

The **Research Alliance** for
New York City Schools

School Organizational Contexts, Teacher Turnover, and Student Achievement:

Evidence from Panel Data

Working Paper

Matthew A. Kraft
Brown University

William H. Marinell
Darrick Yee
Harvard University

March 2016

Abstract

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 contexts 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, academic expectations, teacher relationships, and school safety are all independently associated with corresponding reductions in teacher turnover. Increases in school safety and academic expectations for students also correspond to increases in student achievement. These results are robust to a range of potential threats to validity, suggesting that our findings are likely driven by an underlying causal relationship.

Keywords:

School Culture
School Climate
School Learning Environment
Working Conditions
Organizational Contexts
Teacher Turnover
Student Achievement
Middle School

About this research:

This study was conducted by Matthew Kraft (Brown University), William Marinell (Harvard University), and Darrick Yee (Harvard University) in collaboration with the Research Alliance for New York City Schools, with support from the William T. Grant Foundation. The findings and conclusions are those of the authors and not necessarily of the Research Alliance or its funders.

“The whole is more than the sum of its parts.”
– Aristotle, *Metaphysica*.

1. Introduction

There is increasing consensus among research and policy circles that teachers affect students' academic achievement more than any other school-related factor. In response, researchers are devoting considerable attention to investigating how to best measure teacher effectiveness (e.g. Kane, McCaffrey, Miller, & Staiger, 2013), while state and local policymakers are overhauling teacher evaluation systems and reforming human resource practices. Included in this reform agenda are initiatives to provide teachers with individualized feedback (Papay, 2012), incentivize their individual effort (Podgursky & Springer, 2007) and replace those who are persistently low performing (Hanushek, 2009). While these reforms may play a critical role in improving students' educational outcomes, they do not address a core aspect of teachers' work that influences their effectiveness and career decisions: the organizational contexts in which they teach (Johnson, 2009; Kennedy, 2010).

In teaching, as in any occupation where professionals perform their work in organizational contexts, productivity is shaped by both individual and organizational factors (Hackman & Oldham, 1980; Johnson, 1990; Kanter, 1983). 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 implemented and perceived within schools (Honig, 2006). To maximize teachers' efforts and students' achievement, researchers and policymakers must complement the extensive teacher effectiveness literature with a commensurate body of work measuring schools' organizational contexts and examining their relationships with important student and teacher outcomes.

Organizational theory and recent evidence suggest that school contexts affect student achievement through a variety of indirect and direct channels. Indirect effects likely operate through the influence of organizational contexts on teachers' career decisions and interactions with students. Studies consistently find chronic teacher turnover in schools with dysfunctional contexts and a lack of organizational supports (Simon & Johnson, 2015). High rates of teacher turnover impose large financial costs on schools (Barnes, Crowe, & Schaefer, 2007; Birkeland & Curtis, 2006; Milanowski & Odden, 2007) and reduce student achievement (Ronfeldt, Loeb, & Wyckoff, 2013) by, in

part, undercutting efforts to build capacity and coordinate instruction among a staff. Studies also repeatedly find that novice teachers are less effective, on average, than the more experienced teachers they often replace (Rockoff, 2004; Harris & Sass, 2011; Papay & Kraft, 2015).

Schools with supportive professional environments are not only more likely to retain their teachers; evidence suggests they also maximize teachers' and students' learning opportunities. Over time, teachers improve their ability to raise student achievement more when they work in school environments characterized by meaningful opportunities for feedback, productive peer collaboration, responsive administrators, and an orderly and disciplined environment (Kraft & Papay, 2014). The strong association between measures of school safety and average student achievement suggests that students are unable to concentrate on academics when they fear for their physical well-being (Steinberg, Allensworth, & Johnson, 2011). Students' motivation, effort, perseverance, and beliefs about their potential for academic success are also shaped directly by the academic expectations schools set for all students (Wentzel, 2002; Jussim & Harber, 2005).

In recent years, the proliferation of school surveys administered to teachers, students and parents has provided new opportunities to quantify dimensions of school organizational contexts and directly examine their relationship with teacher turnover and student achievement. Scholars have explored these relationships using data from schools in California, Chicago, Massachusetts, New York City, and North Carolina (Allensworth, Ponisciak & Mazzeo, 2009; Johnson, Kraft & Papay, 2012; Marinell & Coca, 2013; Boyd et al., 2011; Bryk et al., 2010; Ladd, 2011; Loeb, Darling-Hammond, & Luczak, 2005). Taken together, this growing body of work has established that organizational contexts are stronger in schools with lower teacher turnover and higher student achievement. However, this literature has been largely limited to cross-sectional analyses and longitudinal case studies preventing analysts from answering 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?

In this study, we provide the first direct evidence to inform answers to these critical questions 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 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 panel data allow us to identify these estimates using within-school variation over time by accounting for any time-invariant differences across schools and unobserved time-shocks through school and year fixed effects.

We focus our analyses in this paper on NYC middle schools for several reasons. The middle grade years are crucial in students' academic and social-emotional development and play a critical role in influencing students' high school and post-secondary outcomes (Balfanz, 2009; Balfanz, Herzog, & Mac Iver, 2007; Murdock, Anderman, & Hodge, 2000; Neild & Balfanz, 2006; Roderick, 1994). Despite this, evidence suggests that middle schools may be the most troubled of the broad school-level categories. Middle schools have uncommonly high rates of teacher turnover (Marinell & Coca, 2013; NCTAF, 2007), teachers often consider middle school assignments as less desirable than comparable elementary or high school assignments (Neild, Useem, & Farley, 2005), and middle school teachers may receive less tailored preparation than elementary and high school teachers (Neild, Farley-Ripple & Byrnes, 2009), which may compromise their effectiveness as individual practitioners.

We find that four distinct dimensions of middle schools' organizational contexts emerge from teachers' responses to the NYC School Survey: leadership & professional development, high academic expectations for students, teacher relationships & collaboration, and school safety & order. Exploiting within-school variation across time, we find robust relationships between select dimensions of the school context, teacher turnover, and student achievement in mathematics and English language arts (ELA) in a school. Improvements in the leadership & professional development, academic expectations, teacher relationships & collaboration, and safety & order within a school over time are all independently associated with decreases in teacher turnover. We also find compelling evidence that improvements in schools' safety & order and increases in academic expectations for students predict corresponding improvements in mathematics achievement. Finally, we 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.

