School Organizational Contexts, Teacher Turnover, and Student Achievement: Evidence From Panel Data

Matthew A. Kraft
Brown University
William H. Marinell
Darrick Shen-Wei Yee
Harvard University

We study the relationship between school organizational contexts, teacher turnover, and student achievement in New York City (NYC) middle schools. Using factor analysis, we construct measures of four distinct dimensions of school climate captured on the annual NYC School Survey. We identify credible estimates by isolating variation in organizational contexts within schools over time. We find that improvements in school leadership especially, as well as in academic expectations, teacher relationships, and school safety are all independently associated with corresponding reductions in teacher turnover. Increases in school safety and academic expectations also correspond with student achievement gains. These results are robust to a range of threats to validity suggesting that our findings are consistent with an underlying causal relationship.

MATTHEW A. KRAFT is an assistant professor of education and economics at Brown University, Box 1938, Providence, RI 02912, USA; e-mail: *mkraft@brown.edu*. His research interests include human capital policies in education, the economics of education, and applied quantitative methods for causal inference. His primary work focuses on efforts to improve educator and organizational effectiveness in K–12 urban public schools.

WILLIAM H. MARINELL is the chief of learning and innovation in the Southbridge Public Schools—the third district taken over by the state of Massachusetts. Prior to joining Southbridge, Marinell was the director of analytics at the Center for Education Policy Research at Harvard University and, preceding that, a senior research scientist at New York University's Research Alliance for NYC Schools.

DARRICK SHEN-WEI YEE is an EdD candidate in the Quantitative Policy Analysis in Education Program at the Harvard Graduate School of Education. His research focuses on quantitative evaluation, federal education policy, and the statistical properties of state accountability metrics.

> The whole is more than the sum of its parts.
>
> —Aristotle, *Metaphysica*

Educational reforms over the past decade have increasingly focused on efforts to recruit, select, develop, evaluate, and retain effective teachers. This attention on the individual teacher was catalyzed in large part by research documenting the large magnitude and variability in teacher effects on student achievement (see Hanushek & Rivkin, 2010). However, in teaching, as in any occupation where professionals work within an organization, productivity is shaped by both individual and organizational factors (Hackman & Oldham, 1980; Johnson, 2009; Kanter, 1983; Kennedy, 2010). Organizational contexts in schools are both teachers' working conditions and students' learning environments. Furthermore, organizational factors largely dictate the success of policies designed to increase individual teachers' effectiveness by shaping how these policies are perceived and implemented within schools (Honig, 2006).

Organizational theory and empirical evidence suggest that school contexts affect teachers and students through multiple pathways. Decades of qualitative research have illuminated the important and interrelated features of schools' climates, cultures, and contexts (Johnson, 1990; Lortie, 1975). Schools are dynamic organizations where a "constellation of features" interacts to shape teachers' motivation, job satisfaction, and sense of success (Johnson, 1990; Johnson & Birkeland, 2003). Much of the early empirical research focused specifically on the association between student achievement gains and a single feature of the school environment, such as school leadership style (Leithwood & Jantzi, 2006; Miller & Rowan, 2006) or teacher collaboration (Goddard, Goddard, & Tschannen-Moran, 2007; Lee & Smith, 1996). More recent work has analyzed longitudinal data using structural equation modeling (SEM) to explore direct and indirect pathways between, and reciprocal relationships among, leadership styles, organizational capacity, teacher practices, and student achievement (Dumay, Boonen, & Van Damme, 2013; Heck & Hallinger, 2010; Supovitz, Sirinides, & May, 2010).

Not surprisingly, efforts to measure multiple dimensions of the school organizational context and estimate their causal impact on teachers and students in a common model have been limited by the complexity of these constructs and relationships. In recent years, the proliferation of school surveys administered to teachers, students, and parents have provided new opportunities to quantify a wide range of dimensions of school organizational contexts and examine their relationship with teacher turnover and student

achievement using large-scale data. Studies of school organizational contexts in California, Chicago, Massachusetts, New York City, and North Carolina document consistent evidence of the relationship between school contexts and teacher turnover as well as emerging evidence of the relationship between school contexts and student achievement (for a review, see Simon & Johnson, 2015). However, questions still remain about whether these relationships are in fact causal.

Previous studies have relied almost universally on a single year of school context data. This approach, which leverages variation in context measures across schools to examine differences in student achievement levels or gains, is unable to rule out a range of plausible alternative explanations. Although the University of Chicago Consortium on School Research (CCSR) has collected repeated waves of its 5 Essentials Schools Survey, studies leveraging these data focus primarily on explaining differences in achievement trends across schools with a single cross-section of survey data (Allensworth, Ponisciak, & Mazzeo, 2009; Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). One important exception is Heck and Hallinger's analysis of how changes in distributed and collaborative leadership are associated with changes in school capacity, sociocurricular organization, and academic achievement (Hallinger & Heck, 2010a, 2010b; Heck & Hallinger, 2009). Here, the authors' primary focus is on testing their conceptual model of direct and indirect relationships between constructs using SEM rather than on causal inference. As they explain, SEM can produce misleading results due to omitted and confounding variables (Heck & Hallinger, 2009).

In this study, we provide the first direct evidence to inform answers to two questions central to policy and practice: Would strengthening organizational contexts in schools decrease teacher turnover and increase student achievement? And, which dimensions should we focus on for improvement? We accomplish this by leveraging panel data from the New York City Department of Education's (NYC DOE) School Survey. Starting in 2007, the NYC DOE has administered an annual school survey to teachers, parents, and students—one of the largest survey administration efforts conducted in the United States outside of the decennial population census. We focus our analyses on NYC middle schools because these adolescent years constitute a crucial period in students' academic and social-emotional development (Balfanz, Herzog, & Mac Iver, 2007; Murdock, Anderman, & Hodge, 2000) and because of the acute problems middle schools face with teacher satisfaction and turnover (Marinell & Coca, 2013). We identify distinct, malleable dimensions of NYC middle schools' organizational contexts using teachers' responses to the annual School Survey and estimate the relationship between these measures, teacher turnover, and student achievement.

Our study extends earlier research on school contexts in at least two ways. First, our panel data set allows us to address many of the most important potential threats to the internal validity of previous studies. To date,

studies of school contexts have not accounted for a host of potentially unmeasured between-school differences in student, teacher, and school characteristics that might be correlated with school context measures, teacher turnover, and student achievement. For example, if students who exhibit higher levels of motivation and effort (characteristics that are typically unobserved by researchers) are more likely to attend schools with stronger school contexts, this would induce a spurious positive relationship between school context measures and student performance. Our primary identification strategy resolves this potential threat by removing any time-invariant differences across schools in the quality of their organizational contexts and their teacher and student outcomes. Such variation may be driven by selection bias caused by student and teacher sorting across schools. We focus our analyses on the relationships between changes in schools' organizational contexts, teacher turnover, and student achievement within schools over time. We then demonstrate the robustness of our results to a range of potential threats to a causal interpretation including common source bias, reverse causality, and omitted variable bias.

Second, our analyses inform both theory and practice by directly comparing the relative magnitude of the relationships of multiple school context dimensions with both turnover and student achievement. Previous studies have focused largely on one dimension or one outcome. Items contained on the NYC School Survey captured four distinct dimensions of schools' organizational contexts: leadership and professional development, high academic expectations for students, teacher relationships and collaboration, and school safety and order. Our analyses illustrate that these dimensions matter to teachers and students in different ways. Among the four dimensions captured by the NYC School Survey, improvements in the leadership and professional development factor have the strongest relationship with decreases in teacher turnover, although all four dimensions have independent and statistically significant negative associations with turnover. In contrast, improvements in schools' safety and order and increases in academic expectations for students are the only two significant predictors of corresponding improvements in mathematics achievement, with safety and order as the dominant measure. We conclude by discussing the implications of these findings for theory and practice.

Literature Review

School Contexts and Teachers

Qualitative studies describe in vivid terms how teachers' career decisions are shaped by the contexts in which they work. Johnson and Birkeland's (2003) longitudinal interview study of 50 new teachers revealed that the most important factor influencing these teachers' career decisions

was whether they felt they could be effective with their students. A variety of working conditions in schools shape this success, such as the nature of collegial interactions, the support of administrators, and school-wide approaches to discipline. Drawing on interviews with teachers in high-poverty urban schools, Kraft et al. (2015) found that teachers consistently described the ways in which the quality of instructional support from administrators and approaches to schoolwide discipline affected their ability to deliver high-quality instruction.

A growing body of empirical research now documents the strong positive relationships between supportive school contexts and teacher retention. Analyses of the nationally representative Schools and Staffing Survey and Teacher Follow-Up Survey were among the first to illustrate that organizational factors such as school leadership and student discipline were predictive of teacher retention decisions (Ingersoll, 2001; Shen, 1997). Several studies document the primary role school administrators have in supporting teachers and influencing their decisions to remain at their school (Boyd et al., 2011; Grissom, 2011). In a review of the recent literature on teacher turnover, Simon and Johnson (2015) identified six empirical studies that examined the relationship between dimensions of the school context and teacher turnover (Allensworth et al., 2009; Boyd et al., 2011; Johnson, Kraft, & Papay, 2012; Ladd, 2011; Loeb, Darling-Hammond, & Luczak, 2005; Marinell & Coca, 2013). Together, these studies present compelling evidence that school context measures are stronger predictors of teacher turnover than individual teacher traits or the average characteristics of students in a school. Although each study examined a different set of dimensions, several consistently emerged as the strongest predictors of turnover, including the quality of school leadership, the degree of order and discipline in a school, and the support that collegial relationships provide.

