Educational Evaluation and Policy Analysis
December 2014, Vol. 36, No. 4, pp. 476–500
DOI: 10.3102/0162373713519496

Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience

Matthew A. Kraft
John P. Papay
Brown University

Although wide variation in teacher effectiveness is well established, much less is known about differences in teacher improvement over time. We document that average returns to teaching experience mask large variation across individual teachers and across groups of teachers working in different schools. We examine the role of school context in explaining these differences using a measure of the professional environment constructed from teachers responses to state-wide surveys. Our analyses show that teachers working in more supportive professional environments improve their effectiveness more over time than teachers working in less supportive contexts. On average, teachers working in schools at the 75th percentile of professional environment ratings improved 38% more than teachers in schools at the 25th percentile after 10 years.

Keywords: *teacher development, returns to experience, school context, professional environment, professional development*

Research documenting the primary importance of effective teachers has shaped education policy dramatically in the last decade, resulting in a broad range of reforms targeted at increasing teacher quality. Federal, state, and local policy initiatives have sought to attract and select highly qualified candidates, evaluate their performance, and reward and retain those teachers judged to be most effective. This narrow focus on individuals discounts the important role of the organizational context in shaping teachers' career decisions and facilitating their success with students. In response, some scholars have argued that reforms targeting teacher effectiveness would achieve greater success by also working to improve the organizational context in schools (Johnson, 2009; Kennedy, 2010).

Mounting evidence suggests that the school context in which teaching and learning occurs can have important consequences for teachers and students. Recent studies document the influence of school contexts on teachers' career decisions, teacher effectiveness, and student achievement (Boyd et al., 2011; Johnson, Kraft, & Papay, 2012; Ladd, 2011; Loeb, Darling-Hammond, & Luczak, 2005). These studies capitalize on new measures of the school context constructed from student, teacher, and principal responses to district and state-wide surveys. We build on this work by investigating how the school context influences the degree to which teachers become more effective over time. We refer to these changes in effectiveness of individual teachers over time as "returns to teaching experience."

Studies on the returns to teaching experience find that, on average, teachers make rapid gains in effectiveness early in their careers, but that additional experience is associated with more modest improvements (e.g., Boyd, Lankford,

Loeb, Rockoff, & Wyckoff, 2008; Harris & Sass, 2011; Papay & Kraft, 2013; Rockoff, 2004; Wiswall, 2013). Using a rich administrative data set from Charlotte–Mecklenburg Schools (CMS), we demonstrate that this average profile masks considerable heterogeneity among teachers, as well as systematic differences in the average returns to experience among teachers in different schools. We also find that this variation in returns to teaching experience across schools is explained, in part, by differences in schools' professional environments. Teachers who work in more supportive environments become more effective at raising student achievement on standardized tests over time than do teachers who work in less supportive environments. These findings challenge common assumptions made by education policymakers and highlight the role of the organizational context in promoting or constraining teacher development.

In the following section, we review the literature on returns to teaching experience and describe the relationship between organizational contexts and worker productivity. We then describe our data and our measure of the professional environment. Next, we explain our empirical framework for measuring changes in effectiveness over a teacher's career, present our findings, and explore the sensitivity of these finding to our modeling assumptions. We further examine alternative explanations for the relationship we observe between returns to teaching experience and the professional environment in schools. Finally, we conclude with a discussion of our results and their policy implications.

Organization Theory and Productivity Improvement in Schools

Heterogeneity in the Returns to Teaching Experience

Studies find that novice and early-career teachers are less effective than their more experienced peers (Clotfelter, Ladd, & Vigdor, 2007; Rockoff, Jacob, Kane, & Staiger, 2011; Wayne & Youngs, 2003) and that, on average, individual teachers make rapid gains in effectiveness during the first several years on the job (Boyd, Lankford, et al., 2008; Rockoff, 2004). However, it remains less clear how much teachers continue to improve later in their careers (Harris & Sass, 2011; Papay & Kraft, 2013; Wiswall, 2013). Scholars hypothesize that these returns to teaching experience result from the acquisition of new human capital, including content knowledge, classroom management techniques, and methods of instructional delivery. Teachers learn how to create and modify instructional materials (Kaufman, Johnson, Kardos, Liu, & Peske, 2002) and better meet the diverse instructional needs of students (Johnson & Birkeland, 2003) as they gain experience on the job.

Clearly, though, these average patterns obscure potential heterogeneity in returns to teaching experience. Just as there are large differences in the effectiveness of teachers at any given level of experience, there are differences in the rate at which individual teachers improve throughout their careers. Kane, Rockoff, and Staiger (2008) find initial evidence of this heterogeneity in New York, as alternatively certified and uncertified teachers improve their effectiveness over time more rapidly than their traditionally certified counterparts. Early evidence on an urban teacher residency program also suggests that program graduates underperform all other novice teachers but improve rapidly over time and eventually outperform their peers after several years in the classroom (Papay, West, Fullerton, & Kane, 2012). Two recent studies suggest that differential returns to experience are related to school characteristics. Loeb, Kalogrides, and Beteille (2012) document how, on average, teachers improve at faster rates in schools with higher value-added scores. Sass, Hannaway, Xu, Figlio, and Feng (2012) find faster improvement among teachers at schools with fewer low-income students.