2. Related Literature

Decades of qualitative research has illuminated the important and complex ways in which schools' organizational contexts affect teachers' motivation, job satisfaction, and sense of success (Lortie, 1975; Johnson, 1990; Johnson & Birkeland, 2003). As Johnson (1990) described, schools are complex organizations where a "constellation of features" interact to shape the work context for teachers. Johnson and Birkeland's (2003) longitudinal interview study of 50 new teachers revealed that the most important factor influencing teachers' career decisions was whether they felt they could be effective with their students. Teachers described how a variety of working conditions in schools, such as the nature of collegial interactions, the support of administrators, and school-wide approaches to discipline, either supported or undercut their own efforts. Drawing on interviews with teachers in high-poverty urban schools, Kraft et al. (2015) found that teachers consistently described the ways in which instructional support from administrators and school 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; Weiss, 1999). In their 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, Ponisciak, & Mazzeo, 2009; Johnson, Kraft, & Papay, 2012; Marinell & Coca, 2013; Boyd et al., 2011; Ladd, 2011; Loeb, Darling-Hammond, & Luczak, 2005). 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 these studies examine somewhat differing sets of school context dimensions, several of these dimensions consistently emerge as the strongest predictors, including the quality of school leadership, the degree of order and discipline in a school, and the support that collegial relationships provide.

While the relationship between school organizational contexts and teacher turnover is well established in the literature, we know much less about how these contexts relate to student achievement. To our knowledge, only three studies have examined

the relationship between the school context and students' academic outcomes. Ladd (2009) demonstrated that both teachers' perceptions of school leadership and the amount of common planning time (as captured on the North Carolina Teacher Working Conditions Survey) predicted a school's value-added in mathematics. Using data from a very similar statewide survey in Massachusetts, Johnson, Kraft, & Papay (2012) found that teachers' perceptions of their schools' working conditions were strong predictors of the median student growth percentile in their school in the following two years, after controlling for various student-, teacher-, and school-level characteristics. Finally, Bryk and his colleagues (2010) leveraged expansive data on the organizational practices and learning climates in Chicago Public Schools to explain why some elementary schools made substantial gains, while others did not, after the district decentralized control over schools.

These prior studies of the relationship between school contexts, teacher turnover, and student achievement took 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 & Johnson created measures by grouping and 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 effect 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.

Taken as a whole, this literature documents consistent evidence of the relationship between measures of the school organizational context and teacher turnover, as well as emerging evidence of a direct relationship between school contexts and student achievement. However, questions still remain about whether these relationships are, in fact, causal given that prior studies have largely relied on a single year of teacher survey data to construct measures of the school context.¹ This reliance on cross-sectional school context measures prevents researchers from ruling out several

plausible alternative explanations for the relationships they find. This paper provides new analyses that address many of the most important potential threats to these previous findings.

3. Research Design

3.1 Data

We combine four sources of data from the NYC DOE: student assessment data, school administrative data, human resources data, and teachers' responses to the NYC DOE's School Survey. Student assessment data includes information on students' demographic characteristics and their scaled scores on the State's intermediate-level exams in mathematics and ELA—the New York State Testing Program's standardized assessments, which are administered by the Office of State Assessment. School administrative records contain data on school type, grade configuration, enrollment, and other school characteristics. Human resources files maintained by the district capture demographic data on all personnel employed by the district, as well as job and assignment codes, salaries, and information on degrees and experience (as reflected in the salary schedule).

We complement these district administrative files with teachers' responses to the School Survey—an anonymous survey that is administered annually and designed to capture teachers' opinions across four broad reporting categories: Academic Expectations, Communication, Engagement, and Safety & Respect. The vast majority of survey items ask teachers to respond to statements by indicating the extent to which they agree or disagree on a four point Likert scale. Two items ask teachers the extent to which they feel supported by their principals or other teachers at their school, asking them to choose among four response anchors ranging from “to no extent” to “to a great extent.” We link these survey data to the school administrative files by aggregating our organizational context measures to the school level in each year. This approach is both necessary (as the anonymous nature of the survey prevents us from linking responses to human resources records) and aligned with our theoretical conceptualization of the constructs we aim to measure (i.e., school-wide organizational contexts). In addition, our approach is consistent with the prior literature, reduces classical measurement error that contributes to teacher-level variation, and is less susceptible to potential threats posed by self-report biases and reverse causality, as we discuss below. Of course, teachers in the same schools may

experience their organizational contexts differently. Because we lose this variation with our school-wide measures, we explore whether the degree to which teachers' perspectives differ within a school is an independent predictor of turnover and achievement as an extension of our primary analyses.

3.2 Sample

We create three primary analytic datasets using the data described above. Across all our datasets, we restrict our samples to include only students and teachers in middle schools with traditional grade 6-8 configurations, excluding schools spanning additional grades such as K-8 or 6-12. We also excluded schools in years when they were new (and still phasing-in to full grade 6-8 enrollment), expanding to include additional grades, or in the process of phasing out grades toward closure, in order to avoid conflating turnover with structurally induced employment patterns. This results in an analytic sample of 278 unique NYC DOE middle schools.

First, we construct a teacher-year level panel dataset that contains teachers' responses to the School Survey. Our analytic sample spans five years, from the 2007-08 school year through the 2011-12 school year, and contains over 31,000 responses. Although the district first administered the School Survey in 2006-07, we excluded responses to this initial administration from our analyses because of the relatively low response rate (46%) among middle school teachers. As shown in Table 1 on page 29, response rates jumped to 63% in 2007-08 and rose incrementally each following year, rising to 84% by 2011-12.²

Second, we construct a separate teacher-year level panel dataset, built from human resources files, for use in our turnover analyses. Importantly, this panel includes data through 2012-13, allowing us to create measures of teacher turnover for the 2011-12 academic year. We restrict this dataset to include only full-time middle school classroom or special education teachers (using the NYC DOE's approach to identifying active teachers), resulting in 53,991 teacher-year observations and 16,404 unique teachers. In Table 2 on page 29, we present descriptive characteristics of the middle school teachers in our sample. Almost 70% of the teachers are female, and a majority are White, while 22% are Latino, 13% are Black and 5% are Asian. Teacher experience varies across the sample, with second-stage teachers (i.e., those with 4-10 years of experience) making up the majority. Eighty-five percent of teachers in our sample hold a Master's or other graduate degree.

Third, we construct a student-year level panel dataset from NYC DOE student administrative records for use in our achievement analyses. We provide descriptive statistics on the students in our sample in Table 3 on page 30. Students who attend NYC public middle schools come from diverse backgrounds. Almost 40% of students are Latino, 27% are Black, 18% are Asian and 15% are White. Twelve percent of students in our sample receive special education services, and 7.5% are classified as English language learners.