School Contexts and Students

Literature reviews and meta-analyses of studies examining the relationship between school leadership and student outcomes find mixed results but generally conclude that the relationship is primarily or entirely mediated through principals' influence on teachers' practices and the school learning environment (Hallinger & Heck, 1998; Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004; Waters, Marzano, & McNulty, 2003; Witzeirs, Bosker, & Kruger, 2003). Principals play a key role in promoting professional growth among teachers by serving as instructional leaders who provide targeted feedback and facilitate opportunities for teachers to reflect on their practice (Blase & Blase, 1999; May & Supovitz, 2010).

Research also suggests that the quality of teacher relationships and collaboration are related to student achievement gains. Studies find evidence of the positive, albeit weak, association between the frequency of teacher

collaboration and student achievement (Goddard et al., 2007; Lee & Smith, 1996). Jackson and Bruegmann (2009) found that teachers, especially novices, improve their ability to raise standardized tests scores when they work in a school with more effective grade-level colleagues. Most recently, Ronfeldt, Farmer, McQueen, and Grissom (2015) demonstrated that collaboration in instructional teams was predictive of more rapid teacher improvement over time and student achievement gains.

Safe and orderly learning environments appear to be particularly salient for supporting student achievement. The large achievement gaps associated with measures of school safety in Chicago schools illustrate the value of environments where teachers and students are able to concentrate on teaching and learning (Steinberg, Allensworth, & Johnson, 2011). Bryk and his colleagues (2010) showed how schools that make substantial gains in student learning are more than twice as likely to have safe and orderly school climates. Further evidence of the importance of school safety and order comes from compelling econometric analyses of the causal effect of disruptive students on their peers' academic achievement (Carrell & Hoekstra, 2010; Figlio, 2007).

Evidence of the importance of teacher expectations for student achievement date back to Rosenthal and Jacobson's (1968) seminal study of the Pygmalion effect. Many replication studies have since confirmed that experimental manipulations of teachers' beliefs about their students' abilities affect student learning (Raudenbush, 1984). In a series of studies, Hoy and his colleagues demonstrated the predictive power of academic emphasis and optimism, two school-level measures of press for academic achievement, on student achievement (Hoy & Hannum, 1997; Hoy & Sabo, 1998; Hoy, Tarter, Hoy, 2006). Studies examining variation in charter school effects find that a culture of high expectations is likely a key ingredient in the success of high-performing charter schools (Dobbie & Fryer, 2013).

Only two studies we are aware of directly examined the relationships between multiple measures of school contexts and achievement, both using cross-sectional data. Ladd (2009) demonstrated that teachers' perceptions of school leadership and the amount of common planning time predicted a school's value-added in mathematics. Johnson et al. (2012) found that measures of the quality of collegial relationships, shared governance, and school culture were the strongest predictors of schools' median student growth percentile over the following two years, after controlling for a range of student-, teacher-, and school-level characteristics.

Research Design

Data and Sample

We draw from four sources of data provided by the NYC DOE:

Table 1
Middle School Teacher Response Rates to the New York City Department of Education School Survey

	n	Average	25th Percentile	75th Percentile
2008	5,369	63	45	84
2009	6,063	74	60	92
2010	6,574	77	65	92
2011	6,699	82	71	95
2012	6,994	84	74	96

Note. "Average" is the overall response rate for the indicated year. "Percentiles" indicate school-level response rates for schools at the 25th and 75th percentile of the school response rate distribution.

1. Teacher responses to the NYC School Survey, from which we construct our measures of school context;
2. NYC human resources data, which we use as the source for teacher turnover data and teacher-level covariates;
3. NYC student assessment data, which serve as the basis for our achievement outcome data and student-level covariates;
4. NYC school administrative data, which we employ to identify our analysis sample and provide school-level covariates.

The NYC School Survey is administered annually to teachers, students, and parents in the district. The original version of the teacher survey we analyze was developed for the NYC DOE by a contracting firm and intended to capture teachers' opinions across four broad reporting categories: academic expectations, communication, engagement, and safety and respect. Individual teacher responses are anonymized but linked to the teacher's school. Although the district first administered the School Survey in 2006–2007, we excluded responses of this initial administration from our analyses because of the relatively low response rate (46%) among middle school teachers. As shown in Table 1, response rates jumped to 63% in 2007–2008 and rose incrementally each following year, rising to 84% by 2011–2012.[1] We describe our approach to constructing school context measures using these data in the following section.

Human resources data capture demographic information on all personnel employed by the district, as well as job codes, school assignment, salaries, and information on degrees and experience. For our primary turnover analyses, we construct our outcome measure of turnover by coding a teacher as 0 ("stayer") if he or she remained in the same school the following year *and* retained a teaching job code and 1 otherwise. For our supplemental analyses, we distinguish among teachers in this latter group by coding them as "transfers" if they continued teaching but moved to

a different school and "leavers" if they left teaching in NYC public schools altogether. We restrict the data set to include only full-time middle school classroom or special education teachers, using the NYC DOE's approach to identifying active teachers.

Student assessment data include information on students' demographic characteristics and their scale scores on the New York State Testing Program's standardized assessments in mathematics and English language arts (ELA); these scores serve as the basis for our two primary student outcomes. We standardize students' scale scores within grade, year, and subject in order to place these scores on a "pooled" scale and control for secular, statewide trends in score variances and means.

Finally, school administrative records contain data on school type, grade configuration, enrollment, and other school characteristics. We restrict our samples to include only middle schools with traditional Grades 6–8 configurations, excluding schools spanning additional grades such as K–8 or 6–12. In order to avoid conflating turnover with structurally induced employment patterns, we also excluded schools in years when they were new (and still phasing in to full Grades 6–8 enrollment), expanding to include additional grades, or in the process of phasing out grades toward closure.

We produce two analytic data sets from these sources: one for teacher turnover and one for student achievement. Both data sets center on a panel of 278 unique middle schools across the five school years from 2007–2008 through 2011–2012 and include school-level teacher and student statistics, such as log-enrollment, free/reduced-price lunch percentage, race/ethnicity and gender percentages for both teachers and students, and other demographic data. The teacher turnover data set comprises 53,991 records associated with 16,404 unique teachers and includes teacher-level turnover outcomes and characteristics. The student achievement data set contains more than 600,000 student test records in each subject (math and ELA) from approximately 334,000 unique students as well as standard student-level demographic characteristics.

We present summary statistics for the turnover and achievement data sets in Tables 2 and 3, respectively. In the turnover data, records are most commonly associated with teachers who are female (69.6%), White (57.9%), and with 4 to 10 years of experience (56.8%). Across all teacher-years, about 6.2% of teachers transferred schools but retained teaching jobs the following year, while an additional 8.9% were no longer teaching in NYC public schools the following year, for a total turnover rate of 15.1%; these rates are generally consistent with prior studies of NYC schools (Marinell & Coca, 2013) and studies of other large, urban school districts (Papay et al., 2015). In the achievement data, nearly 40% of records are associated with Hispanic students, with African American (26.9%), Asian (17.7%), and White students (15.3%) making up the bulk of the remainder. A

	Mean
Female	.696
Race/ethnicity	
Asian	.053
Hispanic	.224
African American	.129
White	.579
Teaching experience	
Novice	.053
2–3 years of experience	.119
4–10 years of experience	.568
11–20 years of experience	.128
>20 years of experience	.131
Highest level of education	
Bachelor's degree	.149
Master's degree	.428
Master's degree plus 30 credits	.420
Turnover status	
Transfer	.062
Leave	.089
Total turnover	.151

Note. Sample includes 53,991 teacher-year observations and 16,404 unique teachers. Total turnover is the sum of movers (teachers who remain active classroom teachers but transfer to another New York City Department of Education school in the subsequent year) and leavers (i.e., teachers who are no longer teaching in NYC public schools.)

Table 3
New York City Middle School Student Characteristics, 2008–2012

	Mean
Female	.498
White	.153
African American	.269
Asian	.177
Hispanic	.394
Special education	.121
Free/reduced-price lunch eligible	.645
English language learner	.075
Mathematics achievement	.029
English language arts achievement	−.004

Note. Sample includes 645,995 student-year observations and 334,050 unique students. Mathematics and English language arts achievement are standardized *z*-scores.

substantial majority of students (64.5%) are classified as eligible for free or reduced-price lunch.

Constructing Measures of School Context

The NYC School Survey data set we employ for our analyses spans five school years (2007–2008 through 2011–2012) and contains more than 31,000 teacher survey responses. Each survey response consists of more than 40 item responses coded on a 4-point Likert scale. We use these responses to construct a small number of distinct school context measures in an effort to improve interpretability, reduce the likelihood of Type I error, and mitigate problems associated with item multicollinearity.

