with medium perceived agency principals as the reference), principal characteristics, the school context, and student performance and teacher value-added scores in the year before the principal assumed their position at the school (Equation 2):

$$\mathrm{SA}_i = \beta + \delta_1 \mathrm{LowPA}_i + \delta_2 \mathrm{HighPA}_i + \alpha' P_i + \gamma' S_i + \theta' X_i + \lambda' T_i + \varepsilon_i. \qquad (2)$$

We specify Equation 2 as an ordered logistic regression for those action measures based on survey questions with a discrete response scale and as an ordinary least squares regression for a continuous action measure. As 20% of principals in our sample are their school's founding principal, they are missing values of prior student performance and teacher value-added. We therefore present results from models with and without these performance measures.

In these models, the coefficients of key interest are those for the indicators for low and high perceived agency (δ_1 and δ_2, respectively), which capture differences in strategic actions relative to medium-agency principals. We conduct a Wald test on the equivalence of δ_1 and δ_2 to assess whether low- and high-agency principals differ in their strategic actions.

Principal Interviews. We use interview data to provide insight into how principals use district policies to improve the teacher workforce at their schools. We do not use the interviews to make broad claims about the role of principal perceived agency in policy implementation as the interview sample is not fully representative of the survey sample.

We code the interviews in several stages. During Stage 1, the research team reads all the interviews and generates a list of codes stemming from our conceptual framework and the survey data (Guba & Lincoln, 1994; see Supplemental Table B1 in the online version of the journal for codebook). We create initial definitions and decision rules for each code and compile them in a codebook used by the team throughout the analysis. We revise the codebook in biweekly meetings based on emerging themes and questions. The team of five raters finalizes codes when the raters reach 80% inter-rater agreement on all codes (Miles et al., 2013).

During the second stage of analysis, we code all interviews using Dedoose software. A team member who did not conduct the interview codes

each interview, increasing team-wide exposure to low-inference data. We code interviews at the stanza level, which consist of question–answer exchanges and relevant follow-up questions. Any codes applied to the stanza capture the full exchange between the participant and interviewer (Saldaña, 2013). Codes are not mutually exclusive; a stanza could be coded as a "strategy" along with "teacher characteristic-tenure status." This allows us to create data matrices about strategy by teacher characteristic (e.g., strategies for supporting effective teachers). Fifteen percent of all interviews are double coded with more than 85% agreement across all codes (Miles et al., 2013).

We then engage in an analytic memoing process. Using multiple passes through the coded data by two or more researchers, we create a memo for each principal, systematically analyzing all coded instances across the interview and rereading the interview as a whole (Dyson & Genishi, 2005). We organize memos around our three research questions, paying attention to confirming and disconfirming evidence (Creswell & Miller, 2000).

After completing the coding and memoing processes, we tag each interview with characteristics of the school and principal, culled from the administrative and survey data. Descriptors include principal perceived agency for different groups of teachers generated from the survey data, strategies reported in the survey, school characteristics, and principal characteristics. This allows us to connect interviews to the analysis of the survey responses to provide fuller, more nuanced answers to our three research questions about principal perceived agency to improve teaching effectiveness. Quotes from interviews represent principals identified as high or low perceived agency for a particular group of teachers from the survey data.

Results

Principal Perceived Agency Over Different Groups of Teachers

RQ1: To what extent do principals perceive they have agency to influence the teaching effectiveness in their schools? How does perceived agency vary by the attributes of teachers?

Principals differ in their perceived agency for improving teaching effectiveness: Some feel empowered and capable of shifting the composition and facilitating the development of the teachers; others report feeling less able to affect such change (Figure 2). Although the majority of principals fall into the medium perceived agency group, the distribution of the remaining principals between the low and high perceived agency groups varies across groups of teachers.

Principals indicate greater perceived agency over the improvement of pretenure teachers than over posttenure teachers and over the improvement of teachers who meet or exceed their expectations than over teachers whose performance is below their expectations.[3] As shown in Figure 2, fewer than half as many principals indicate high-agency over posttenure teachers compared with pretenure teachers, and almost 3 times as many principals express high-agency over teachers meeting or exceeding their expectations than teachers not meeting performance expectations.

Interviews corroborate these survey results. Many principals indicate they are better able to support the development of some groups of teachers than other groups. Several principals note that the weaker, posttenure teachers at their schools are impervious to all district efforts at improvement. Principals discuss the relative ease of developing teachers prior to the consequential tenure decision, when they are "impressionable" and "open," and they recount struggling to work with already tenured teachers who they feel they can neither remove nor, in many cases, improve. One principal summarizes the particular benefits of working with pretenure teachers:

> I actually have embraced this idea of hiring first-year teachers. You don't just find veteran, experienced teachers looking for a brand-new job in the South Bronx. I think we've designed the system around very heavily supporting first and second year teachers. Now, as we've done it, we sort of feel like, "Hey, those are actually the people who become our superstar teachers," because they didn't have any bad habits yet or anything else . . . Because they don't know anything yet, they're really open to learning. And if they don't work out, we can tell them after one or two or three years. Once people get tenure, it becomes much more difficult.

For this principal, and many others interviewed, it becomes much more challenging to

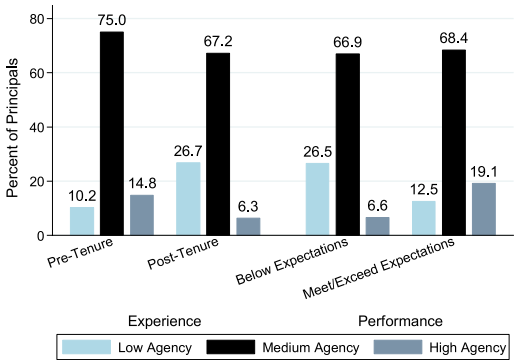

FIGURE 2. *Distribution of principal perceived agency, by perceived agency measure.*
Source. Supplemental Table C3 (given in the online version of the journal).

improve the teaching effectiveness of the posttenure teacher workforce.

A small group of principals surveyed and interviewed express high agency over those not meeting expectations and articulate a clear commitment to fostering ongoing improvement of posttenure teachers. In describing how they conceptualize their role with respect to teachers, these principals discuss the need to make tenure a meaningful milestone, but also to support the development of more experienced teachers. One principal articulates the need to support ongoing growth for posttenure teachers:

> Our veteran and also our effective teachers, our strong teachers, appreciated having feedback more than anybody else in the building. 'Cuz generally they get left out like, "Oh, you're not on my priority list." Then they're the ones that are just so ready to develop. I think I read a study once about people leaving the profession, that one of the number one reasons why they left is that they felt that they were in isolation, and they weren't challenged anymore. I could see that, "Okay, you've reached the threshold. Now we're not worried about developing you anymore."