3.3 Outcome Measures

We construct two sets of primary outcomes for the regression analyses described below: teacher turnover and student achievement on standardized State tests. To create the turnover outcomes, we examine teachers' school assignment and active status over time and categorize teachers into three mutually exclusive groups at the end of each year: *stayers* (i.e., teachers who remain as classroom teachers in their current school in the following year); *movers* (i.e., teachers who remain active classroom teachers but transfer to another NYC DOE school in the subsequent year); and *leavers* (i.e., teachers who leave their school and are not teaching in any other school in the NYC public school system the next year). In our data, leavers consist of teachers who secure non-teaching assignments or were no longer employed by the NYC public school system. We use a single overall measure of turnover in our primary analyses, which takes on a value of one for teachers who transfer to a different school (movers) or leave teaching in an NYC DOE school (leavers). The average turnover rate among middle school teachers in our data is 15%, with approximately 6% transferring to other schools and 9% taking non-teaching positions or leaving the NYC public school system (Table 2). These turnover rates are consistent with those from prior analyses of NYC public schools (Marinell & Coca, 2013) and studies of other larger urban school districts (Papay, Bacher-Hicks, Page & Marinell, 2015).

We use students' scaled scores on the New York State Testing Program's standardized assessments in mathematics and English language arts as our two primary student outcomes. Multiple choice questions comprise the majority of items on the state tests in both subjects, which also include several short response items and one to four extended response tasks. We standardize students' scaled scores within grade, year, and subject, in order to place these scores on a common scale and pool our analyses.

3.4 Constructing Measures of School Context

We draw upon both theory-driven and data-driven approaches to inform our construction of school context measures. To begin, we selected a subset of items that best matched our intended purposes of measuring malleable aspects of the school context. 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 administrators and teachers do not primarily control through their decisions and practices. This process allows us to focus our analyses on a parsimonious set of policy-relevant dimensions of the school organizational context. Our intention with this initial analysis is to provide administrators and policymakers with evidence that can inform their efforts to improve student and teacher outcomes by strengthening school contexts. This goal requires that we focus on dimensions of school context that practitioners can easily influence. However, we recognize that our decision to focus on school context dimensions that practitioners can more directly influence may have resulted in our omitting other dimensions, such as parent engagement, that may be related to teachers' career decisions and students' achievement. Thirty-three items met these criteria.

Prior research on the NYC DOE School Survey found that a theorized factor structure based on the survey's broad reporting categories did not fit the data well (Nathanson, et al., 2013).³ Given this, we do not endorse the survey's nominal reporting categories and instead include 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, et al. (2011). First, we conduct a principal components analysis on our panel dataset of teachers' responses to identify the number of meaningful, independent dimensions of school context that were captured by our 33 items. Second, we apply 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 are produced using these rotated loadings.

Our analysis of the principal components leads 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 (Ladd, 2011; Kane, et al., 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 do not share common variance in a teacher-level dataset.⁴ 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.⁵ We then standardized each of these school-year-level averages across all school-years to facilitate comparisons across factors.

Our context measures capture four broad organizational features of schools: leadership & professional development (Leadership), high academic expectations for students (Expectations), teacher relationships & collaboration (Relationships), and school safety & order (Safety). We arrived at these labels by characterizing the dominant items on each factor. We provide a complete list of the items and factor loadings in Appendix Table A1 on page 41.

Items asking teachers about the quality of school leadership, professional development opportunities, and feedback in a school loaded strongly onto the first factor, which we call Leadership. As shown Appendix Table A1, the Leadership dimension has the most items with large factor loadings (i.e., exceeding 0.50) and explains 21% of the variance across our 33 teacher item responses. Items with the strongest loadings onto the Leadership factor fell into two broad categories: 1) items that inquired directly about attributes of principals' leadership; and 2) 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 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 is dominated by a block of items that fall into a thematic category capturing the rigor of academic expectations for students in a school. This Expectations dimension explains 18% of the variance across teachers' survey responses. 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.

Lastly, we capture a fourth dimension that measures student behavior and the level of school safety. The Safety dimension explains 11% of the total variance. 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.⁶

3.5 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. We removed any time-invariant differences across schools in the quality of their organizational contexts and their teacher and student outcomes because we view this variation as likely to be driven by the strong selection biases that student and teacher sorting across schools

can induce. For example, if students who exhibit higher levels of motivation and effort (characteristics that are frequently unobservable to researchers) are more likely to attend schools with stronger school contexts, this could induce a positive bias in our estimates.

Our preferred modeling approach is to fit models at the most fine-grained level of measurement for our outcomes of interest, to include school fixed-effects to isolate within-school variation, and to 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 upon request, we confirm that our results are consistent if not somewhat larger and even more precisely estimated when we fit 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 her school the following year in a teacher-year level panel dataset. 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:

$$Pr(\text{Turnover}_{jts} = 1 \mid \text{Dimensions}_{st}, T_{ist}, \bar{T}_{st}, \bar{S}_{st}, \pi_s, \gamma_t) = \frac{1}{1 + e^{-k}}$$

where $k = \beta' \text{Dimensions}_{st} + \eta' T_{ist} + \delta' \bar{T}_{st} + \varphi' \bar{S}_{st} + \pi_s + \gamma_t$

Here, a teacher's decision to not return to her school is modeled as a function of our primary question predictors, Dimensions_{st} , a vector of four school-context factors described above. 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, Loeb, & Wyckoff, 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:

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

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 characters, and school-level student characteristics as described above. 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.

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.

4. Findings

4.1 School Contexts in NYC Middle Schools

In order to provide readers with intuitive measures of how NYC DOE middle school teachers perceive the school context in which they work, we calculate the percentage of teachers who either agree or strongly agree with each survey item in a given school-year. Overall, we find that middle school teachers view their schools relatively favorably; mean agreement rates among teacher responses in our full panel range from 74% to 93% across the 33 survey items we analyze. We report agreement rates for each item in Appendix Table A1. 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 (based on the Leadership dimension) and that they taught in safe and orderly schools (per the Safety dimension).

Descriptive statistics and distributions demonstrate the substantial variation across our four school context measures. Figure 1 on page 35 shows the probability density plots of our four unstandardized predictors in a school-year dataset. These plots suggest that the distributions of our measures all broadly approximate normal curves with varying degrees of negative skewness. 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.

We decompose the variation in our school context measures to examine the degree to which variation in these measures are driven by differences across schools or changes within schools over time. An analysis of variance in a school-year dataset 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 on page 30. Several important patterns emerge. The two school context measures that characterize adult relationships, either between administrators and teachers (Leadership) or among teachers (Relationships), are both uncorrelated with average student achievement on New York State tests and largely unrelated to student demographic characteristics. Leadership has a moderate 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.

4.2 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 on page 31, 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 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: 1) we replicate previous findings that, on average, schools with higher-quality contexts experience lower turnover, 2) we show that improvements in the organizational context within a school over time are associated with corresponding decreases in teacher turnover, 3) 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 4) 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-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 are each associated with a 1.9 percentage point decrease in turnover when each measure is included separately. When all measures are included simultaneously in the model, these estimates are slightly attenuated, but 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. T-tests confirm that all of these differences are statistically significant.