Prior studies of the relationship between school contexts, turnover, and student achievement take two primary approaches to constructing school context measures. Researchers have typically taken either a theory- or data-driven approach to identifying multiple dimensions of school context. Four of the studies reviewed by Simon and Johnson (2015) created measures by averaging teachers' responses to survey items that were intended to capture conceptually distinct dimensions of the school context identified by theory and prior research. While this approach is grounded in a strong theoretical framework and has intuitive appeal, in practice, these multiple measures often capture a large degree of common variance, limiting researchers' ability to isolate the independent relationship of any specific dimension from others. Alternatively, Loeb and her coauthors (2005) and Ladd (2011) constructed unique dimensions of the school context based on factor analysis methods that minimize the shared variance across factors. Such an approach allowed them to fit models with multiple measures of the school context that do not suffer from multicollinearity. However, this data-driven approach can come at the cost of reduced conceptual clarity around exactly what each factor is measuring.

We drew on both theory-driven and data-driven approaches to inform our construction of school context measures. First, we screened out items that were not common across survey versions from 2008 to 2012. Second, we removed items about school context features that we judged as primarily capturing factors largely outside of the control of school staff. This process allowed us to focus our analyses on a set of 33 items measuring school characteristics that, in principle, school leaders can directly influence. However, we recognize that this decision may have resulted in our omitting other dimensions, such as community support, that may be related to teachers' career decisions and students' achievement.

Prior research on the NYC DOE School Survey found that a theorized factor structure based on the survey's four broad reporting categories did not fit the data well (Nathanson et al., 2013). Given this, we did not endorse the survey's nominal reporting categories and instead included all 33 items in

a two-step process intended to reduce the dimensionality of the survey data, following Loeb et al. (2005), Ladd (2011), and Kane, Taylor, Tyler, and Wooten (2011). First, we conducted a principal components analysis on our panel data set of teachers' responses to identify the number of meaningful, independent dimensions of school context captured by our 33 items.[2] Second, we applied an orthogonal (varimax) rotation to the resulting principal component loadings in order to maximize the "distinctiveness" of each measure and improve interpretability. Our final school context measures were produced using these rotated loadings.

Our analysis of the principal components led us to retain four orthogonal dimensions of variation from the school survey item responses. Similar to prior analyses of teacher working conditions surveys (e.g., Kraft & Papay, 2014), almost every item was equally weighted on the first principal component, which explained half of the total variance. Visual examination of the scree plot suggested that the items captured several other potential dimensions but without a clear "breaking point" between components after the first. We therefore chose to apply the Kaiser-Guttman stopping criterion, retaining the four principal components that had eigenvalues greater than one. Together, these four principal components explained 64% of the total item variance.

Following past research, we constructed measures of the four dimensions suggested by our exploratory principal components analysis by applying a varimax rotation to the principal component loadings (Kane et al., 2011; Ladd, 2011). The varimax rotation helped to make the pattern of loadings more pronounced by maximizing the variance of loadings for each factor, producing a simpler structure and factors that may be easier to interpret while preserving the pairwise orthogonality and total variance of the original components. Our choice of an orthogonal rotation allowed us to construct measures of distinct dimensions of the school context that are uncorrelated at the teacher level.[3]

Our context measures capture four broad organizational features of schools, which we describe as leadership and professional development (Leadership), high academic expectations for students (Expectations), teacher relationships and collaboration (Relationships), and school safety and order (Safety). We arrived at these labels by characterizing the dominant items on each factor and provide a complete list of the items and factor loadings in Appendix Table A1.

Items asking teachers about the quality of school leadership, professional development opportunities, and feedback in a school loaded strongly onto the Leadership factor. As shown in the Appendix Table A1, the Leadership dimension has the most items with large factor loadings (i.e., exceeding .50) and explains 21% of the variance across our 33 teacher item responses. Items with the strongest loadings fell into two broad categories: (a) items that inquired directly about attributes of principals' leadership

and (b) items that focused on aspects of schools' professional development. We view these latter items as capturing, in part, the ability of a principal or school leadership team to identify and provide high-quality professional development opportunities. Specifically, items that loaded strongly onto this factor inquired about whether principals communicate a clear vision, encourage open communication on important school issues, are effective managers, support their teachers, provide time for collaboration, and provide feedback on teachers' instruction. The items pertaining to professional development asked about the efficacy of professional development opportunities generally and whether teachers received training in their content areas and in using data to inform their instruction.

Our second factor explains 18% of the total survey response variance and is dominated by a block of items that fall into a thematic category capturing the rigor of academic expectations for students in a school. The items with the largest loadings on this factor inquired about the extent to which schools set high expectations for all students, set high standards for student work, have clear measures of progress for student achievement, help students develop challenging learning goals, and support students in achieving these goals.

The third factor, Relationships, primarily reflects items that capture the nature of teacher relationships and collaboration in a school and explains 14% of the total variance. The items with the largest loadings on this factor inquired about the extent to which teachers feel supported by their colleagues, work together to improve their instructional practice, trust one another, respect peers who take on leadership roles, and respect colleagues who are the most effective teachers.

Finally, we identify a fourth dimension, explaining 11% of total variance, that associates strongly with items related to student behavior and the level of school safety. The items that contribute the strongest loadings on this factor inquire about whether the school is characterized by crime and violence or students being threatened or bullied, whether order and discipline are maintained, whether adults within the school are disrespectful to students, and whether teachers feel safe at their school and can get the help they need to address student misbehavior.

We calculated factor scores for each teacher in each year and then averaged these scores to the school-year level to obtain our primary predictors.[4] We then standardized each of these school year–level averages across all school years to facilitate comparisons across factors.

Empirical Methods

Our primary identification strategy involves isolating within-school variation over time in order to estimate the relationship between changes in schools' organizational contexts, teacher turnover, and student achievement.

In previous studies, researchers have identified these parameters using cross-sectional variation in organizational context measures across schools. The limitation of this approach is that it cannot account for a host of potentially unmeasured, between-school differences in student, teacher, and school characteristics that might be correlated with the school context measures as well as teacher turnover and student achievement.

Our preferred modeling approach is to fit models at the most fine-grained level of measurement for our outcomes of interest, include school fixed effects to isolate within-school variation, and adjust our standard errors accordingly for the nested nature of our data. This approach is somewhat conservative in that it excludes all between-school variation from our estimates, and it does not leverage the added precision of directly modeling our nested error structure. In supplemental analyses available on request, we confirm that our results are consistent, if not somewhat larger and even more precisely estimated, when we fit random effects multilevel models where Level 1 is time, Level 2 is either individual teachers or students, and Level 3 is schools.

In our first set of analyses, we model the probability an individual teacher does not return to his or her school the following year in a teacher-year level panel data set. We model this binary outcome for teacher j in school s at time t using a logistic regression model, which can be expressed as follows:

$$Pr(Turnover_{jst}=1|Dimensions_{st}, T_{jst}, \bar{T}_{st}, \bar{S}_{st}, \pi_s, \gamma_t) = \frac{1}{1+e^{-k}},$$

where $k = \beta' \ Dimensions_{st} + \eta' T_{jst} + \delta' \bar{T}_{st} + \varphi' \bar{S}_{st} + \pi_s + \gamma_t.$

Here, a teacher's decision to not return to his or her school is modeled as a function of our primary question predictors, *Dimensions*$_{st}$, a vector of four school-context factors described previously. T_{ist} is a vector of individual teacher characteristics. $\bar{T}_{st}$ captures these same teacher characteristics averaged within school and year. $\bar{S}_{st}$ is a vector of student characteristics averaged within school and year, and π_s and γ_t are school and year fixed effects, respectively. Average teacher characteristics include controls for gender, race, experience, and degrees. School characteristics include controls for the proportion of students by gender, race, free/reduced-lunch eligibility, special education status, and English language learner status, as well as log-enrollment and an indicator for schools that provide free lunch to all students (i.e., universal feeding schools).

In our second set of analyses, we build on a large body of prior work in the education and economic literatures in which researchers model student achievement as a function of student, teacher, and school factors (e.g.,

Koedel, Mihaly, & Rockoff, 2015; Ronfeldt et al., 2013; Todd & Wolpin, 2003). We fit an ordinary least squares regression model where student test scores for student i in grade g in school s at time t are modeled as follows:

$$A_{igst} = f(A_{igst-1}) + \beta' \ Dimensions_{st} + \eta' S_{igst} + \delta' \bar{T}_{st} + \varphi' \bar{S}_{st} + \pi_s + \alpha_{gt} + \varepsilon_{igst}.$$

Here, student achievement, A_{igst}, in either mathematics or ELA is modeled as a cubic function of prior year achievement in both subjects, a vector of individual student characteristics, S_{igst}, and vectors of school context dimensions, school-level teacher characteristics, and school-level student characteristics described previously. We again include school fixed effects, π_s, and add grade-by-year fixed effects, α_{gt}, to account for grade-by-year specific shocks, such as differences in test scales and content. Individual student characteristics include measures for gender, race, free or reduced-price lunch eligibility, special education status, and English language learner status. Across both models, we weight our estimates by the teacher response rate for each school in each year to account for the varying precision of our school context measures. This serves to guard against the possibility that differential teacher response rates could bias our estimates.

Across both models, our primary parameters of interest are the estimates of β associated with each of our four school context measures. Our estimates of β will necessarily understate the true magnitude of these relationships given that the measurement error inherent in teachers' survey responses reduces the reliability of our school context measures. We account for the potential of correlated errors by clustering our standard errors at the school-year level.