Several principals articulated the refrain that even experienced and skilled teachers need support and actionable feedback. One describes teaching as: "a journey not a destination. 'Cause the bottom line, this doesn't stop when you get tenure. The expectation is you have to maintain that and grow." Another principal requires posttenure teachers to serve as new teacher mentors or "model teachers" to create a sense "that there's always a ladder within our building, where good people can get better and be great."

Our analyses make clear that principals' perception of agency vary based on the tenure status and performance of the teacher. On average, principals express less agency over teachers they perceive to be weaker, or not meeting their expectations, who are also those most likely in need of support from school leadership. Principals also express a greater sense of agency over pretenure teachers. Given that the vast majority of teachers are posttenure (75%), this lower perceived agency for improving tenured teacher may hinder the implementation of policies designed to improve all teachers, regardless of their performance and tenure status.

Principal Perceived Agency, Principal Attributes, and School Context

RQ2: Does principal perceived agency vary systematically with the attributes of principals and their schools?

For each of the four perceived agency measures, we estimate ordered logistic regressions with and without student and teacher performance measured in the year prior to the principal's arrival at the school (Table 3). The results show only one consistent pattern: principals in schools with higher concentrations of Hispanic and Black students report lower perceived agency over pretenure teachers. This pattern is open to multiple interpretations and may say more about the principals than it does about the schools. Having a measure of principal effectiveness would help narrow the possible explanations. Although it is unclear how to interpret these results, we include these contextual variables in subsequent models to allow us to explore our relationships of interest controlling for these potential confounds.

While we also find few systematic relationships between contextual variables and perceived agency across the survey sample, many principals detail in interviews how their school's contextual factors circumscribe their perceived agency, though also not in systematic ways. Several principals point out that their ability to shift the composition of their teacher workforce

TABLE 3

Selected Coefficients From Ordered Logistic Regression Models of Principal Perceived Agency

	Dimension: Experience				Dimension: Performance			
	Pretenure		Posttenure		Below expectations		Meet/exceed expectations	
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Students Black	−0.017†	−0.027*	0.001	−0.007	−0.011	−0.024*	−0.008	−0.012
	(0.009)	(0.012)	(0.009)	(0.011)	(0.009)	(0.011)	(0.009)	(0.012)
% Students Hispanic	−0.026*	−0.036*	0.001	−0.010	−0.005	−0.016	−0.017†	−0.024†
	(0.011)	(0.014)	(0.010)	(0.012)	(0.010)	(0.013)	(0.010)	(0.013)
Enrollment	−0.017	0.004	−0.038	−0.011	0.013	0.034	−0.090*	−0.068
	(0.043)	(0.049)	(0.039)	(0.044)	(0.040)	(0.044)	(0.040)	(0.045)
Performance included		X		X		X		X
Observations	255	202	255	203	256	203	255	202
Pseudo R^2	.057	.072	.017	.025	.021	.036	.049	.079

Note. Standard errors in parentheses. All models also included percentage of students eligible for free/reduced-price lunch, school characteristics (grades served and borough), and principal attributes (age, gender, race/ethnicity, years principal at the school, and whether taught at the school). The performance covariates are measured in the year prior to the principal's arrival at the school and are the percent of students proficient in neither math nor English language arts and the percent of teachers with value-added below the 25th percentile in math and English language arts. No coefficient on any of these covariates was statistically significant.

†$p < .1$. *$p < .05$.

is limited by their perceptions of the teacher labor market, the desirability of the school for students and teachers, and superintendent support. For example, one principal noted,

> there's a teacher shortage, but it's different for me because I'm in one of the most fantastic buildings, and it's not because of me. It's just a really nice location, really nice families, really good scores, really great teachers. Some schools, if they lose a teacher who is average, all they can get back is a subaverage teacher.

In contrast, several low-agency principals discuss lowering expectations for teachers because of what they perceive to be a lack of otherwise qualified applicants to their schools.

Principals describe district superintendents as a key contextual factor contributing to how much agency they feel around compositional change at their school. Some note feeling hamstrung by district regulations, suggesting that making tenure decisions "sometimes feel like a numbers game." Others suggest the superintendent is the one with the power—"ultimately, it is not my decision"—and that they could not go against the superintendent's decision:

> I have to present an argument to the superintendent if I've seen the growth, but the superintendent also recommends on her own. Like there was a teacher I felt that his practice was growing and the superintendent says, "It's not enough for me." I can't go against what the superintendent says.

Others feel more agency because they are "extremely supported" by their superintendent and have "aligned expectations" within the district. Another details,

> Our district is very, very coherent . . . the principals we do walkthroughs with each other in different buildings, and everybody is pretty much doing it a little bit differently, but overall we are moving teacher practice not just as a school, but as an entire district.

Others acknowledge the central role of district superintendents but still feel a sense of control in shaping the teacher workforce in their school. For example, one notes, "each superintendent approaches [this] really differently. Part of it is learning the politics of how they are going to make the decision." With this knowledge, the principal can present a case in such a way that the superintendent's decision is likely to match the

principal's preference.[4] Overall, we find only minor systematic differences in principal perceived agency across principals and schools with different characteristics in the survey data. In interviews, principals did attribute their perceived agency to more nuanced contextual factors, including central office leadership and support. The discrepancy between the survey and interview may result from differences in the interviews surfacing contextual factors that are less readily quantifiable and not necessarily aligned with measured attributes.

Principal Perceived Agency and Strategic Policy Implementation

RQ3: Do principals with different levels of perceived agency use different policy implementation strategies?

To address this issue, we analyze the relationship between principal perceived agency and strategic policy actions around tenure and *Advance* evaluations, using both survey and interview data. We test these relationships with two models, with and without controls for student achievement and teaching effectiveness at the school the year prior to the principal's arrival. Both models control for student, school, and principal attributes. As we will show, the differences between principals with low and medium and between those with medium and high perceived agency are frequently insignificant while the differences between low and high principals are often statistically significant. Given our goal of hypothesis generation, we focus our discussion primarily on the low-versus-high differences but show all differences in the tables.

We focus on perceived agency over pretenure teachers and teachers performing below expectations in our analysis of the tenure review process given the policy's design. As all teachers participate in the *Advance* teacher development and evaluation system each year, we examine all four perceived agency measures (pre- and posttenure teachers, teachers meeting or performing below expectations). We then triangulate patterns in our survey data with those culled from the interviews.