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. A one standard deviation increase in Leadership at the 10th percentile is associated with a 1.7 percentage point decrease in teacher turnover, while the same marginal effect at the 90th percentile falls to 1.4 percentage points. Given the average turnover rate among middle schools 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 0.8, 0.8, and 0.6 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 in 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.

4.3 Student Achievement Growth

In Table 6 on page 32, 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 and our full set of covariates). Our results reveal four key findings: 1) we replicate and extend previous research finding that schools with higher-quality school contexts have students who experience larger achievement gains, 2) we show that improvements in the school context within a school over time are associated with corresponding increases in student achievement gains, 3) we find that the relationship between the school context and student achievement gains at a school is stronger in mathematics than in ELA, and 4) 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 0.056 and 0.032 standard deviation (SD) increase in mathematics and ELA, respectively. A one standard deviation increase in Expectations is associated with a 0.033 and 0.020 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 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 0.032 and 0.013 SD increase in student achievement in mathematics and ELA, respectively, when controlling for all other school context measures. An increase in 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 relationship between Leadership and student gains is no longer statistically significant in either subject when we focus on changes within school over time. 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.

The magnitudes of these within-school relationships are small but meaningful when placed in context. Our results are estimated from models that condition 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 0.053 SD in mathematics and 0.019 SD in ELA. Our estimates are substantially smaller than the size of teacher effects on student achievement, which are approximately 0.15 SD in mathematics and 0.10 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; Garet et al., 2011; Yoon et. al., 2007). Ronfeldt, Loeb, and Wykcoff (2013) find that a one standard deviation decrease in teacher turnover corresponds with an even smaller 0.02 and 0.01 SD increase in student achievement in mathematics and ELA, respectively.

4.4 Variation in Teachers' Perspectives on the School Context

We extend our primary analyses above by exploring whether the degree of agreement among teachers about their school contexts is itself a measure of the functionality of

an organization. For example, imagine two schools where, on average, teachers rate their Leadership as middling. In one of these schools teachers generally agree the Leadership is of mediocre quality, while in the other school some teachers are very unsatisfied but others are highly impressed. Do the differences in agreement between teachers capture important information? We explore this by calculating the variance of teachers' estimated factor scores within each school in a given year, expressed as the standard deviation of each score. We then standardize these variability measures in order to place them on a comparable scale as the average school context measures. We find a moderate to strong negative correlation (-.43 to -.56) between the mean and variance of a given school context measure in a school-year dataset. This is the consequence of a ceiling effect where schools with higher average ratings have compressed distributions compared to schools with lower average ratings.

We explore whether the degree of agreement among teachers about their school contexts is predictive of turnover and achievement, conditional on average school context ratings, while recognizing the limitations of this analysis due to the mechanical relationship between our measures of central tendency and variability. Across the twelve models we fit (4 school context dimension * 3 outcomes), we only find that variation in teachers' perceptions of the Leadership in their schools is a significant predictor of teacher turnover ($b=0.009$, $p=0.001$ [predicted marginal effect at 50th percentile]). The small magnitude of this finding combined with the null results across all other models suggests that the variability in teachers' perceptions does not contain substantial information that is not captured by school-wide means.

5. Threats to Validity

The findings presented above show clearly that, among NYC middle schools, improvements in a school's context over time are associated with corresponding decreases in teacher turnover and increases in student achievement gains. Our panel data allow us to control for all time-invariant differences across schools and common shocks across time, thus removing two of the largest potential threats to interpreting the observed relationship between the quality of school organizational contexts, teacher turnover, and student achievement as causal. 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.

5.1 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 in section 4.1 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 primary driven by random noise.⁷

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-squared values from models in which we regress a given school context dimension on a full set of school indicator variables in a school-year dataset. These estimates are reported in Column 1 of Table 7 (page 33). 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-squared 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.

5.2 Common Source Bias

A central concern is the endogenous relationship between teachers' responses on the NYC DOE school survey and their decisions 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 onto the dominant items from the Expectations factor and nine questions that map onto the Safety factor.⁸ Measures of students' and teachers' perceptions of Expectations in their school have a correlation of 0.22, while perceptions of Safety have a correlation of 0.68 in a school-year dataset. Results using these alternative student-based measures are consistent with our main findings and of comparable magnitudes. As we show in Table 8 on page 33, 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.

5.3 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 on page 34, 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 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 2008-11 and 2009-12 to ensure these results are not due to the restricted four-year panel datasets 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.

5.4 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, & Taber's (2005) classic analysis of selection bias in Catholic school effect estimates. 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 on page 34, 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.

The internal validity of our results is also strengthened by their close alignment to findings 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 their 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, Relationships has a strong association with teachers' career decisions, while having no detectable relationship with student achievement gains when controlling for other school context dimensions. If our results were driven by self-report bias, non-response 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.

6. Conclusion

6.1 Implications for Policy and Practice

This paper 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 that could explain away an underlying causal relationship. These results further illustrate the importance of both individual and organizational effectiveness when designing reforms aimed at raising student achievement.

The longstanding challenges of closing achievement gaps and turning around chronically under-performing schools demand both individual and organizational solutions (Johnson, 2012). Recent scholarship and federal education policy have generated considerable momentum behind reform efforts aimed at remaking teacher evaluation systems and placing an effective teacher in every classroom (Steinberg & Donaldson, 2015). 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 (e.g., 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 teachers who have the misfortune of trying to deliver high-quality instruction and improve their craft amidst organizational dysfunction, continually reshuffling the staff in search of teachers who can be successful in spite of organizational limitations 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, sources of data on school contexts exist, and schools and districts are increasingly using them. The New Teacher Center's Teaching, Empowering, Leading, and Learning Survey (TELLS) and the University of Chicago Consortium on School Research's 5 Essential Supports Survey are two examples of widely used surveys designed to capture rich data on school organizational contexts. In addition, many local and state education agencies now administer student surveys as part of teachers' evaluations; these surveys could also be used to gather important information about aspects of the school environment, such as the level of safety or the rigor of schools' academic expectations for students.

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 districts producing customized school reports that describe average levels and trends in teachers' perceptions of schools' organizational contexts, as well as relative comparisons with similar schools. For example, the NYC DOE is currently revising its School Survey to capture a broader range of information about school contexts with improved measurement properties. The district is also developing customized school reports, which will be used to inform school leaders' improvement plans. Ultimately, the DOE hopes to incorporate these reports into system-wide school improvement efforts that would involve pairing schools that have strong organizational contexts with those that are struggling.

District leaders and principals could also use school context reports to identify and target efforts aimed at strengthening specific organizational weaknesses in their schools. Prior work has found that principals can strengthen organizational contexts by establishing school-wide behavior policies and systems, by developing opportunities for teacher collaboration and meaningful feedback, and by articulating high expectations for students paired with a range of student support services (Kraft, et al., 2015). While principals are undoubtedly critical to this work, districts should also consider the role that central administrators and school support organizations can play in both helping assess school contexts and in supporting principals and school leadership teams in strengthening the organizational characteristics of schools. For example, districts might use data and analysis on school contexts to inform the curriculum for principals' in-service training.