Findings

School Contexts in NYC Middle Schools

In order to provide readers with intuitive measures of how NYC DOE middle school teachers perceive their school contexts, we calculate the percentage of teachers who either agree or strongly agree with each survey item and present the results in the Appendix Table A1. Overall, we find that mean agreement rates among teacher responses in our full panel range from 74% to 93% across the 33 survey items we analyze. We see that teachers were most likely to agree with the items that load most strongly on the Expectations dimension, followed closely by the Relationships dimension. Teachers were somewhat less willing to agree that their schools benefited from high-quality leadership and professional development and that they taught in safe and orderly schools.

Descriptive statistics and distributions reveal substantial variation across our four school context measures. The distributions of these measures all
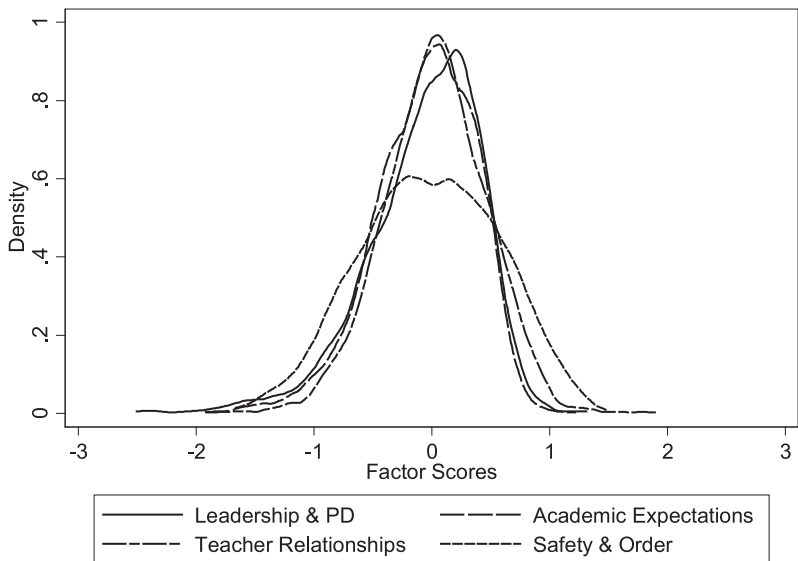
Figure 1. **Probability density functions of our four unstandardized school context measures in a school-by-year data set (*n* = 1,150). The underlying unit of measurement is teachers' responses to the School Survey on a 4-point Likert scale, which were standardized in a teacher-year data set.**

broadly approximate normal curves with varying degrees of negative skewness (see Figure 1). Our measures of Leadership, Expectations, and Relationships have similar distributions, with Relationships demonstrating very little skewness, while both Leadership and Expectations have somewhat longer left-hand tails. In comparison, our measure of the Safety in schools is much more variable, with more density in the lower tail of the distribution capturing schools with very low levels of Safety as perceived by teachers. An analysis of variance in a school-year data set reveals that nearly half of the total variation in teachers' assessments of the Leadership (50%), Expectations (49%), and Relationships (46%) dimensions is within schools, over time. We find notably less within-school variation in the Safety measure (28%), which makes sense given that this dimension, more than any other, is influenced by factors outside of a school's control.

We next examine the correlation between our school context measures, average student characteristics in a school, and outcomes in Table 4. Several important patterns emerge. The two school context measures that characterize adult relationships, either between administrators and teachers (Leadership) or among teachers (Relationships), are uncorrelated with

The Correlations Between School Context Measures and Average Student Characteristics

	Leadership and Professional Development	Academic Expectations	Teacher Relationships and Collaboration	Safety and Order
Percentage free/reduced-price lunch eligible	.062*	−.222*	.001	−.283*
Percentage special education	.033	−.178*	.067*	−.361*
Percentage English language learner	.054	−.173*	.051	−.085*
Percentage African American	.000	−.084*	−.075*	−.342*
Percentage Hispanic	.006	−.216*	.103*	−.091*
Percentage turnover	−.244*	−.356*	.062*	−.296*
Mathematics achievement	.027	.409*	−.010	.610*
English language arts achievement	.028	.406*	−.025	.526*

Note. n = 1,150. Values indicate correlations between school-year means. Mathematics and English language arts achievement are the school-year average student test scores on the New York State exams standardized within grade, subject, and year.
*p < .05.

average student achievement on New York State tests and largely unrelated to student demographic characteristics. Leadership has a moderately negative correlation with teacher turnover and is unrelated to achievement levels, while Relationships, unexpectedly, has a small positive correlation with turnover and is unrelated to achievement levels. In contrast, the two measures that characterize educators' interactions with students, Expectations and Safety, are consistently correlated with student performance on state tests as well as student demographic characteristics. Both of these measures have moderate negative correlations with turnover and moderate to strong positive correlations with academic achievement.

Teacher Turnover

We estimate the relationship between school context measures and our outcomes of interest using models both without and with school fixed effects as well as with all measures included separately and simultaneously. This produces four sets of results for each outcome of interest and helps to illustrate important differences across estimation strategies. In Table 5, we present predicted marginal effects of a one standard deviation increase in measures of the school context on the probability a teacher does not return the next year (hereafter "turnover"). Our model relates school context

Table 5

Predicted Marginal Effects of Dimensions of the School Context on the Probability of Teacher Turnover

	Percentile of the School Context Distribution at Which the Predicted Marginal Effect Is Estimated					
	10th (1a)	50th (1b)	90th (1c)	10th (2a)	50th (2b)	90th (2c)
	Panel A: Coefficients estimated for each factor separately					
Leadership and professional development	−.021***	−.019***	−.017***	−.02***	−.018***	−.016***
	(.003)	(.002)	(.002)	(.003)	(.003)	(.002)
Academic expectations	−.022***	−.019***	−.017***	−.011**	−.010**	−.010**
	(.003)	(.003)	(.002)	(.004)	(.003)	(.003)
Teacher relationships and collaboration	−.008**	−.008**	−.007**	−.008*	−.007*	−.007*
	(.003)	(.003)	(.002)	(.003)	(.003)	(.003)
Safety and order	−.021***	−.019***	−.017***	−.014***	−.013***	−.012***
	(.003)	(.003)	(.002)	(.004)	(.004)	(.003)
	Panel B: Coefficients estimated for each factor simultaneously					
Leadership and professional development	−.014***	−.013***	−.012***	−.017***	−.016***	−.014***
	(.003)	(.002)	(.002)	(.003)	(.003)	(.002)
Academic expectations	−.011***	−.011***	−.01***	−.006*	−.006*	−.006*
	(.003)	(.003)	(.002)	(.003)	(.003)	(.003)
Teacher relationships and collaboration	−.007**	−.007**	−.006**	−.008**	−.008**	−.007**
	(.003)	(.002)	(.002)	(.003)	(.003)	(.003)
Safety and order	−.013***	−.012***	−.011***	−.009*	−.008*	−.008*
	(.003)	(.003)	(.002)	(.004)	(.004)	(.003)
School fixed effects		—			Yes	
n (teachers)		53,991			53,991	

Note. Standard errors reported in parentheses are clustered by school-year. Estimates are derived from logistic regression models weighted by teacher response rates on the school survey. Coefficients are reported as predicted turnover probabilities. All models include vectors of individual teacher characteristics, average teacher and student characteristics at the school-year level, and year fixed effects. Individual and average teacher covariates include controls for teacher gender, race, experience, and degrees. Average student characteristics include controls for student enrollment (logged) and average student gender, race, free/reduced-price lunch status, special education status, and English language learner status.
*p < .05. **p < .01. ***p < .001.

1427

dimensions to teacher turnover via an S-shaped logistic curve, which allows the marginal effects to differ across the distribution of our predictors. We present estimates of the relationship between predictors and the predicted probability of turnover at the 10th, 50th, and 90th percentile of the distribution of each school context measure to characterize this curvilinear relationship. The results we present in Table 5 illustrate four important findings: (a) We replicate previous findings that, on average, schools with higher quality contexts experience lower turnover; (b) we show that improvements in the organizational context within a school over time are associated with corresponding decreases in teacher turnover; (c) we find that increases in school context measures have larger marginal effects on turnover for schools that start at lower levels of school context quality; and (d) we document the independent relationship between multiple dimensions of the school context and teacher turnover.

We derive estimates presented in Panels A and B, Columns 1a through 1c, following the primary modeling approaches used in the literature. Here we exploit both between- and within-school variation in school context measures and find that all four of our measures of school context are negatively associated with turnover. A one standard deviation increase in Leadership, Expectations, and Safety at the 50th percentile is associated with a 1.9 percentage point decrease in turnover for each measure when included separately. When all measures are included simultaneously in the model, these estimates are slightly attenuated; however, their sign, relative magnitudes, and statistical significance remain unchanged. These jointly conditional estimates show that a uniform one standard deviation increase at the 50th percentile across all four school context measures is associated with a 4.3 percentage point decrease in turnover.

These baseline estimates are comparable to prior studies that relied solely on cross-sectional variation. Boyd and his colleagues (2011) found that a uniform one standard deviation increase across six working condition measures was associated with approximately a 6 percentage point decrease in the probability first-year teachers did not return to their school. When the authors examined this relationship among non–first year teachers, their estimates were attenuated and closely approximate our own. Ladd (2011) found that a one standard deviation increase in the quality of school leadership was associated with a 5.6 percentage point decrease in self-reported planned departures but only a 1.8 percentage point decrease for actual departures—an estimate very similar to our own.