Teacher Tenure Review. Principals who indicate they feel low agency to improve the effectiveness of pretenure teachers make more use of extensions than do high-agency principals, extending roughly 14 percentage points more teachers (Columns 1 and 2, Table 4). This difference is more than half a standard deviation in the use of extensions. Having extended a teacher's probationary period, however, high-agency principals then leverage the extension period in ways more in keeping with the policy design than do low-agency principals. The district encourages principals to use the extension option for teachers who may not currently meet performance expectations but show the potential to do so, when given additional supports. High-agency principals are more likely to provide extended teachers supports and, alternatively, to counsel extended teachers out (Column 4). Principals with high perceived agency over teachers performing below expectations report counseling out significantly more teachers than low-agency principals (Columns 5 and 6). Perceived agency is not significantly related to the other strategic actions for implementing teacher tenure review (see Supplemental Table C6 in the online version of the journal).

For principals to leverage the tenure review process to improve teaching effectiveness, they must be comfortable with the system's expectations for their role in that process. Principals are expected to gather the requisite information to make a tenure recommendation during the typical 3-year probationary period. Low-agency principals, however, are less likely to report having sufficient information to make a tenure decision (Columns 1 and 2) and are less likely to indicate that the 3-year probationary period permits an accurate assessment of teachers (Columns 3 and 4). Overall, there is consistent evidence that low perceived agency principals feel less control than high-agency principals (Columns 5 and 6, Table 5). Across all three measures, low-agency principals are significantly less comfortable with the tenure review process than are medium-agency principals.

The interviews support these findings. Principals who are high perceived agency over pretenure tenure teachers on the survey talk in interviews about being "decisive about teacher quality" and report knowing someone is "not meant to be a teacher" fairly early in their career. All but two of the principals who are high

TABLE 4

Selected Estimated Coefficients From Regression Models of Strategic Actions for Teacher Tenure Review on Perceived Agency Over Pretenure Teachers and Teachers Performing Below Expectations

	Probationary period extension rate (pretenure agency)		Offered extended teachers additional supports and/or counseled them out (pretenure agency)		Number of teachers counseled out (below expectations agency)	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Low agency	0.109†	0.065	−0.645	−0.566	−0.461	−0.583†
	(0.059)	(0.063)	(0.514)	(0.643)	(0.280)	(0.309)
High agency	−0.028	−0.088†	0.537	2.089†	0.991*	0.887
	(0.045)	(0.051)	(0.911)	(1.227)	(0.505)	(0.577)
F test: High versus low agency	†	*		*	**	*
Observations	208	158	118	92	252	199
Pseudo R^2	.155	.249	.089	.197	.050	.062
Performance included		X		X		X

Note. Standard errors in parentheses. All models also included student, school, and principal attributes. See Supplemental Table C2 (given in the online version of the journal) for descriptive statistics and survey question wording for these measures.
$^{\dagger}p < .1$. *$p < .05$. **$p < .01$.

TABLE 5

Selected Estimated Coefficients From Regressions of Principal Views on Their Role in Implementing Tenure Review Process on Perceived Agency Over Pretenure Teachers

	I had the information I needed to make tenure decisions		The current probationary period allowed for an accurate assessment of teachers		I have control over the tenure decision process	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Low agency	−1.849***	−2.379***	−1.044*	−1.379**	−1.122**	−1.070*
	(0.449)	(0.516)	(0.406)	(0.449)	(0.411)	(0.450)
High agency	0.312	0.441	0.252	0.822†	−0.274	−0.035
	(0.368)	(0.440)	(0.354)	(0.433)	(0.355)	(0.428)
F test: High versus low agency	***	***	*	***	†	†
Observations	239	189	238	188	240	190
Pseudo R^2	.081	.122	.042	.067	.046	.051
Performance included		X		X		X

Note. Standard errors in parentheses. All models also included student, school, and principal attributes. See Supplemental Table C5 (given in the online version of the journal) for descriptive statistics and survey question wording for these measures.
$^{\dagger}p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

perceived agency over pretenure teachers note in interviews that they counsel out ineffective teachers well before a tenure decision, making statements such as "the children shouldn't have a third year of this." They report being very direct with teachers, making plain "this is not the career for you."

These principals' sense of agency and comfort with authority is reflected in their discussion of the tenure review process. Unlike the low perceived agency principals who report feeling constrained by the superintendent's decision-making authority around tenure, the principals with high perceived agency over pretenure teachers discuss

their comfort articulating their central role in the tenure process. One principal notes,

> Sometimes principals are afraid to have the real conversation about why you're not giving someone tenure. Make it around these technical things and defer to superintendents like, "The superintendent was in your room and said this,"—In my mind, if you really sit down with the teacher and say, "Here's what's keeping *me* from giving you tenure," and then the person, if you're really willing to invest in them and work with them, they will turn that around. Then at the end of that, they'll be better.

The common theme across these principals' interviews is the need for directness and clarity with teachers about the extension decision, coupled with additional supports. Principals report telling extended teachers, "if you continue performing at this rate, I will never recommend you for tenure," and "if you don't get [tenure] in four years, you're not meant to be a teacher." At the same time, the principals are equally forceful about the need for supports for extended teachers because extra time alone is unlikely to realize improvement. One described, "let's give it one more year, but let's really push for progress . . . Let's figure out the specific things you need to improve and make sure we help you get there." Principals with high perceived agency over pretenure teachers describe using tenure extensions to clearly signal the need for continued improvement, while using the time strategically to target areas for growth.

In contrast, the principals who are low agency over pretenure teachers are more passive about the tenure process and the use of extensions, with less clarity about why they extend teachers and/or what they do to support those who are extended. One goes so far as saying, "it's not totally clear to me how tenure even works," and many focus on the procedural elements of the tenure review, such as collating tenure binders. Several of these principals put the onus on the extended teachers to develop strategies for improvement: "We would allow the teachers to take on professional development in the areas to support their own growth, but they need to identify those"; "They do know if they have any issues, they can e-mail an administrator." When asked how she supports teachers who have been extended, one principal responds, "it's up to the teacher to look for the support. We can just do so much, so I also want to see if the teacher's taking any initiative." The principals with low perceived agency over pretenure teachers describe their role in the tenure process, both before and after extensions, as less directive and less supportive. Collectively, the survey and interview data suggest high-agency principals are better able to leverage the tenure review process as it was designed: to improve teaching effectiveness in their school through both the development of extended teachers and the differential retention of teachers, based on their perceived effectiveness.

Teacher Evaluation System. The centerpiece of NYC's evaluation system is the feedback provided to teachers following observations of their classrooms conducted by principals, assistant principals, and superintendents. Scheduling both the observations and meetings to provide the feedback requires principals to prioritize this work as they have many other leadership responsibilities and limited time.