Another potential extension of these findings—and a potential application of data from teacher and student surveys—would be to incorporate evidence of schools' organizational context into principals' evaluations. The principalship 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). Incorporating data on organizational contexts into principal evaluation systems could provide school leaders with important feedback on their organizational leadership and simultaneously hold them accountable for promoting and sustaining the types of school environments that we find are related to teacher retention and student achievement. By extension, districts might also use data on school contexts to inform strategic staffing decisions, potentially offering incentives to principals with proven abilities to drive organizational change to accept placements at schools with organizational deficiencies.

Pursuing any of these organizational reforms as a singular reform strategy will not be sufficient to close achievement gaps or turn around failing schools, given the moderate magnitude of our results. However, such initiatives can and should be a meaningful part of larger reform efforts to increase teacher retention and student achievement.

6.2 Implications for Future Research

We see several important directions for future research on school organizational contexts. The process of developing instruments that can capture a comprehensive set of reliable school context measures is still in its initial stages. Researchers and practitioners continue to invest in efforts to enhance the precision and conceptual

clarity of these measures. We analyzed four school context dimensions that we viewed as being more immediately under the direct control of administrators and teachers. However, other school context dimensions we excluded from our analyses or that were not captured by the NYC DOE School Survey in the years we explored may be equally as important to teachers' and students' experiences in school. Rigorous qualitative studies suggest that school context features such as teacher leadership and shared governance, as well as time for common planning and collaboration, may also be high-leverage organizational practices. Researchers need to synthesize findings from qualitative and quantitative studies of schools as organizations to develop and then test a more comprehensive set of measures. Educators will also benefit from additional studies that examine whether and why the relationships between school contexts and teacher and student outcomes differ across K-12 school levels.

Our findings also highlight the need for more in-depth qualitative research that examines how and 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 network dynamics. How do administrators successfully lead collective action in their school buildings to strengthen organizational practices? For example, what do successful administrators do to ensure that behavioral norms are applied consistently and high expectations are always upheld? Case studies of schools in which efforts to strengthen organizational capacity are unsuccessful will also provide important insights about obstacles that block collective action. Such research can be the basis for developing the features of successful organizational reform strategies.

Advancing our understanding of the potential for organizational reforms to drive student learning gains will require researcher-practitioner partnerships to experimentally test the efficacy of interventions that target specific school context dimensions. For example, experiments to improve the organizational capacity of firms in the private sector have produced compelling evidence of the large causal effect of productive organizational practices (e.g., Bloom, et al., 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 the ongoing efforts to strengthen

teacher effectiveness and create schools where all students are supported to reach high academic standards.

Tables and Figures

Table 1: Middle School Teacher Response Rates to the NYC DOE 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

Notes: Cells report percentages.

Table 2: New York City Middle School Teacher Characteristics, 2008-2012

	Mean
Female	0.696
Race/Ethnicity	
Asian	0.053
Latino	0.224
Black	0.129
White	0.579
Teaching Experience	
Novice	0.053
2-3 years experience	0.119
4-10 years experience	0.568
11-20 years experience	0.128
> 20 years experience	0.131
Highest Level of Education	
Bachelor's Degree	0.149
Master's Degree	0.428
Master's Degree plus 30 Credits	0.420
Turnover Status^a	
Movers	0.062
Leavers	0.089
Total Turnover	0.151

Notes: Sample includes 53,991 teacher-year observations and 16,404 unique teachers. ^aTotal turnover is the sum of movers (teachers who remain active classroom teachers but transfer to another NYC DOE school in the subsequent year); and leavers (i.e., teachers who leave their school and are not teaching in any other school in the NYC public school system the next year).

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

	Mean
Female	0.498
White	0.153
Black	0.269
Asian	0.177
Latino	0.394
Special Education	0.121
Free/Reduced Price Lunch Eligible	0.645
English Language Learner	0.075
Mathematics Achievement (z-score)	0.029
ELA Achievement (z-score)	-0.004

Notes: Sample includes 645,995 student-year observations and 334,050 unique students.

Table 4: The Correlations between School Context Measures and Average Student Characteristics

	Leadership & Professional Development	Academic Expectations	Teacher Relationships & Collaboration	Safety & Order
% Free/Reduced Price Lunch Eligible	0.062*	-0.222*	0.001	-0.283*
% Special Education	0.033	-0.178*	0.067*	-0.361*
% English Language Learner	0.054	-0.173*	0.051	-0.085*
% Black	0.000	-0.084*	-0.075*	-0.342*
% Latino	0.006	-0.216*	0.103*	-0.091*
% Teacher Turnover	-0.244*	-0.356*	0.062*	-0.296*
Mathematics Achievement (z-score)	0.027	0.409*	-0.010	0.610*
ELA Achievement (z-score)	0.028	0.406*	-0.025	0.526*

Notes: * $p < 0.05$. $n = 1150$. Values indicate correlations between school-year means. Mathematics and ELA Achievement are the school-year average student test scores on the New York State exams standardized within grade, subject and year.

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					
	10 th	50 th	90 th	10 th	50 th	90 th
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
Panel A: Coefficients estimated for each factor separately						
Leadership & Professional Development	-.021*** (0.003)	-.019*** (0.002)	-.017*** (0.002)	-.02*** (0.003)	-.018*** (0.003)	-.016*** (0.002)
Academic Expectations	-.022*** (0.003)	-.019*** (0.003)	-.017*** (0.002)	-.011** (0.004)	-.010** (0.003)	-.010** (0.003)
Teacher Relationships & Collaboration	-.008** (0.003)	-.008** (0.003)	-.007** (0.002)	-.008* (0.003)	-.007* (0.003)	-.007* (0.003)
Safety & Order	-.021*** (0.003)	-.019*** (0.003)	-.017*** (0.002)	-.014*** (0.004)	-.013*** (0.004)	-.012*** (0.003)
Panel B: Coefficients estimated for each factor simultaneously						
Leadership & Professional Development	-.014*** (0.003)	-.013*** (0.002)	-.012*** (0.002)	-.017*** (0.003)	-.016*** (0.003)	-.014*** (0.002)
Academic Expectations	-.011*** (0.003)	-.011*** (0.003)	-.01*** (0.002)	-.006* (0.003)	-.006* (0.003)	-.006* (0.003)
Teacher Relationships & Collaboration	-.007** (0.003)	-.007** (0.002)	-.006** (0.002)	-.008** (0.003)	-.008** (0.003)	-.007** (0.003)
Safety & Order	-.013*** (0.003)	-.012*** (0.003)	-.011*** (0.002)	-.009* (0.004)	-.008* (0.004)	-.008* (0.003)
School Fixed Effects		-			Yes	
n (teachers)		53,991			53,991	

Notes: ***p<0.001, **p<0.01, *p<0.05. 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.