We present our preferred estimates from models that include school fixed effects in Columns 2a through 2c. These estimates document the meaningful and statistically significant relationship between all four measures of the school context and teacher turnover within schools over time. Across these results, Leadership emerges as having the strongest relationship with turnover among the four school context measures. In Panel A, where each

school context measure is included separately, a one standard deviation increase in Leadership at the 50th percentile is associated with a 1.8 percentage point decrease in teacher turnover. This estimate is more than twice the magnitude of the coefficients associated with Expectations and Relationships and 38% larger than the estimate for Safety.

In Panel B, we show that even when we restrict estimates to within-school variation over time and control for all dimensions simultaneously, all four school context measures remain independent and significant predictors of teacher turnover. Similar to Boyd et al. (2011) and Ladd (2011), we find that Leadership is the dominant predictor among our school context measures. The relative stability of our estimate for Leadership when we include additional school context measures suggests that the relationship between Leadership and turnover is largely direct instead of being mediated by other school context factors. General linear hypothesis tests confirm that the coefficient on Leadership in the logistic model is at least marginally statistically significantly different from both those on Expectations (p = .03) and Relationships (p = .06) but not for Safety (p = .18). Given the average turnover rate among middle school teachers in NYC is 15.1%, a one standard deviation increase in the quality of Leadership alone is associated with approximately an 11% reduction in turnover. The marginal effects of Safety, Relationships, and Expectations become .8, .8, and .6 percentage points, respectively. When we disaggregate our turnover outcome into movers and leavers, we find that school context measures are strong predictors of teachers' decisions to transfer schools but, as one might hypothesize, only weakly associated with teachers' decisions to leave the classroom or district altogether.

These results suggest that improving the school context in which teachers work could play an important role in a multifaceted plan to reduce teacher turnover among middle schools. As our results in Columns 2a through 2c of Panel B document, improvements in all four dimensions of the school context are independently related to reductions in turnover. If a school at the 50th percentile of the distribution in Leadership, Expectations, Relationships, and Safety was able to improve each of these dimensions of the school context by one standard deviation (up to the 84th percentile), our estimates suggest that this could reduce turnover by 3.8 percentage points, a 25% reduction in average turnover rate.

Student Achievement Gains

In Table 6, we examine the relationship between changes in the quality of schools' organizational contexts and student achievement in mathematics and ELA. Given that our models condition on students' prior achievement, we characterize our models as capturing the relationship between school contexts and student achievement gains (i.e., how well a student performs given what we would have predicted based on their prior performance

Table 6

The Relationship Between Dimensions of the School Context and Student Achievement

	Mathematics		English Language Arts	
	(1)	(2)	(3)	(4)
	Panel A: Coefficients estimated for each factor separately			
Leadership and professional development	.014**	.013**	.008*	.004
	(.004)	(.005)	(.003)	(.004)
Academic expectations	.033***	.013**	.020***	.002
	(.005)	(.005)	(.004)	(.004)
Teacher relationships and collaboration	.007	.009$^+$	−.002	.003
	(.004)	(.005)	(.004)	(.004)
Safety and order	.056***	.032***	.025***	.013**
	(.005)	(.005)	(.003)	(.004)
	Panel B: Coefficients estimated for each factor simultaneously			
Leadership and professional development	−.005	.006	−.002	.002
	(.004)	(.005)	(.004)	(.004)
Academic expectations	.026***	.009*	.018***	.000
	(.005)	(.005)	(.004)	(.004)
Teacher relationships and collaboration	.001	.008$^+$	−.006	.004
	(.004)	(.005)	(.004)	(.004)
Safety and order	.052***	.030***	.022***	.013**
	(.005)	(.006)	(.004)	(.004)
School fixed effects	—	Yes	—	Yes
n (students)	641,515	641,515	626,913	626,913

Note. Standard errors reported in parentheses are clustered by school-year. Estimates are derived from ordinary least squares regression models weighted by teacher response rates on the School Survey. All models include cubic functions of prior achievement in math and English language arts, vectors of individual student characteristics, vectors of average student and teacher characteristics at the school-year level, and grade-by-year fixed effects. Individual student characteristics include controls for gender, race, free/reduced-price lunch status, special education status, and English language learners. Average teacher covariates include controls for teacher gender, race, experience, and degrees. Average student covariates include all individual measures as well as student enrollment (logged).

$^+ p < .10.$ $^* p < .05.$ $^{**} p < .01.$ $^{***} p < .001.$

and our full set of covariates). Our results reveal four key findings: (a) We replicate and extend previous research findings that schools with higher quality school contexts have students who experience larger achievement gains, (b) we show that improvements in the school context within a school over time are associated with corresponding increases in student achievement gains, (c) we find that the relationship between the school context and student achievement gains at a school is stronger in mathematics than in ELA, and (d) we illustrate that the relationship between the school context and student achievement gains varies considerably across dimensions.

Estimates from our baseline models presented in Columns 1 and 3 reveal meaningful, positive associations between Safety, Expectations, and Leadership with student achievement gains in both subjects. We find that Safety has the strongest relationship with student gains across both subjects, where a one standard deviation change is associated with a .056 and .025 standard deviation (*SD*) increase in mathematics and ELA, respectively. A one standard deviation increase in academic expectations is associated with a .033 and .020 *SD* difference in mathematics and ELA achievement, respectively. However, our estimates of the association between Relationships and achievement are near zero and not statistically significant in all models. As we saw with teacher turnover, including all four dimensions of the school context in our baseline models somewhat attenuates our estimates.

Results from our preferred models reported in Panel B, columns 2 and 4 demonstrate that improving the Safety and Expectations in a school are associated with corresponding improvements in student achievement gains. We find that a one standard deviation increase in the Safety of a school is associated with a .030 and .013 *SD* increase in student achievement in mathematics and ELA, respectively, when controlling for all other school context measures. General linear hypothesis tests confirm that the magnitude of the coefficient on Safety is statistically different than those from the other three school context dimensions for math but only statistically significantly different from Expectations for ELA. An increase in the Expectations for students in a school is also associated with achievement gains in mathematics but not in ELA. This pattern of stronger relationships between schooling and student academic growth in mathematics compared to ELA is a consistent finding in the education research literature (e.g., Rich, 2013).

The attenuation of our estimates when we include school fixed effects provides some evidence that prior cross-sectional estimates may be upwardly biased due to positive sorting patterns. Consistent with prior theoretical models (Hallinger & Heck, 1998; Supovitz et al., 2010), our results suggest that the link between Leadership and student achievement is likely indirect and mediated through the school's culture and teachers' practices. While the within-school relationship between Leadership and student gains is statistically significant in the unconditional model for math (Panel A Column 2), it is attenuated and no longer statistically significant when we

include all school context measures simultaneously (Panel B Column 2). In exploratory analyses not reported here, we also find that the relationships between school context measures and student achievement do not appear to be mediated through teacher turnover. Our estimates are nearly identical when controlling for turnover in either the current or previous year.

The magnitudes of these within-school relationships are small but meaningful when placed in context. Our results are estimated from models conditioned on all four school context measures and thus have the potential to be realized simultaneously, assuming an underlying causal relationship. This suggests that schools able to improve each of these organizational contexts simultaneously by one standard deviation could increase student achievement growth by .053 *SD* in mathematics and .019 *SD* in ELA. Our estimates are substantially smaller than the size of teacher effects on student achievement, which are approximately .15 *SD* in mathematics and .10 *SD* in ELA per standard deviation difference in teacher effectiveness (Hanushek & Rivkin, 2010). However, very few programs have been shown to impact teacher effectiveness at scale (e.g., Garet et al., 2008, 2011; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Ronfeldt et al. (2013) find that a one standard deviation decrease in teacher turnover corresponds with an even smaller .02 and .01 *SD* increase in student achievement in mathematics and ELA, respectively.

Threats to Validity

Interpreting our results as suggestive of a causal relationship requires us to adopt several additional assumptions. Here, we examine these assumptions as well as a range of plausible alternative explanations to better understand the underlying relationships between school context dimensions, turnover, and student achievement.

Within-School Variation in School Context Measures

Our preferred modeling approach restricts the variation in our measures of the school context to variation within schools over time. However, it could be that the variation within schools among these measures is too limited to identify credible estimates of the relationship between these measures and our outcomes of interest. It is also possible that sufficient variation exists but that this variation is largely due to measurement error. The analysis of variance estimates we presented earlier revealed that there exists substantial variation in school context measures within schools over time. We turn our attention here to study whether this variation appears to be largely systematic or primarily driven by random noise.[5]

We accomplish this by estimating the proportion of within-school variation that can be explained by school-specific linear trends. We do this within a fixed effects framework where we first obtain R^2 values from models in

Table 7
Exploratory Analyses of the Within-school Variation in School Context Measures

	R^2 With School Fixed Effects	R^2 With School Fixed Effects and School-Specific Linear Trends	Proportion of Within-School Variance Explained by School-Specific Linear Trends
Leadership and professional development	.615	.778	.423
Academic expectations	.620	.781	.423
Teacher relationships and collaboration	.640	.828	.522
Safety and order	.789	.888	.472

Note. Column 1 contains the R^2 values of models that predict each measure of the school context using a full set of school fixed effects. Column 2 contains the R^2 values of models that predict each measure of the school context using a full set of school fixed effects and school-specific linear trends. Column 3 is the product of the following calculation: (R^2 Model 2 – R^2 Model 1)/(1 – R^2 Model 1). Estimates in Column 3 are also identical to models where the residuals from the specifications in Column 1 (demeaned school context measures) are regressed on school fixed effects and school-specific slopes.

which we regress a given school context dimension on a full set of school indicator variables in a school-year data set. These estimates are reported in Column 1 of Table 7. We then augment these models to also include school-specific linear trends by interacting each school indicator with a linear term for year. We show the resulting R^2 estimates from these augmented models in Column 2. In Column 3, we present our quantity of interest, the proportion of within-school variation explained by school-specific linear trends. The numerator in this ratio is the additional variation explained by the school-specific linear time trends (i.e., Column 2 – Column 1); the denominator is the proportion of total variance that is within schools over time (i.e., 1 – Column 1). This exercise suggests that linear trends explain between 42% and 52% of the total within-school variation. Thus, there appears to be both substantial and meaningful within-school variation in school context measures that can support our within-school identification strategy.