Our data reveal that principals with high perceived agency strategically allocate their time and resources in the provision of feedback. We measure feedback by the number of conversations principals have with each of the four specific subgroups of teachers (pre- and posttenure teachers, teacher meeting or performing below expectations) about their instructional practice. Principals who express more agency over a given group of teachers have more conversations about instruction with those teachers than do low-agency principals (Table 6). For example, principals with high perceived agency over teachers performing below expectations have more conversations with teachers performing below expectations than principals with low perceived agency over this group (bottom panel, Columns 1 and 2). The pattern is the same for pre- and posttenure teachers (top panel, Columns 1 and 3) and teachers meeting or exceeding expectations (bottom panel, Columns 3 and 4), although the differences are statistically significant at the 10% level. In sum, principals with high perceived agency strategically allocate their time to teachers whom they believe they can influence.

Interview data support the survey findings that perceived agency is associated with different implementation approaches to the teacher evaluation and development system, *Advance*. In interviews, low-agency principals (across teacher sub

TABLE 6

Selected Estimated Coefficients From Regressions of the Number of Conservations With Specific Subgroups of Teachers About Their Instructional Practice on Perceived Agency for That Subgroup

	(1)	(2)	(3)	(4)
	Dimension: Experience			
Variable	Pretenure		Posttenure	
Low agency	−0.557	−0.563	−0.446	−0.609†
	(0.441)	(0.478)	(0.280)	(0.311)
High agency	0.431	0.405	0.519	0.441
	(0.372)	(0.430)	(0.560)	(0.618)
F test: High versus low agency	†		†	
Observations	248	196	253	201
Pseudo R^2	.072	.071	.072	.079
	Dimension: Performance			
Variable	Below expectations		Meet/exceed expectations	
Low agency	−0.733*	−0.831*	−0.214	−0.320
	(0.298)	(0.332)	(0.383)	(0.426)
High agency	1.000	1.305	0.617†	0.655†
	(0.715)	(0.810)	(0.326)	(0.377)
F test: High versus low agency	*	**	†	†
Observations	241	191	253	200
Pseudo R^2	.106	.117	.074	.066
Performance included		X		X

Note. Standard errors in parentheses. All models also included student, school, and principal attributes. See Supplemental Table C2 (given in the online version of the journal) for descriptive statistics and survey question wording for these measures.
†$p < .1$. *$p < .05$. **$p < .01$.

groups) also describe "doing more" observations and feedback, rather than strategically allocating time and resources to provide feedback to teachers in ways that maximize the feedback's impact. This was evident in comments such as "I am in classrooms a lot" and "I do all the *Advance* observations, every single one, so teachers know that I have my finger on the pulse." The low-agency principals do not describe a particular strategy to engaging in observations and feedbacks. The general approach is one of "more is more."

As with the teacher tenure review, principals with low perceived agency across teacher subgroups report being less confident than high-agency principals in their ability to meet the teacher evaluation system's expectation that they provide useful, honest, and concrete feedback to teachers about their classroom performance. We present the results for perceived agency over pretenure teachers in Table 7, although the findings are consistent across the measures of perceived agency over posttenure teachers, teachers meeting expectations, and teachers performing below expectations. Compared with high-agency principals, there are more teachers with whom low-agency principals feel it is challenging to discuss content-specific issues (top panel, Columns 1 and 2), to identify concrete steps to improve the teacher's practice (top panel, Columns 3 and 4), and to provide negative feedback about the teacher's teaching (bottom panel, Columns 1 and 2). Low-agency principals also worry more that providing negative feedback will undermine their relationships with other teachers (bottom panel, Columns 3 and 4). In fact, low-agency principals are less confident in their role in the teacher evaluation system than medium-agency principals.

Interviews reinforce the survey findings that principals with perceived agency over a particular group of teachers use *Advance* observations

TABLE 7

Selected Estimated Coefficients From Regressions of Principal Views of Their Role in Implementing the Teacher Evaluation System on Perceived Agency Over Pretenure Teachers

	(1)	(2)	(3)	(4)
Variable	I find it challenging to talk with the teacher about content-specific issues when the teacher is teaching a subject I did not teach		I find it challenging to identify concrete steps to help the teacher improve his or her practice	
Low agency	0.858*	0.873†	1.184**	1.376**
	(0.428)	(0.481)	(0.452)	(0.514)
High agency	−0.578	−0.550	−0.135	−0.177
	(0.421)	(0.488)	(0.456)	(0.538)
F test: High versus low agency	*	*	*	*
Observations	253	200	253	200
Pseudo R^2	.083	.098	.063	.092
Variable	I find it challenging to give the teacher negative feedback about the teacher's teaching		I worry that providing negative feedback will lead the teacher to undermine my relationship with other teachers	
Low agency	1.267**	1.178*	1.369***	1.542***
	(0.413)	(0.463)	(0.400)	(0.449)
High agency	−0.360	−0.509	−0.305	−0.209
	(0.460)	(0.567)	(0.415)	(0.492)
F test: High versus low agency	**	*	**	**
Observations	252	199	252	199
Pseudo R^2	.071	.084	.081	.103
Performance included		X		X

Note. Standard errors in parentheses. All models also included student, school, and principal attributes. See Supplemental Table C5 (given in the online version of the journal) for descriptive statistics and survey question wording for these measures.
†$p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

more strategically to provide formative feedback to those teachers (pretenure teachers or those not meeting expectations, for example) that they perceive as benefiting from it most. A common theme across interviews with high perceived agency principals is the district mandated observations of teachers are helpful, but not sufficient for realizing improvement. Many principals with high agency for pretenure teachers, for instance, say that the observation requirements for *Advance* are "inadequate" and that the bar for "effective practice" is far too low for those early in their teaching career. That said, these principals are still able to use *Advance* in strategic ways to support their own goals. Most say they observe pretenure teachers far more than required by *Advance*, but note their "typical observations" are often much shorter than the *Advance* requirements. Another principal with high perceived agency over pretenure teachers tells these early career teachers that *Advance* encourages "informal, unannounced observations" (though the policy does not specify this particular approach), and this encourages the need to be "ready every day of the year." One principal with high perceived agency over teachers performing below expectations says he "only need[s] 1–2 minutes to know if a weaker teacher was engaged in effective instruction" and that "pop-ins" are the most efficient way of gathering information about teachers who are not meeting his expectations. Again, principals with high perceived agency over particular groups of teachers express a

decisiveness and strategic use of time in implementation of the *Advance* evaluation and feedback provision.

In contrast to some principals who report maximizing their observation time writ large, many of the principals with high perceived agency over teachers not meeting expectations suggest "being really thoughtful and careful about what [they] need to do and what could be done equally well, maybe even better, by someone else." One principal with high perceived agency over teachers not meeting her expectations notes,

> I just spend the time observing the ones who are really struggling, the ones who are not hitting the bar, who can't get the kids to sit down, who can't plan the engaging, innovative activities. My AP's they can do the teachers who don't need as much.