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 & Professional Development	.014** (0.004)	.013** (0.005)	.008* (0.003)	0.004 (0.004)
Academic Expectations	.033*** (0.005)	.013** (0.005)	.020*** (0.004)	0.002 (0.004)
Teacher Relationships & Collaboration	0.007 (0.004)	.009+ (0.005)	-0.002 (0.004)	0.003 (0.004)
Safety & Order	.056*** (0.005)	.032*** (0.005)	.025*** (0.003)	.013** (0.004)
Panel B: Coefficients estimated for each factor simultaneously				
Leadership & Professional Development	-0.005 (0.004)	0.006 (0.005)	-0.002 (0.004)	0.002 (0.004)
Academic Expectations	0.026*** (0.005)	0.009* (0.005)	0.018*** (0.004)	0.000 (0.004)
Teacher Relationships & Collaboration	0.001 (0.004)	0.008+ (0.005)	-0.006 (0.004)	0.004 (0.004)
Safety & Order	0.052*** (0.005)	0.030*** (0.006)	0.022*** (0.004)	0.013** (0.004)
School Fixed Effects	-	Yes	-	Yes
n (students)	641,515	641,515	626,913	626,913

Notes: ***p<0.001, **p<0.01, *p<0.05. 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 ELA, 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).

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

	R-Squared with School Fixed Effects ^a	R-Squared with School-Fixed Effects and School-Specific Linear-Trends ^b	Proportion of Within-School Variance Explained by School-Specific Linear Trends ^c
Leadership & Professional Development	0.615	0.778	0.423
Academic Expectations	0.620	0.781	0.423
Teacher Relationships & Collaboration	0.640	0.828	0.522
Safety & Order	0.789	0.888	0.472

Notes: a. This column contains the R-squared values of models that predict each measure of the school context using a full set of school fixed effects. b. This column contains the R-squared values of models that predict each measure of the school context using a full set of school fixed effects and school-specific linear trends. c. This column shows the product of the following calculation: (R-squared model 2 - R-squared model 1)/(1-R-squared model 1). Estimates in this column are also identical to models where the residuals from the specifications in the first column of numbers [demeaned school context measures] are regressed on school fixed effects and school-specific slopes.

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*** (0.003)	-.012*** (0.004)	.034*** (0.005)	.025*** (0.005)	.011** (0.004)	0.007 (0.004)
Safety & Order	-.016*** (0.003)	-.012** (0.004)	.064*** (0.006)	.024*** (0.006)	.040*** (0.005)	0.007 (0.005)
Panel B: Coefficients estimated for each factor simultaneously						
Academic Expectations	-.012*** (0.003)	-.010* (0.004)	0.009+ (0.005)	0.021*** (0.006)	-0.007 (0.004)	0.005 (0.005)
Safety & Order	-.010** (0.003)	-0.008 (0.005)	0.059*** (0.006)	0.016* (0.007)	0.043*** (0.005)	0.005 (0.006)
School Fixed Effects	-	Yes	-	Yes	-	Yes
n	53,835	53,835	639,750	639,750	625,152	625,152

Notes: ***p<0.001, **p<0.01, *p<0.05. 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.

Table 9: Exploratory Tests for Reverse Causality

	n	Outcomes (time t+1)			
		Leadership & Professional Development	Academic Expectations	Teacher Relationships & Collaboration	Safety & Order
Predictors (time t)					
Turnover	42,416	-0.002 (0.007)	-0.010 (0.006)	0.002 (0.007)	0.001 (0.005)
Residualized Gain Score: Mathematics	502,746	-0.001 (0.003)	0.002 (0.003)	-0.002 (0.003)	0.002 (0.003)
Residualized Gain Score: ELA	491,383	0.007*** (0.002)	0.003 (0.002)	0.004* (0.002)	0.002 (0.002)
School Fixed Effects		Yes	Yes	Yes	Yes

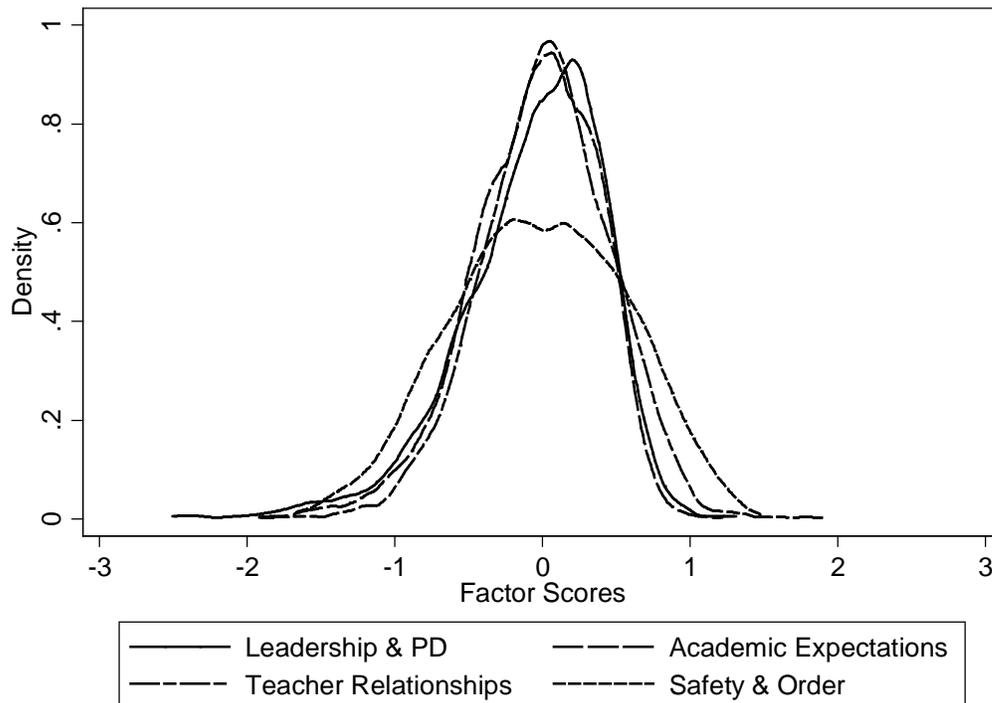
Notes: ***p<0.001, **p<0.01, *p<0.05. Each cell represents results from a separate regression with standard errors clustered by school-year reported in parentheses. Estimates are derived from OLS regression models. Residualized gain scores are the residuals from a regression model of students' achievement scores in a given subject regressed on cubic functions of prior test scores in mathematics and ELA.