Common Source Bias

A central concern is the endogenous relationship between teachers' responses on the NYC DOE School Survey and their decisions about whether to return to their school in the following year. It could be that teachers who have decided they are leaving their school focus more on the negative aspects of their experiences and rate their school lower than they

would otherwise. It is also possible that teachers' responses are shaped by their perceptions of how well students are performing in a given year, although this threat is less plausible given the much greater challenge of predicting student achievement gains compared to student achievement levels. We address these concerns by replacing measures of the school context based on teachers' responses to the School Survey with measures constructed using students' responses. This breaks the potential link between teachers' self-reported perspectives of the school context, their direct control over turnover, and their influence over measures of student achievement.

Although the items on the student survey differ from those on the teacher survey, there are seven questions that map on to the dominant items from the Expectations factor and nine questions that map onto the Safety factor.[6] Measures of students' and teachers' perceptions of the Expectations in their school have a correlation of .22, while perceptions of Safety have a correlation of .68 in a school-year data set. Results using these alternative student-based measures are consistent with our main findings and of comparable magnitudes. As we show in Table 8, the relationship between Expectations and turnover is significant and even slightly larger, while the estimated coefficient for Safety when predicting turnover is identical but less precisely estimated. Both Expectations and Safety remain significant predictors of student achievement gains in mathematics as well. These findings are strong evidence of the validity of our school context measures based on teachers' perceptions as well as of the robustness of the relationship between the school context, turnover, and student achievement.

Reverse Causality

Another important challenge is determining the direction of the relationship between school context measures, turnover, and student achievement. Teachers typically complete the School Survey in March, several months before the end of the year, when they likely make career decisions and when students take standardized tests. Thus, the necessary temporal order between our predictors and outcomes for a causal relationship is satisfied by our modeling approach.

We test for reverse causality by conducting a set of falsification tests where we predict our school context measures in the following year (time $t + 1$) using measures of whether a teacher turned over at the end of the current year (time t) as well as current-year student achievement gain scores in mathematics and ELA. We calculate gain scores as the residual from a simple model where students' test scores (time t) are regressed on our vector of cubic functions of prior test scores in mathematics and ELA (time $t - 1$), following West et al. (2016). In Table 9, we show that neither turnover nor students' gains in mathematics are significant predictors of any of our four school context measures. We do find a small but significant relationship

Table 8
The Relationship Between Dimensions of the School Context Constructed From Student Surveys, Teacher Turnover, and Student Achievement

	Turnover		Mathematics		English Language Arts	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Coefficients estimated for each factor separately					
Academic expectations	−.016***	−.012***	.034***	.025***	.011**	.007
	(.003)	(.004)	(.005)	(.005)	(.004)	(.004)
Safety and order	−.016***	−.012**	.064***	.024***	.040***	.007
	(.003)	(.004)	(.006)	(.006)	(.005)	(.005)
	Panel B: Coefficients estimated for each factor simultaneously					
Academic expectations	−.012***	−.010*	.009+	.021***	−.007	.005
	(.003)	(.004)	(.005)	(.006)	(.004)	(.005)
Safety and order	−.010**	−.008	.059***	.016*	.043***	.005
	(.003)	(.005)	(.006)	(.007)	(.005)	(.006)
School fixed effects	—	Yes	—	Yes	—	Yes
n	53,835	53,835	639,750	639,750	625,152	625,152

Note. Standard errors reported in parentheses are clustered by school-year. Estimates with turnover as an outcome are predicted marginal effects at the 50th percentile of a given school context dimension. See Tables 5 and 6 for further details about model specifications.
$+p < .10.$ $*p < .05.$ $**p < .01.$ $***p < .001.$

Table 9
Exploratory Tests for Reverse Causality

	n	Outcomes (time *t* + 1)			
		Leadership and Professional Development	Academic Expectations	Teacher Relationships and Collaboration	Safety and Order
Predictors (time *t*)					
Turnover	42,416	−.002	−.010	.002	.001
		(.007)	(.006)	(.007)	(.005)
Residualized gain score: Mathematics	502,746	−.001	.002	−.002	.002
		(.003)	(.003)	(.003)	(.003)
Residualized gain score: English language arts	491,383	.007***	.003	.004*	.002
		(.002)	(.002)	(.002)	(.002)
School fixed effects		Yes	Yes	Yes	Yes

Note. Each cell represents results from a separate regression with standard errors clustered by school-year reported in parentheses. Estimates are derived from ordinary least squares regression models. Residualized gain scores are the residuals from a regression model of students' achievement scores in a given subject regressed on cubic functions of prior test scores in mathematics and ELA.
*$p < .05$. **$p < .01$. ***$p < .001$.

between student gains in ELA and teachers' ratings of Leadership and Relationships in the following year. These results suggest that teachers may respond to gains in ELA when evaluating their school leadership and relationships with their peers. However, our primary findings do not include evidence of corresponding relationships between these three measures in the opposite direction; thus, these results pose little threat to our analyses.

In supplemental analyses not presented here, we replicate these results when we exchange our outcome and predictors but predict school context measures from time *t* using outcome measures from time *t* − 1. We also confirm that our primary results hold in panels that cover 2008–2011 and 2009–2012 to ensure these results are not due to the restricted four-year panel data sets for which lagged and lead measures are available. Overall, these results show that our estimates are unlikely to be primarily driven by the influence of teacher turnover and student achievement gains on teachers' perceptions of their school contexts.

Omitted Variable Bias

A final threat is the possibility of omitted variables that are correlated with changes in measures of the school context and our outcomes of interest within schools over time. We address this threat by including a rich set of

individual student or teacher characteristics as well as time-varying average student and teacher characteristics for each school in a given year. However, these covariates constructed from administrative data are far from exhaustive. We attempt to gain some intuition about the potential magnitude and direction of any omitted variable bias following Altonji, Elder, and Taber's (2005) classic analysis of selection bias in estimating Catholic school effects. We accomplish this by examining the degree to which our estimates change when we remove student and teacher individual demographic controls as well as average student and teacher characteristics. As shown in Table 10, estimates from models that exclude our rich set of time-varying measures are almost identical to our preferred estimates and don't appear to differ in any systematic way. If potential omitted variables are correlated with these observed variables and have similar relationships with our outcomes, it would not appear that their omission would bias our results substantially.

Perhaps the most important potential omitted variable threat is our lack of a more comprehensive set of school context measures. Our analyses are limited to the number and type of school-context dimensions suitable for our analysis in the NYC DOE School Survey. The four dimensions we measure are among the most common features of the school context found in the research literature and ones with strong theoretical groundings. Important dimensions like teacher trust and time for collaboration are partially but not fully reflected in our measure of Relationships. However, we identified three primary dimensions that are omitted in our analyses based on our review of the literature: parent and community support, teacher leadership, and resources and facilities. We cannot definitely rule out that the relationships we find may reflect an underlying relationship between one or more of these omitted dimensions and our outcomes.

Our concerns over the omission of the dimensions described previously is assuaged somewhat by the close alignment of our results with those from qualitative studies in which teachers report or explain their primary reasons for leaving a school. Ingersoll (2001) found that among a nationally representative sample, teachers in urban high-poverty schools cited dissatisfaction with their job caused by student discipline problems as the primary reason for leaving a school. Pallas and Buckley (2012) administered a survey to middle school teachers in almost half of the NYC DOE middle schools included in our analytic sample. They found that teachers cited a lack of student discipline and a lack of support from administrators as the two most important reasons they weighed when considering leaving a school. Consistent with these findings, our measures of the Leadership and Safety are the strongest predictors of teacher turnover.

Further evidence of the internal validity of our findings comes from the dynamic relationship we find between specific dimensions of the school context and our outcomes. For example, Leadership has a strong association with teachers' career decisions while having no direct relationship with

Table 10

The Relationship Between Dimensions of the School Context, Teacher Turnover, and Student Achievement With and Without Control Variables

	Turnover		Mathematics		English Language Arts	
	(1)	(2)	(3)	(4)	(5)	(6)
	Coefficients estimated for each factor simultaneously					
Leadership and professional development	−.016***	−.018***	.006	.008+	.002	.004
	.003	.003	(.005)	(.005)	(.004)	(.004)
Academic expectations	−.006*	−.006+	.009*	.009+	.000	.002
	.003	.003	(.005)	(.005)	(.004)	(.004)
Teacher relationships and collaboration	−.008**	−.007*	.008+	.010*	.004	.004
	.003	.003	(.005)	(.005)	(.004)	(.004)
Safety and order	−.008*	−.008*	.030***	.028***	.013**	.013**
	.004	.004	(.006)	(.006)	(.004)	(.005)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
n	53,991	53,991	641,9515	641,9515	626,913	626,913

Note. Standard errors reported in parentheses are clustered by school-year. Estimates with turnover as an outcome are predicted marginal effects at the 50th percentile of a given school context dimension. Covariates is shorthand for all student and teacher individual demographic controls as well as average student and teacher characteristics. We retain controls for prior achievement in both subjects when modeling student achievement because we are most interested in the achievement of students conditional on their past performance.