Another principal with high perceived agency over teachers not meeting his expectations notes that "teachers are the most important investment that we make, so the feedback has to be very, very strategic and actionable, especially for the ones who are not quite there yet." These principals echo that feedback, rather than observations, are the true lever for improvement for teachers they perceive as weaker and needing of support, but that having a mandated system for observation has been a useful tool for their instructional leadership.

"Principal discretion" in policy implementation is common across the interviews with principals with high perceived agency across different groups of teachers with whom they work, echoing recent work by Donaldson and Woulfin (2018) in Connecticut and earlier work on principals' use of evaluation policies (Donaldson, 2013; Goldring et al., 2015; Youngs, 2007). Many principals describe using the policies to advance their own agendas for their schools and suggest that tenure and *Advance* work in tandem with other systems and policies to affect teacher improvement. For instance, *Advance* is described as a tool and framework for informing coaching, new teacher mentoring, and ongoing professional development efforts. One principal with high perceived agency across all four teacher subgroups expresses this most clearly, "you're asking me about the policies, like they are their own separate things, but like *Advance* and all those rubrics are just a tool for helping me get all my teachers better every single day." Instead of implementing *Advance* as a discrete system for assessing teachers at the end of each school year, these principals report using the observation rubrics as ongoing frameworks for high-quality practice and useful tools for promoting more formative conversations about instructional improvement.

The survey and interview data are consistent: Principals with high perceived agency for particular groups of teachers take different strategic policy actions in working with those teachers. It is possible, however, given the self-reported nature of the data, that the differences are all the perception of the principal with no real differences. Although we are unable to test this directly, on the district's 2015–2016 school survey, teachers consistently rate high-agency principals' leadership more favorably than low-agency principals' leadership (Table 8, Columns 2 and 3), although the difference is only statistically significant with respect to teachers performing below expectations. This is suggestive evidence that teachers' perceptions of effective principal leadership is positively associated with principals' own perceptions of their agency to improve teaching effectiveness.

Discussion and Implications

Over the last decade, policymakers, practitioners, and researchers have embraced a variety of reforms intended to improve teaching effectiveness. Without exception, these reforms, while demonstrating pockets and periods of success, have failed to realize their goals at scale. This lack of success is typically identified as a failure of policy design, with associated recommendations to abandon the policy approach (e.g., Stecher et al., 2018). A rich literature in policy implementation (Fixsen et al., 2005) and recent research on implementation of teacher evaluation specifically (Donaldson & Woulfin, 2018; Marsh et al., 2017; Stecher et al., 2018) suggests this diagnosis may be wrong. The policy itself may be effective, if it is well-resourced and embraced by practitioners (see, for example, Dee & Wyckoff, 2015). This article explores this proposition with a focus on the role that principals play in two prominent policies intended to improve teaching effectiveness.

TABLE 8

Average School-Aggregate Teacher Ratings of Principal Leadership Effectiveness by Principal Perceived Agency

	Dimension: Experience		Dimension: Performance	
	Pretenure	Posttenure	Below expectations	Meet/exceed expectations
Perceived agency group	(1)	(2)	(3)	(4)
Low agency	3.043 (0.511)	3.092 (0.413)	3.103 (0.420)	3.053 (0.449)
Medium agency	3.155 (0.391)	3.162 (0.397)	3.155 (0.403)	3.156 (0.390)
High agency	3.201 (0.407)	3.285 (0.476)	3.330 (0.376)	3.204 (0.442)
T test: High versus low agency		†	*	
Observations	256	256	257	256

Note. Standard errors in parentheses. Teachers responded on a 4-point scale (*strongly disagree* to *strongly agree*) to the following 17 statements: I feel respected by the principal at this school; the principal at this school is an effective manager who makes the school run smoothly; the principal has confidence in the expertise of the teachers at this school; I trust the principal/school leader at his or her word (to do what he or she says that he or she will do); at this school, it is ok to discuss feelings, worries, and frustration with the principal; the principal takes a personal interest in the professional development of teachers; the principal looks out for the personal welfare of the staff members; the principal places the needs of children ahead of personal interests; the principal and assistant principal function as a cohesive unit; the principal/school leader at this school makes clear to the staff his or her expectations for meeting instructional goals; the principal/school leader at this school communicates a clear vision for this school; the principal/school leader at this school understands how children learn; the principal/school leader at this school sets high standards for student learning; the principal/school leader at this school sets clear expectations for teachers about implementing what they have learned in professional development; the principal/school leader at this school carefully tracks student academic progress; the principal/school leader at this school knows what's going on in my classroom; and the principal/school leader at this school participates in instructional planning with teams of teachers.
$^{\dagger}p < .1$. $^{*}p < .05$.

We hypothesize that unless principals believe they can improve specific aspects of teaching effectiveness in their schools, they are unlikely to engage in strategic actions around policy implementation (Donaldson, 2013; Marsh et al., 2017; Youngs, 2007; Youngs & King, 2002). We find that principals express differential agency over different groups of teachers. Many do not believe that they can improve teaching effectiveness or exit ineffective teachers. As a result, they are unlikely to embrace policies with these aims. Others do believe they have agency over ineffective teachers. In general, principals felt less agency over improving posttenure teachers. However, even in this case, a group of principals we surveyed and interviewed perceived a sense of agency over posttenure teachers.

We find that perceived agency is not systematically associated with readily measurable characteristics of principals or the schools in which they work, which is consistent with prior research (Tschannen-Moran & Gareis, 2007). This does not imply that principal beliefs are fixed or unrelated to school or district level factors. Indeed, a large literature on student, teacher, and principal beliefs suggest they are malleable and influenced by a range of school and district policies and interventions (e.g., Klassen et al., 2011; Rimm-Kaufman & Sawyer, 2004; Yeager & Walton, 2011). Research by Tschannen-Moran and Gareis's (2007) suggests that a complex network of school-level supports, including teachers, students, parents, and staff can cultivate principal self-efficacy over time. They also find that district level from superintendents and central office staff serve as significant predictors of the development of principal self-efficacy. Our interview data also suggest district personnel, including regional superintendents, can enhance perceived agency over policy implementation, in particular tenure policy. These reports resonate with prior literature and speak to the need for coordinated central office support to create cultures in which principals feel empowered (Donaldson & Mavrogordato, 2018; Donaldson & Woulfin, 2018). In this way, perceived agency is likely a combination of principals' beliefs about themselves and their ability to implement a particular

policy, coupled with beliefs that within a principal's context their behaviors or strategic actions will affect change (e.g., a district superintendent will affirm your determination about a teacher's tenure decision). What is less clear is how to develop these beliefs in principals. If principal perceived agency is indeed something that we want to promote, we need more empirical work to understand what interventions or supports would cultivate principal perceived agency over all the teachers in their school.