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 & Professional Development	-.016*** 0.003	-.018*** 0.003	0.006 (0.005)	0.008+ (0.005)	0.002 (0.004)	0.004 (0.004)
Academic Expectations	-.006* 0.003	-.006+ 0.003	0.009* (0.005)	0.009+ (0.005)	0.000 (0.004)	0.002 (0.004)
Teacher Relationships & Collaboration	-.008** 0.003	-.007* 0.003	0.008+ (0.005)	0.010* (0.005)	0.004 (0.004)	0.004 (0.004)
Safety & Order	-.008* 0.004	-.008* 0.004	0.030*** (0.006)	0.028*** (0.006)	0.013** (0.004)	0.013** (0.005)
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes		Yes		Yes	
n	53,991	53,991	641,9515	641,9515	626,913	626,913

Notes: ***p<0.001, **p<0.01, *p<0.05. 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.

Figure 1: Probability Density Functions of our Four Unstandardized School Context Measures in a School-By-Year Dataset (n=1,150)



Notes: The underlying unit of measurement is teachers' responses to the School Survey on a four point Likert scale which were standardized in a teacher-year dataset.

Notes

- ¹ Allensworth, Ponisciak, & Mazzeo (2009) have survey data from two waves of the University of Chicago Consortium on School Research teacher survey. However, the authors only report findings from analyses that estimate the relationship between measures of the school context from the spring of 2005 to teacher turnover between 2005-06 and 2006-07. Bryk and his colleagues (2010) also analyze multiple survey waves collected by the Chicago Consortium but focus primarily on explaining differences in achievement trends across schools with one year of baseline survey data.
- ² 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 2007-08 and increased steadily to 58% by 2011-12.
- ³ The DOE's reporting categories for the survey were Academic Expectations, Communication, Engagement and Safety & Respect.
- ⁴ 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 multi-collinearity. 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 0.88 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.
- ⁵ 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).
- ⁶ We reverse code items with negative valence so higher ratings are associated with safer and more orderly environments.
- ⁷ We thank Sean Corcoran for his helpful suggestions that motivated these analyses.
- ⁸ 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: University of Chicago Consortium on Chicago School Research.
- 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.
- Balfanz, R. (2009). *Putting Middle Grades Students on the Graduation Path: A Policy and Practice Brief*. Ohio: National Middle School Association.
- 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.
- Barnes, G., Crowe, E., & Schaefer, B. (2007). *The Cost of Teacher Turnover in Five School Districts: A Pilot Study*. Washington, DC: National Commission on Teaching and America's Future.
- Betebenner, D.W. (2009). "Norm and Criterion-Referenced Student Growth." *Educational Measurement: Issues and Practice*, 28, 42-51.
- Betebenner, D.W., (2011). *A Technical Overview of the Student Growth Percentile Methodology: Student Growth Percentiles and Percentile Growth Projections/ Trajectories*. Dover, New Hampshire: The National Center for the Improvement of Educational Assessment.
- Birkeland, S.E. & Curtis, R. (2006). *Ensuring the Support and Development of New Teachers in the Boston Public Schools: A Proposal to Improve Teacher Quality and Retention*. Boston, MA: Boston Public Schools.
- 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., Easton, J. Q., & Luppescu, S. (2010). *Organizing Schools for Improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.
- 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., Uekawa, K., Falk, A., Bloom, H. S., Doolittle, F., Zhu, P., Silverberg, M., & Szejnberg, L. (2008). *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement*. Washington, D. C.: U.S. Department of Education, National Center for Education Statistics.
- Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2011). *Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation*. Washington, D. C.: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance.
- 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. and Oldham, G. R. (1980). *Work Redesign*. Reading, MA: Addison-Wesley.
- Hanushek, E. A. (2009). "Teacher Deselection." In D. Goldhaber & J.

- Hannaway, (Eds.), *Creating a New Teaching Profession* (pp. 165-180). Washington D.C.: Urban Institute Press.
- Hanushek, E. A., and Rivkin, S. G. (2010). "Generalizations About Using Value-Added Measures of Teacher Quality." *American Economic Review*, 100(2), 267-271.
- Harris, D., & Sass, T. (2011). "Teacher Training, Teacher Quality, and Student Achievement." *Journal of Public Economics*, 95, 798-812.
- Honig, M. [Ed.] (2006). *New Directions in Education Policy Implementation: Confronting Complexity*. Albany, NY: State University of New York Press.
- Ingersoll, R.M. (2001). "Teacher Turnover and Teacher Shortages: An Organizational Analysis." *American Educational Research Journal*, 38(3), 499-534.
- Johnson, S.M. (1990). *Teachers at Work: Achieving Success in Our Schools*. New York: Basic Books.
- Johnson, S.M. (2009). *How Best to Add Value? Striking a Balance Between the Individual and the Organization in School Reform*. Washington, D.C.: Economic Policy Institute.
- Johnson, S. M. (2012). "Having it Both Ways: Building the Capacity of Individual Teachers and Their Schools." *Harvard Educational Review*, 82(1), 107-122.
- 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.
- Jussim, L., & Harber, K. D. (2005). "Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies." *Personality and Social Psychology Review*, 9(2), 131-155.
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. Research Paper. MET Project Research Paper, Bill & Melinda Gates Foundation.
- 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*. New York, NY: Simon & Schuster.
- Kennedy, 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?" Working Paper No. 33. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research

- 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.
- 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., & 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. steinhardt.nyu.edu/research_alliance/publications/ttp_synthesis
- Milanowski, A.T., & Odden, A.R. (2007). "A New Approach to the Cost of Teacher Turnover." Working Paper 13. Seattle, WA: School Finance Redesign Project, Center on Reinventing Public Education.
- Murdock, T. B., Anderman, L. H. and 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., Segeritz, M., Cole, R., Lent, J., McCormick, M., & Kemple, J. (2013). *NYC School Survey 2008-2010: Assessing the Reliability and Validity of Progress Report Measures*. New York: The Research Alliance for New York City Schools. steinhardt.nyu.edu/research_alliance/publications/nyc_school_survey
- National Commission on Teaching and America's Future [NCTAF]. (2007). *The Cost Of Teacher Turnover in Five School Districts: A Pilot Study*. Arlington, VA.
- Neild, R. C. and Balfanz, R. (2006). *Unfulfilled Promise: The Dimensions and Characteristics of Philadelphia's Dropout Crisis, 2000-2005*. Philadelphia, PA: Philadelphia Youth Network.
- Neild, R. C., Farley-Ripple, E. N., and Byrnes, V. (2009). "The Effect of Teacher Certification on Middle Grades Achievement in an Urban District." *Educational Policy*, 23(5), 732-760.
- Neild, R.C., Useem, E., and Farley, E. (2005). *Quest for Quality: Recruiting and Retaining Teachers in Philadelphia*. Philadelphia, PA: Research for Action.
- New York City Department of Education (2008). "Educator Guide. The New York City Progress Report. Elementary/Middle School." Retrieved from schools.nyc.gov/NR/exeres/32B0EF5C-2B4D-455F-9836-FA0020E1F5B3,frameless.htm?NRMODE=Published
- 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: Research Alliance for New York City Schools. steinhardt.nyu.edu/research_alliance/publications/ttp_thoughts_of_leaving
- Papay, J. P. (2012). "Refocusing the Debate: Assessing the Purposes and Tools of Teacher Evaluation." *Harvard Educational Review*, 82(1), 123-141.
- 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 ssrn.com/abstract=2607776
- Papay, J.P. & Kraft, M.A. (2015). "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement." *Journal of Public Economics*, 130, 105-119
- Podgursky, M., & Springer, M. G. (2007). "Credentials Versus Performance: Review of the Teacher Performance Pay Research." *Peabody Journal of Education*, 82(4), 551-573