$+p < .10.$ $*p < .05.$ $**p < .01.$ $***p < .001.$

student achievement gains when controlling for other school context dimensions. If our results were driven by self-report bias, nonresponse bias, or omitted variable bias, we would expect these biases to be common across all our measures of the school context. Instead, our results differ markedly across our four measures and with each outcome of interest.

Conclusion

This article contributes to a growing body of empirical literature that examines the organizational contexts in which teachers work and students learn. Our analyses suggest that when schools strengthen the organizational contexts in which teachers work, teachers are more likely to remain in these schools, and student achievement on standardized tests increases at a faster rate—findings that are robust to a range of potential threats. School administrators' leadership skills emerge as particularly salient for whether teachers decide to remain in their schools. The degree to which students and teachers feel their school is a safe, orderly learning environment is of central importance for student achievement in the NYC middle schools we studied. These results further illustrate the role of both individual and organizational effectiveness when designing reforms aimed at raising student achievement.

Implications for Policy and Practice

Recent scholarship and federal policy have generated considerable momentum behind reform efforts aimed at remaking teacher evaluation systems and placing an effective teacher in every classroom. However, teachers do not work in a vacuum; their schools' organizational contexts can undermine or enhance their ability to succeed with students. When aspects of the school context—for example, a principal who is an ineffective instructional leader, a school that lacks a consistent disciplinary code—are partly, or largely, to blame for poor performance, efforts to measure and strengthen individual teacher effectiveness are unlikely to be adequate remedies in themselves (Bryk et al., 2010). For schools where teachers are trying to deliver high-quality instruction and improve their craft amid organizational dysfunction, continually reshuffling the staff is also unlikely to result in a successful school turnaround.

To complement the vast literature on teachers' individual effectiveness, the education sector needs a commensurate body of research and policy reform agenda aimed at measuring and strengthening schools' organizational contexts. Similarly, school and district leaders need reliable data about the strengths and weaknesses of both individual teachers and school organizations as a whole to inform systematic efforts to improve student performance. Encouragingly, districts are increasingly administering school context surveys, such as the New Teacher Center's Teaching, Empowering, Leading, and Learning Survey and CCSR's 5 Essential Supports Survey, as

well as student surveys, which could also be used to gather important information about students' perspectives on the school environment.

The challenge for researchers and policymakers is to develop effective ways to use these data to inform schools' organizational development. One promising approach might entail producing customized school reports that describe levels and trends in teachers' perceptions of schools' organizational contexts as well as relative comparisons with similar schools. District leaders and principals could use these reports to identify and target efforts aimed at strengthening specific organizational weaknesses. Given that the principalship itself offers one of the highest leverage points for shaping the organizational practices and culture of a school (Boyd et al. 2011; Bryk et al., 2010; Grissom, 2011), another potential extension of these findings could be to use measures of schools' organizational contexts to inform principal hiring and placement. For example, districts could offer incentives to principals with proven abilities to improve specific dimensions of organizational contexts (e.g., safety or teacher collaboration) to accept placements at schools that rate poorly on these dimensions. Districts could also encourage learning partnerships between schools with complementary strengths and weaknesses. Given the moderate magnitude of our results, such reforms will not eliminate achievement gaps or excessive teacher turnover on their own. However, our findings suggest that such initiatives should be an important element of reform efforts aimed at increasing teacher retention and student achievement.

Finally, we caution against hastily incorporating survey-based measures of organizational contexts into accountability systems as part of the expanded accountability metrics required under the Every Student Succeeds Act. Attaching high-stakes consequences to school context measures would create perverse incentives for principals to pressure teachers, students, and parents to rate their schools favorably and could undermine the value of these measures as diagnostic tools.

Implications for Future Research

We see several important directions for future research. The process of developing a comprehensive and reliable set of school context measures is still in its initial stages. We analyzed four school context dimensions that we viewed as being more immediately under the control of administrators and teachers. However, other dimensions excluded from our analyses or not captured by the NYC DOE School Survey in the years we analyzed, such as common planning time for collaboration, may be equally important to teachers' and students' experiences in school. Researchers should continue to invest in efforts to enhance the precision, conceptual clarity, and coverage of surveys intended to capture information about school organizational contexts. Our findings also highlight the need to develop even more nuanced

conceptual models that reflect how specific school context dimensions shape teachers' and students' experiences in school in different ways.

We also see the need for further qualitative research that examines why some efforts to strengthen organizational contexts are successful while others are not. Changing the culture and collective practices of a teaching staff is an interpersonal process that involves complex social dynamics. How do administrators successfully lead collective action in their schools to strengthen organizational practices? What do administrators do to ensure that behavioral norms are applied consistently and high expectations are always upheld? Such research can be the basis for developing the features of successful organizational reform strategies.

Advancing our understanding will also require researcher-practitioner partnerships to develop and evaluate the efficacy of interventions to strengthen schools' organizational contexts. Experiments to improve the organizational capacity of private sector firms have produced compelling evidence of the large causal effect of productive organizational practices (e.g., Bloom, Eifert, Mahajan, McKenzie, & Roberts, 2013). In the education sector, Fryer's (2014) randomized evaluation of the effect of introducing evidence-based practices from highly effective charter schools into low-performing public schools provides initial evidence of the promise of organizational reforms. Further experimentation with interventions designed to strengthen the organizational contexts in schools should play a critical role in ongoing efforts to strengthen teacher effectiveness and create schools where all students are supported to reach high academic standards.

Appendix Table A1
Raw Factor Loadings for School Context Dimensions

	Leadership and Professional Development	Academic Expectations	Teacher Relationships and Collaboration	Safety and Order	Percentage Agree or Strongly Agree
Unstandardized factor variance	7.02	5.92	4.45	3.69	na
Percentage of total variation explained	21.00	18.00	14.00	11.00	na
Survey Items		Raw Factor Loadings			
1. The principal at my school . . . communicates a clear vision for our school.	**.609**	**.508**	.139	.200	.857
2. The principal at my school . . . encourages open communication on important school issues.	**.656**	.432	.174	.186	.782
3. Curriculum, instruction, and assessment are aligned within and across the grade levels at this school.	.442	**.577**	.227	.161	.876
4. The principal at my school . . . is an effective manager who makes the school run smoothly.	**.658**	.439	.113	.265	.788
5. My school . . . has high expectations for all students.	.315	**.729**	.197	.260	.889
6. My school . . . sets high standards for student work in their classes.	.086	**.748**	.330	.191	.931
7. My school . . . has clear measures of progress for student achievement throughout the year.	.356	**.739**	.228	.179	.897
8. My school . . . helps students develop challenging learning goals.	.368	**.765**	.224	.174	.853
9. My school . . . helps students find the best ways to achieve their learning goals.	.396	**.733**	.238	.201	.853
10. My school . . . offers a wide enough variety of courses to keep students engaged.	.439	.482	.179	.270	.785
11. To what extent do you feel supported by: my principal?	**.693**	.332	.151	.214	.820

(continued)

Appendix Table A1 **(continued)**

	Leadership and Professional Development	Academic Expectations	Teacher Relationships and Collaboration	Safety and Order	Percentage Agree or Strongly Agree
12. To what extent do you feel supported by: other teachers at my school?	.092	.128	**.733**	.089	.919
13. School leaders . . . provide time for collaboration among teachers.	**.548**	.363	.350	.135	.874
14. Teachers in my school . . . respect colleagues who take the lead in school improvement efforts.	.259	.259	**.764**	.129	.882
15. Teachers in my school . . . trust each other.	.236	.208	**.782**	.169	.809
16. Teachers in my school . . . respect colleagues who are the most effective teachers.	.202	.213	**.809**	.132	.889
17. School leaders . . . give me regular and helpful feedback about my teaching.	**.611**	.426	.289	.144	.766
18. Teachers in my school . . . work together on teams to improve their instructional practice.	.287	.386	**.642**	.113	.865
19. Teachers in my school . . . use student achievement data to improve instructional decisions.	.327	.485	.474	.104	.932
20. Overall, my professional development experiences this school year have . . . provided me with teaching strategies to better meet the needs of my students.	**.727**	.239	.301	.092	.773
21. I have sufficient materials to teach my class(es), including: books, audio/visual equipment, maps, and/or calculators.	**.558**	.179	.237	.265	.768
22. I received helpful training on the use of student achievement data to improve teaching and learning this year.	**.665**	.310	.281	.117	.784

(continued)

Appendix Table A1 (continued)

	Leadership and Professional Development	Academic Expectations	Teacher Relationships and Collaboration	Safety and Order	Percentage Agree or Strongly Agree
23. Overall, my professional development experiences this school year have . . . provided me with content support in my subject area.	**.732**	.221	.291	.084	.740
24. Teachers and school leaders in my school use information from parents to improve instructional practices and meet student learning needs.	.487	.413	.306	.251	.753
25. My school communicates effectively with parents when students misbehave.	**.520**	.335	.237	.395	.831
26. Order and discipline are maintained at my school.	.497	.315	.179	**.583**	.750
27. I can get the help I need at my school to address student behavior and discipline problems.	**.556**	.282	.209	**.523**	.771
28. I am safe at my school.	.414	.247	.196	**.619**	.909
29. Crime and violence are a problem in my school.[a]	.092	.150	.082	**.794**	.812
30. Students in my school are often threatened or bullied.[a]	.159	.213	.121	**.740**	.660
31. Adults at my school are often disrespectful to students.[a]	-.069	.280	.256	**.527**	.892
32. There is a person or a program in my school that helps students resolve conflicts.	.392	.206	.211	.346	.884
33. My school is kept clean.	.331	.165	.202	.413	.868

Note. n = 31,699 survey responses. Factor loadings above .5 are highlighted in bold text. Loadings produced using varimax rotation following principal component analysis. Resulting factors are orthogonal (pairwise uncorrelated) across teacher-years. Factor loadings of .5 or larger are presented in bold text.