Finally, we find that principal perceived agency is associated with principals' actions to im-prove teaching effectiveness (Donaldson & Mavrogordato, 2018; Goldring & Pasternak, 1994; Halverson et al., 2004). High-agency principals engage in activities associated with improvements in teaching effectiveness much more frequently than low-agency principals. Principals with high and low perceived agency also have quantitatively and qualitatively different approaches to policy implementation. High-agency principals report using the policies in service of their goals, getting information quickly, and making decisive personnel decisions (Donaldson & Mavrogordato, 2018; Donaldson, 2013). High-agency principals report that they use their time more efficiently in both the tenure review process and *Advance* evaluation systems. They are more likely to counsel out weaker teachers before the time-consuming tenure review process. They leverage extensions of the tenure probationary period in strategic ways to signal the need for improvement and provide the supports to help realize these improvements. In contrast, across surveys and interviews, principals with lower perceived agency report struggling to gather information quickly, facilitating hard conversations with weaker teachers, and determining clear steps to promote improvement for those teachers, a finding consistent with recent work by Kraft and Gilmour (2016).

As we suggested in our theory of change, we hypothesize that perceived agency is a necessary but insufficient condition for principals to effect improvement in the teaching workforce. We are not suggesting that a principal's belief that they can influence teacher composition and/or development is sufficient for leveraging changes in the teacher workforce, but our analyses demonstrate that principals with stronger perceived agency are more likely to take actions consistent with improving outcomes for teachers. Whether these principals have the skills to successfully manipulate mechanisms to improve teaching effectiveness remains to be seen. A failure to engage in this process resulting from weak perceived agency preempts the success of these mechanisms.

Our analysis has some limitations. First, the analysis has external and internal validity limitations. The analysis reflects the beliefs and behaviors of NYC middle school principals around two teacher policies. The findings may not generalize to other settings or policies. Nor does this analysis have a strong causal interpretation. By including a variety of controls in our regression analysis, we attempt to limit explanations that compete with principal perceived agency as the key driver behind differences in various actions linked to the policies. So, although we rule out some competing explanations, we caution that factors other than principals' perceived agency may account for some of the relationships we find.

Second, our analysis provides limited insight on what contributes to the meaningful differences in perceived agency that we observe across these principals. We find these differences are largely unrelated to observable characteristics of principals or their schools. They are also seemingly unrelated to principals' background experiences in a particular school (having taught in the school they now lead, for example) or participated in a particular principal preparation program (as described in interviews). Understanding the causes of these differences will require a rich data collection to augment administrative data. Such an analysis will have important implications for improving perceived agency among principals.

Research on principals' beliefs suggests they are malleable and represent a complex interaction between an individual principal and the context in which they work (Tschannen-Moran & Gareis, 2007). Principals in NYC rarely move between schools, so we are unable to empirically separate a principal from the context in which they work or observe principals working in multiple contexts. Understanding more about how beliefs may evolve in different work contexts, as well as designing and testing interventions that might shift beliefs, are important directions for future research.

Data on principal knowledge and skills may explain differential perceived agency. Certain kinds of knowledge (about content, instruction, or students) or skills (including pedagogical, communication, and interpersonal skills) may well be correlated with principal perceived agency, or perceived agency might represent a more distinct disposition. Understanding these relationships is outside the scope of the current data, but it is an important direction for future research. The field lacks a robust set of measures of relevant principal knowledge and skills, and we need to know more about how to select and develop principals with the knowledge, skills, and dispositions that support the development of a high-quality teacher workforce.

Finally, our analyses do not examine the effects of perceived agency on outcomes of the policies, for example, changes in teaching effectiveness through compositional change or development of current teachers. This is an important analysis, which is an important next step for our research in NYC. Because we know so little about how principals implement policy, we chose to broadly describe what we viewed as key elements of a theory of change that connect the design of two important policies intended to improve teaching effectiveness to their intended outcomes. We believe the exploratory analysis presented in this article is a necessary first step by documenting important descriptive patterns.

Developing an effective teacher workforce is likely a productive mechanism to improving student outcomes (e.g., Chetty et al., 2014). The results of this study shed light on potential mechanisms for more effective policy design and implementation for teacher improvement. First, our finding that a small proportion of principals feel high agency over specific aspects of teacher effectiveness offers a key reason why teacher effectiveness policies may not realize their intended impact without targeted supports for principals. Unless those charged with implementing policies embrace those policies, it is unlikely the mechanisms necessary for success will function as planned. The evidence that high-agency principals are more comfortable leveraging evaluation data to provide formative feedback is critical, given prior work that suggests formative feedback from evaluation is key in leveraging teachers improvement (Taylor & Tyler, 2012) and student performance gains (Steinberg & Sartain, 2015). Before concluding that teacher evaluation is ineffective and a waste of time and money, we should better understand the reasons for this outcome.

Ideally, principals would view policies as opportunities to affect their strategic goals around teaching effectiveness, rather than mandates with which they must comply. Our findings suggest that perceived agency might be an important contribution to perceptions of policies and subsequent implementation strategies. Principals, those responsible for principal training, and their superintendents once they become principals can use these results for principal development and selection. Several studies, including ours, suggest that principals need support from district personnel to implement evaluation policies in ways that better align with district goals (Donaldson & Mavrogordato, 2018; Halverson et al., 2004). Our data suggest that those who support principals from preparation into the field, including central office leaders, may benefit by cultivating a sense of agency, coupled with knowledge and skills, to facilitate strong and strategic school leadership.

Additional research is necessary to more fully understand how to select and train principals who strategically embrace policies to improve the quality of instruction in their buildings. Descriptive research can provide a sense of whether our findings generalize to other contexts. Ultimately, rigorous causal research is needed to determine whether increasing principal perceived agency supports policy implementation and boosts teaching effectiveness.

Acknowledgments

We thank Vicki Bernstein and Phil Weinstein at the New York City Department of Education for providing the data used in this article and for answering questions about New York City Department of Education policies. We also thank Ernest Logan and Eloise Messineo of the Council of School Supervisors & Administrators for their assistance with the principal survey. We appreciate helpful comments on earlier drafts from participants at the AERA, APPAM, and AEFP research conferences, as well as from Julie Marsh and reviewers at this journal. Anisah Waite, Rebekah Berlin, Katharine Sadowski, Hannah Mathews, and Jillian McGraw provided exceptional research assistance. The opinions expressed are those of the authors as are any errors.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here is supported by grants from the Carnegie Corporation of New York, the Institute of Education Sciences, U.S. Department of Education, through grant R305B140026 to the Rectors and Visitors of the University of Virginia and support from the National Center for the Analysis of Longitudinal Data in Education Research (CALDER).