- Rich, M. (2013). "In Raising Scores, 1 2 3 Is Easier than ABC." *The New York Times*. Retrieved from http://www.nytimes.com/2013/05/30/education/reading-gains-lag-improvements-in-math.html?_r=0
- Rockoff, J.E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94(2), 247-252.
- Roderick, M. (1994). "Grade Retention and School Dropout: Investigating the Association." *American Educational Research Journal*, 31(4), 729-759.
- Ronfeldt, M., Lankford, H., Loeb, S., & Wyckoff, J. (2013). "How Teacher Turnover Harms Student Achievement." *American Educational Research Journal*, 50(1), 4-36.
- 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: University of Chicago Consortium on Chicago School Research.
- Steinberg, M. P., & Donaldson, M. L. (2015). *The New Educational Accountability: Understanding the Landscape of Teacher Evaluation in the Post-NCLB Era. Education Finance and Policy*.
- 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.
- Weiss, E. M. (1999). "Perceived Workplace Conditions and First-Year Teachers' Morale, Career Choice Commitment, and Planned Retention: A Secondary Analysis." *Teaching and Teacher Education*, 15(8), 861-879.
- Wentzel, K. R. (2002). "Are Effective Teachers Like Good Parents? Interpersonal Predictors of School Adjustment in Early Adolescence." *Child Development*, 73(1), 287-301.
- West, M.R., Kraft, M.A., Fin, A. S., Martin, R., Duckworth, A., Gabrieli, C.F., & Gabrieli, J.D. (2015). Promise and Paradox: Measuring Students' Non-Cognitive Skills and the Impact of Schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148-170.
- Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Appendix Table A1: Raw Factor Loadings for School Context Dimensions

	Leadership & Prof. Dev.	Academic Expectations	Teacher Relationships & Collaboration	Safety & Order	% Agree or Strongly Agree
Unstandardized Factor Variance	7.02	5.92	4.45	3.69	na
Percent 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.	0.609	0.508	0.139	0.200	0.857
2) The principal at my school...encourages open communication on important school issues.	0.656	0.432	0.174	0.186	0.782
3) Curriculum, instruction, and assessment are aligned within and across the grade levels at this school.	0.442	0.577	0.227	0.161	0.876
4) The principal at my school...is an effective manager who makes the school run smoothly.	0.658	0.439	0.113	0.265	0.788
5) My school... has high expectations for all students.	0.315	0.729	0.197	0.260	0.889
6) My school... sets high standards for student work in their classes.	0.086	0.748	0.330	0.191	0.931
7) My school... has clear measures of progress for student achievement throughout the year.	0.356	0.739	0.228	0.179	0.897
8) My school... helps students develop challenging learning goals.	0.368	0.765	0.224	0.174	0.853
9) My school...helps students find the best ways to achieve their learning goals.	0.396	0.733	0.238	0.201	0.853
10) My school... offers a wide enough variety of courses to keep students engaged.	0.439	0.482	0.179	0.270	0.785
11) To what extent do you feel supported by: my principal?	0.693	0.332	0.151	0.214	0.820
12) To what extent do you feel supported by: other teachers at my school?	0.092	0.128	0.733	0.089	0.919
13) School leaders... provide time for collaboration among teachers.	0.548	0.363	0.350	0.135	0.874
14) Teachers in my school... respect colleagues who take the lead in school improvement efforts.	0.259	0.259	0.764	0.129	0.882
15) Teachers in my school... trust each other.	0.236	0.208	0.782	0.169	0.809
16) Teachers in my school... respect colleagues who are the most effective teachers.	0.202	0.213	0.809	0.132	0.889
17) School leaders... give me regular and helpful feedback about my teaching.	0.611	0.426	0.289	0.144	0.766
18) Teachers in my school... work together on teams to improve their instructional practice.	0.287	0.386	0.642	0.113	0.865
19) Teachers in my school... use student achievement data to improve instructional decisions.	0.327	0.485	0.474	0.104	0.932
20) Overall, my professional development experiences this school year have... provided me with teaching strategies to better meet the needs of my students.	0.727	0.239	0.301	0.092	0.773

Survey Items	Leadership & Prof. Dev	Academic Expectations	Teacher Relationships & Collaboration	Safety & Order	% Agree Or Strongly Agree
	Raw Factor Loadings				
21) I have sufficient materials to teach my class(es), including: books, audio/visual equipment, maps, and/or calculators.	0.558	0.179	0.237	0.265	0.768
22) I received helpful training on the use of student achievement data to improve teaching and learning this year.	0.665	0.310	0.281	0.117	0.784
23) Overall, my professional development experiences this school year have... provided me with content support in my subject area.	0.732	0.221	0.291	0.084	0.740
24) Teachers and school leaders in my school use information from parents to improve instructional practices and meet student learning needs.	0.487	0.413	0.306	0.251	0.753
25) My school communicates effectively with parents when students misbehave.	0.520	0.335	0.237	0.395	0.831
26) Order and discipline are maintained at my school.	0.497	0.315	0.179	0.583	0.750
27) I can get the help I need at my school to address student behavior and discipline problems.	0.556	0.282	0.209	0.523	0.771
28) I am safe at my school.	0.414	0.247	0.196	0.619	0.909
29) Crime and violence are a problem in my school.*	0.092	0.150	0.082	0.794	0.812
30) Students in my school are often threatened or bullied.*	0.159	0.213	0.121	0.740	0.660
31) Adults at my school are often disrespectful to students.*	-0.069	0.280	0.256	0.527	0.892
32) There is a person or a program in my school that helps students resolve conflicts.	0.392	0.206	0.211	0.346	0.884
33) My school is kept clean.	0.331	0.165	0.202	0.413	0.868

Notes: n=31,699 survey responses. Factor loadings above 0.5 are highlighted in bold text. Loadings produced using varimax rotation following principal component analysis. Resulting factors are orthogonal (pairwise uncorrelated) across teacher-years.

* 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.