[a]Items have been reverse coded to have positive valence so that the reported agreement rates are the proportion of teachers who disagree with the statement.

Notes

[1]Response rates among middle school students were above 87% in each of the five years we study, while middle school parent response rates started at 39% in 2007–2008 and increased steadily to 58% by 2011–2012.

[2]We assign integer values to each response on the Likert scale and reverse code items with negative valence so higher ratings are associated with safer and more orderly environments.

[3]Measures that are constructed using an oblique rotation can be highly correlated, often preventing researchers from being able to include multiple factors in a single model due to multicollinearity. In our data, factor scores produced from oblique rotations are strongly correlated with our orthogonal factor scores, with pairwise correlations between the substantively similar factors of .88 or higher. Replacing our preferred measures with those constructed from oblique rotations produces nearly identical results to those we present when each factor is included separately and broadly consistent results when all measures are included in the model.

[4]Previous analyses have shown that alternative approaches to aggregating teachers' responses to the school-year level using a Jackknife or leave-out-mean approach produces nearly identical results as sample means (Johnson, Kraft, & Papay, 2012).

[5]We thank Sean Corcoran for his helpful suggestions that motivated these analyses.

[6]We construct these measures by mapping similar items across survey forms and then estimating factor scores for the two dimensions with common items following the same principal-component factor analysis process with an orthogonal rotation.

References

Allensworth, E., Ponisciak, S., & Mazzeo, C. (2009). *The schools teachers leave: Teacher mobility in Chicago Public Schools*. Chicago, IL: Consortium on Chicago School Research, University of Chicago.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151–184.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223–235.

Blase, J., & Blase, J. (1999). Principal's instructional leadership and teacher development: Teacher's perspectives. *Educational Administration Quarterly*, *35*(3), 349–378.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? Evidence from India. *Quarterly Journal of Economics*, *128*(1), 1–52.

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.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Carrell, S. E., & Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics, 2*(1), 211–228.

Dobbie, W., & Fryer, R. G. (2013). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics, 5*(4), 28–60.

Dumay, X., Boonen, T., & Van Damme, J. (2013). Principal leadership long-term indirect effects on learning growth in mathematics. *The Elementary School Journal, 114*(2), 225–251.

Figlio, D. N. (2007). Boys named Sue: Disruptive children and their peers. *Education Finance Policy, 2*(4), 376–394.

Fryer, R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics, 129*(3), 1355–1407.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., . . . Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement.* Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., . . . Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation.* Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement. *Teachers College Record, 109*(4), 877–896.

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.

Hackman, J. R., & Oldham, G. R. (1980). *Work redesign.* Reading, MA: Addison-Wesley.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement, 9*(2), 157–191.

Hallinger, P., & Heck, R. H. (2010a). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership & Management, 30*(2), 95–110.

Hallinger, P., & Heck, R. H. (2010b). Leadership for learning: Does collaborative leadership make a difference in school improvement? *Educational Management Administration & Leadership, 38*(6), 654–678.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100*(2), 267–271.

Heck, R. H., & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal, 46*(3), 659–689.

Heck, R. H., & Hallinger, P. (2010). Collaborative leadership effects on school improvement: Integrating unidirectional- and reciprocal-effects models. *Elementary School Journal, 111*(2), 226–252.

Honig, M. I. (Ed.). (2006). *New directions in education policy implementation: Confronting complexity.* Albany, NY: State University of New York Press.

Hoy, W. K., & Hannum, J. (1997). Middle school climate: An empirical assessment of organizational health and student achievement. *Educational Administration Quarterly*, *33*(3), 290–311.

Hoy, W. K., & Sabo, D. J. (1998). *Quality middle schools: Open and healthy.* Thousand Oaks, CA: Corwin Press

Hoy, W. K., Tarter, J. C., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, *43*(3), 425–446.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499–534.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85–108.

Johnson, S. M. (1990) *Teachers at work: Achieving success in our schools.* New York, NY: Basic Books.

Johnson, S. M. (2009). *How best to add value? Striking a balance between the individual and the organization in school reform* (Briefing Paper No. 249). Washington, DC: Economic Policy Institute.

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., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practice using student achievement data. *Journal of Human Resources*, *46*(3), 587–613.

Kanter, R. M. (1983). *The change masters: innovations for productivity in the American corporation*. New York, NY: Simon & Schuster.

Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher*, *39*(8), 591–598.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, *47*, 180–195.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, *36*(4), 476–500.

Kraft, M. A., Papay, J. P., Charner-Laird, M., Johnson, S. M., Ng, M., & Reinhorn, S. K. (2015). Educating amidst uncertainty: The organizational supports that teachers need to serve students in high-poverty, urban schools. *Educational Administration Quarterly*, *51*(5), 753–790.

Ladd, H. (2009). *Teachers' perceptions of their working conditions: How predictive of policy-relevant outcome?* (CALDER Working Paper No. 33). Washington, DC: CALDER.

Ladd, H. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, *33*(2), 235–261.

Lee, V. E., & Smith, J. B. (1996). Collective responsibility for learning and its effects on gains in achievement for early secondary school students. *American Journal of Education*, *104*(2), 103–147.

Leithwood, K., & Jantzi, D. (2006). Transformational school leadership for large-scale reform: Effects on students, teachers, and their classroom practices. *School Effectiveness and School Improvement*, *17*(2), 201–227.

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. New York, NY: Wallace Foundation.

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, *80*(3), 44–70.

Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.

Marinell, W. H., & Coca, V. M. (2013). *Who stays and who leaves? Findings from a three-part study of teacher turnover in NYC middle schools*. New York, NY: Research Alliance for New York City Schools. Retrieved from http://media.rany cs.org/2013/003

May, H., & Supovitz, J. A. (2010). The scope of principal efforts to improve instruction. *Educational Administration Quarterly*, *47*(2), 332–352.

Miller, R. J., & Rowan, B. (2006). Effects of organic management on student achievement. *American Educational Research Journal*, *43*(2), 219–253.

Murdock, T. B., Anderman, L. H., & Hodge, S. A. (2000). Middle-grade predictors of students' motivation and behavior in high school. *Journal of Adolescent Research*, *15*(3), 327–351.

Nathanson, L., Cole, R., Kemple, J., Lent, J., McCormick, M., & Segeritz, M. (2013). *New York City school survey 2008–2010: Assessing the reliability and validity of progress repot measures*. New York, NY: The Research Alliance for New York City Schools.

New York City Department of Education. (2008). *Educator guide. The New York City progress report. Elementary/middle school*. New York, NY: Author.

Pallas, A. M., & Buckley, C. K. (2012). *Thoughts of leaving: An exploration of why New York City middle school teachers consider leaving their classrooms*. New York, NY: Research Alliance for New York City Schools.

Papay, J. P., Bacher-Hicks, A., Page, L. C., & Marinell, W. H. (2015). The challenge of teacher retention in urban schools: Evidence of variation from a cross-site analysis. Retrieved from http://www.appam.org/assets/1/7/Marinell.pdf

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, *76*(1), 85–97.

Rich, M. (2013, May 29). In raising scores, 1 2 3 is easier than A B C. *The New York Times*.

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, *52*(3), 475–514.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4–36.

Rosenthal, R., & L., Jacobson. (1968). *Pygmalion in the classroom*. New York, NY: Holt.

Shen, J. (1997). Teacher retention and attrition in public schools: Evidence from SASS91. *The Journal of Educational Research*, *91*(2), 81–88.

Simon, N., & Johnson, S. M. (2015). Teacher turnover in high poverty schools: What we know and can do. *Teachers College Record*, *117*(3), 1–36.

Steinberg, M. P., Allensworth, E., & Johnson, D. W. (2011). *Student and teacher safety in Chicago Public Schools: The roles of community context and school social organization*. Chicago, IL: Consortium on Chicago School Research.

Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administrations Quarterly*, *46*(1), 31–56.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485), F3–F33.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement.* Aurora, CO: Mid-Continent Research for Education and Learning.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, *38*(1), 148–170.

Witziers, B., Bosker, R. J., & Kruger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *39*(3), 398–425.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Report No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.