Notes

1. Although theory clearly dictates that "Not at All" responses reflect low perceived agency and "A lot" responses reflect high perceived agency, the distribution of the principal responses also guided our final assignment of responses to agency category. Very few principals responded "Not at All"; thus, we combined them with the "Some" responses, leaving "A good amount" responses as medium perceived agency. We recoded responses of not applicable to missing.

2. We estimated value-added scores separately by subject and year by regressing student test scores on prior test scores (same and opposite subject), student demographics (gender, race/ethnicity, eligibility for free/reduced-price lunch, whether English spoken at home, English language learner status, disability status, and whether changed schools), lagged student absences, grade fixed effects, and teacher fixed effects. We then impose Empirical Bayes shrinkage and standardize the resulting value-added scores within subject and year.

3. Although principals varied in their perceived agency over different populations of teachers, principals who feel greater agency with one set of teachers tend to feel greater agency over other sets of teachers (Supplemental Table C4 in the online version of the journal). Nearly all principals who feel the inability to improve pretenure teachers also question their ability to improve posttenure teachers (84.6%). Among principals who indicate high perceived agency to improve posttenure teachers, most also believe they can improve pretenure teachers (73.3%), and almost all principals (92.3%) who express low agency over teacher meeting or exceeding their expectations also express low agency over teachers performing below their expectations. And among principals who feel high agency to improve the performance of a teacher not meeting their expectations, most (87.5%) also are confident in their ability to improve the performance of teachers meeting or exceeding their expectations.

4. We explore the superintendent's role in predicting perceived agency within the survey sample by, first, including New York City community district fixed effects to account for the superintendent's role and, second, conducting a test of joint significance of the community district fixed effects. The results do not change and the fixed effects are not jointly significant. Our ability to statistically detect the superintendent's role is limited by small within-district sample size and the lack of variation in perceived agency within some districts. We cannot rule out that a particular superintendent is a contributor to a principal's perceived agency. Understanding these dynamics, however, is outside the scope of our study but is an important direction for future research.

References

Balu, R., Béteille, T., & Loeb, S. (2010). Examining teacher turnover: The role of school leadership. *Politique Americaine*, *15*, 55–79.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, *1*(2), 164–180.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis*, *39*(1), 146–170.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, *48*(2), 303–333.

Boyd, D., Lankford, D., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Association Proceedings*, *95*(2), 166–171.

Bryk, A., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. Russell Sage Foundation.

Burch, P., & Spillane, J. P. (2005). How subjects matter in district office practice: Instructionally relevant policy in urban school district redesign. *Journal of Educational Change*, *6*(1), 51–76.

Chetty, R., Friedman, J., & Rockoff, J. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes into adulthood. *The American Economic Review*, *104*(9), 2633–2679.

Coburn, C. E. (2016). What's policy got to do with it? How the structure-agency debate can illuminate policy implementation. *American Journal of Education*, *122*, 465–475.

Cohen-Vogel, L., Osborne-Lampkin, L., & Houck, E. (2013). New data, old patterns: The role of test scores in student assignment. In D. Anagnostopoulos, S. A. Rutledge, & R. Jacobsen (Eds.), *The infrastructure of accountability: Mapping data use and its consequences* (pp. 129–144). Harvard Education Press.

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research*. SAGE.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice*, *39*(3), 124–130.

Dee, T., & Goldhaber, D. (2017). *Understanding and addressing teacher shortages in the United States*. The Hamilton Project, Brookings Institution.

Dee, T., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, *34*(2), 267–297.

Donaldson, M., & Mavrogordato, M. (2018). Principals and teacher evaluation: The cognitive, relational, and organizational dimensions of working with low-performing teachers. *Journal of Educational Administration*, *56*(6), 586–601.

Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly*, *49*(5), 838–882.

Donaldson, M. L., & Woulfin, S. (2018). From tinkering to going "rogue": How principals use agency when enacting new teacher evaluation systems. *Educational Evaluation and Policy Analysis*, *40*(4), 531–556.

Dynarski, M. (2016). *Teacher observations have been a waste of time and money*. Brookings Institution. https://www.brookings.edu/research/teacher-observations-have-been-a-waste-of-time-and-money/

Dyson, A. H., & Genishi, C. (2005). *On the case*. Teachers College Press.

Federici, R. A., & Skaalvik, E. M. (2012). Principal self-efficacy: Relations with burnout, job satisfaction and motivation to quit. *Social Psychology of Education*, *15*(3), 295–320.

Feiman-Nemser, S. (2001). Helping novices learn to teach: Lessons from an exemplary support teacher. *Journal of Teacher Education*, *52*(1), 17–30.

Feng, L., & Sass, T. (2017). The impact of incentives to recruit and retain teachers in "hard-to-staff" subjects. *Journal of Policy Analysis and Management*, *37*(1), 112–135.

Firestone, W. A., Nordin, T. L., Blitz, C. L., Kirova, D., & Shcherbakov, A. (2013). How teacher evaluation is changing educational leadership: Lessons from a two-year pilot program. In *Proceedings of the University Council of Educational Administration*, Indianapolis, IN.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. The University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Fryer, R. (2013). Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, *31*(1), 373–427.

Gates, G., & Gates, M. (2018). *Annual letter 2018: The toughest questions we get*. https://www.gatesnotes.com/2018-Annual-Letter

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, *44*(2), 96–104.

Goldring, E., & Pasternak, R. (1994). Principals' coordinating strategies and school effectiveness. *School Effectiveness and School Improvement*, *5*, 239–253.

Grissom, J. A. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record*, *113*(11), 2552–2585.

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*(5), 1091–1123.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433–444.

Grissom, J. A., Loeb, S., & Nakashima, N. (2014). Strategic involuntary teacher transfers and teacher performance: Examining equity and efficiency. *Journal of Policy Analysis and Management*, *33*(1), 112–140.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 163–194). SAGE.

Halverson, R. R., & Clifford, M. A. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. *Education Administration Quarterly*, *42*(4), 578–619.

Halverson, R. R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In W. K. Hoy & C. G. Miskel (Eds.), *Educational administration, policy, and reform: Research and measurement* (pp. 153–188). Information Age.

Hanushek, E., Kain, J., & Rivkin, S. (2004). Why public schools lose teachers. *Journal of Human Resources*, *39*, 326–354.

Harris, D., Rutledge, S., Ingle, W., & Thompson, C. (2010). Mix and match: What principals really look for when hiring teachers. *Education Finance and Policy*, *5*(2), 228–246.

Iasevoli, B. (2018, February 15). Teacher-evaluation efforts haven't shown results, say Bill and Melinda Gates. *Education Week*. http://blogs.edweek.org/edweek/teacherbeat/2018/02/teacher_evaluation_efforts_haven%27t_shown_results_bill_melinda_gates.html?cmp=soc-edit-tw

Ingle, K., Rutledge, S., & Bishop, J. (2011). Context matters: Principals' sensemaking of teacher hiring and on-the-job performance. *Journal of Educational Administration*, *49*(5), 579–610.

Jacob, B. A. (2011). Do principals fire the worst teachers? *Educational Evaluation and Policy Analysis*, *33*(4), 403–434.

Johnson, S. M., & Birkeland, S. E. (2003). Pursuing a "sense of success": New teachers explain their career decisions. *American Educational Research Journal*, *40*(3), 581–617.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*(10), 1–39.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching*. Bill & Melinda Gates Foundation.

Kardos, S. M., Johnson, S. M., Peske, H. G., Kauffman, D., & Liu, E. (2001). Counting on colleagues: New teachers encounter the professional cultures of their schools. *Educational Administration Quarterly*, *37*(2), 250–290.

Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, *45*(1), 34–70.

Klassen, R. M., Tze, V. M., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review*, *23*(1), 21–43.

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

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.

Ladd, H. F. (2011). Teacher's perceptions of their working conditions: How predictive of planned and actual teacher movement? *Education Finance and Policy*, *33*(2), 235–261.

Ladson-Billings, G. (2009). Opportunity to teach: Teacher quality in context. In D. H. Gitomer (Ed.), *Measurement issues and assessment for teaching quality* (pp. 206–222). SAGE.

Liu, J., & Loeb, S. (2018). *Engaging teachers: Measuring the impact of teachers on student attendance in secondary school* (working paper). Stanford Center for Education Policy Analysis.

Loeb, S., Miller, L. C., & Wyckoff, J. (2015). Performance screens for school improvement: The case of teacher tenure reform in NYC. *Educational Researcher*, *44*(4), 199–212.

Marsh, J. A., Bush-Mecenas, S., Strunk, K. O., Lincove, J. A., & Huguet, A. (2017). Evaluating teachers in the Big Easy: How organizational context shapes policy responses in New Orleans. *Educational Evaluation and Policy Analysis*, *39*(4), 539–570.

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J., Kalra, N., DiMartino, C., & Peng, X. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report*. RAND Corporation.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis*. SAGE.

National Council on Teacher Quality. (2017). *Running in place: How new teacher evaluations fail to live up to promises* (Report). https://www.nctq.org/publications/Running-in-Place:-How-New-Teacher-Evaluations-Fail-to-Live-Up-to-Promises

New York City Department of Education. (2009). *The tenure toolkit, 2009-10*.

Pacheco, A. (2009). Mapping the terrain of teacher quality. In D. H. Gitomer (Ed.), *Measurement issues and assessment for teaching quality* (pp. 161–178). SAGE.

Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. (2016). *Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data* (No. w21986). National Bureau of Economic Research.

Rimm-Kaufman, S. E., & Sawyer, B. E. (2004). Primary-grade teachers' self-efficacy beliefs, attitudes toward teaching, and discipline and teaching

practice priorities in relation to the" responsive classroom" approach. *The Elementary School Journal*, *104*(4), 321–341.

Saldaña, J. (2013). *The coding manual for qualitative researchers*. SAGE.

Sinnema, C., & Robinson, V. (2007). The leadership of teaching and learning: Implications for teacher evaluation. *Leadership and Policy in Schools*, *6*, 319–343.

Smylie, M. A., & Hart, A. W. (1999). School leadership for teacher learning and change: A human and social capital perspective. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (2nd ed., pp. 421–440). Jossey-Bass.

Springer, M. G., Ballou, D., Hamilton, L., Le, V., Lockwood, J. R., McCaffrey, D., Pepper, M., & Stecher, B. (2010). *Teacher pay for performance, experimental evidence from the project on incentives in teaching*. National Center on Performance Incentives, Vanderbilt University.

Springer, M. G., Pane, J., Le, V., McCaffrey, D., Burns, S., Hamilton, L., & Stecher, B. (2012). Team pay for performance: Experimental evidence from the round rock pilot project on team incentives. *Educational Evaluation and Policy Analysis*, *34*(4), 367–390.

Springer, M. G., Swain, W. A., & Rodriguez, L. A. (2016). Effective teacher retention bonuses: Evidence from Tennessee. *Educational Evaluation and Policy Analysis*, *38*(2), 199–221.

Stecher, B. M., Holtzman, D. J., Garet, M. S., Hamilton, L. S., Engberg, J., Steiner, E. D., Robyn, A., Baird, M. D., Gutierrez, I. A., Peet, E. D., de los Reyes, I. B., Fronberg, K., Weinberger, G., Hunter, G. P., & Chamber, J. (2018). *Improving teaching effectiveness: Final report: The intensive partnerships for effective teaching through 2015-16*. RAND Corporation.

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

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

Tschannen-Moran, M., & Gareis, C. R. (2004). Principals' sense of efficacy: Assessing a promising construct. *Journal of Educational Administration*, *42*(5), 573–585.

Tschannen-Moran, M., & Gareis, C. R. (2007). Cultivating principals' self-efficacy: Supports that matter. *Journal of School Leadership*, *17*(1), 89–114.

Van Sciver, J. (1990). Teacher dismissals. *Phi Delta Kappan*, *72*, 318–319.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project.

Yariv, E. (2006). "Mum effect": Principals' reluctance to submit negative feedback. *Journal of Managerial Psychology*, *21*(6), 533–546.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*(2), 267–301.

Youngs, P. (2007). How elementary principals' beliefs and actions influence new teachers' induction experiences. *Educational Administration Quarterly*, *44*(2), 101–137.

Youngs, P., & King, M. B. (2002). Principal leadership for professional development to build school capacity. *Educational Administration Quarterly*, *38*(5), 643–670.

Yu, H., Leithwood, K., & Jantzi, D. (2002). The effects of transformational leadership on teachers' commitment to change in Hong Kong. *Journal of Educational Administration*, *40*(4), 368–389.

Authors

JULIE COHEN is an assistant professor of curriculum and instruction at the Curry School of Education and Human Development at the University of Virginia. She studies teachers and teaching, with a focus on policies that support the development of effective instructional practices.

SUSANNA LOEB is the director of the Annenberg Institute at Brown University. Her research focuses broadly on education policy and its role in improving educational opportunities for students.

LUKE C. MILLER is a research assistant professor of education policy at the Curry School of Education and Human Development at the University of Virginia. His areas of research include teacher labor markets, state education policy, and rural education.

JAMES H. WYCKOFF is Curry Memorial Professor of Education and Public Policy at the University of Virginia. His research focuses on teacher labor markets, teacher quality, and policies to improve teaching in low-performing schools.

Manuscript received September 18, 2018
First revision received March 5, 2019
Second revision received September 13, 2019
Accepted October 16, 2019