School Contexts and Teacher Development

That teachers might improve at different rates in different types of schools is not surprising: For more than a century, scholars of organizational behavior have attempted to explain differences in individual workers' productivity and skill development across work environments. They have developed a rich set of theories to explain how organizational structures, practices, and culture affect the productivity of workers (Hackman & Oldham, 1980; Kanter, 1984). In-depth qualitative studies of schools as workplaces illustrate how

organizational structures can facilitate or limit on-the-job learning for teachers (Johnson, 1990; Lortie, 1975). Together, these organizational theories and qualitative studies predict that school environments where teachers collaborate frequently, receive meaningful feedback about their instructional practices, and are recognized for their efforts will promote teacher improvement at faster rates than schools where such practices are absent.

A growing body of literature on the organizational context in schools has begun to bear out these predictions. Both theory and empirical evidence point to several specific elements of the school organizational context that, when practiced successfully throughout a school, can promote teacher improvement. Principals play a key role in supporting professional growth among teachers by serving as instructional leaders who provide targeted feedback and facilitate opportunities for teachers to reflect on their practice (Blase & Blase, 1999; May & Supovitz, 2011; Waters, Marzano, & McNulty, 2003). A principal's ability to lead effectively and support teachers' practice stands out as a critical influence on teachers' decisions to remain at their school (Boyd et al., 2011; Grissom, 2011).

Several studies find that measures of the social context of work, including principal leadership and peer collaboration, relate to gains in student achievement. Ladd (2009) finds that the quality of school leadership and the availability of common planning time predict school effectiveness, as measured by contributions to student achievement. In a similar study using data from Massachusetts, we find that stronger principal leadership, relationships among colleagues, and positive school culture predict higher median student achievement growth among schools (Johnson et al., 2012). Jackson and Bruegmann (2009) find that teachers, especially novices, improve their ability to raise standardized tests scores when they work in a school with more effective grade-level colleagues. Furthermore, evidence shows that social networks among teachers, particularly those with high levels of expertise and high-depth substantive interactions, enable investments in instructional improvement to be sustained over time (Coburn, Russell, Kaufman, & Stein, 2012).

Over a decade of research by the Consortium on Chicago School Research (CCSR) confirms these findings. Bryk and his colleagues find that for schools to be strong learning environments for students and teachers, adults must work to create a culture of mutual trust and respect (Bryk & Schneider, 2002; Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). They document the fundamental roles of school culture and order and safety in creating an environment where teachers are willing and able to focus on instruction. The large achievement gaps associated with measures of school safety in Chicago schools illustrate the value of environments where teachers and students are able to concentrate on teaching and learning (Steinberg, Allensworth, & Johnson, 2011).

The ways in which schools tailor and implement professional development and evaluation also shape teachers' opportunities for on-the-job learning. Over the past several decades, a growing consensus has emerged around the characteristics of effective professional development programs. Studies find that professional development is most effective when it provides teachers active learning opportunities that are intensive, focused on discrete skills, aligned with curriculum and assessments, and applied in context (Correnti, 2007; Desimone, 2009; Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Wayne, Yoon, Zhu, Cronen, & Garet, 2008). Many programs do not meet these criteria and have largely been found to be ineffective when implemented at-scale (Garet et al., 2008; Glazerman et al., 2008; Jacob & Lefgren, 2004). However, experimental evaluations of programs that do, such as particular literacy coaching models, show measurable improvements in teachers' instructional practice and students' performance on standardized assessments (Matsumura, Garnier, & Resnick, 2010; Neuman & Cunningham, 2009). Allen, Pianta, Gregory, Mikami, and Lun (2011) find that teachers who were assigned randomly to participate in a program that used individualized coaching to improve teacher–student interactions were more effective at raising student test scores in the following year. Furthermore, teacher evaluation can also contribute to such improvement. Taylor and Tyler (2012) find that participating in a rigorous teacher-evaluation program promoted large and sustained improvements in the effectiveness of mid-career teachers.

Together, these studies suggest that a collection of specific elements of the school context can play an important role in facilitating improvements in teacher effectiveness. Here, we examine this relationship directly. Specifically, we pose three primary research questions:

Research Question 1: Do the returns to teaching experience differ across individual teachers?

Research Question 2: Do the average returns to teaching experience differ across schools?

Research Question 3: Do teachers in schools with more supportive professional environments improve more over time than their peers in less supportive environments?

Research Design

Site and Sample

We study teachers and schools in CMS, an urban district in North Carolina that is the 18th largest public school district in the nation. CMS serves over 141,000 students across 174 schools and employs over 9,000 teachers. Teachers in CMS are largely representative of U.S. teachers as a whole. More than 82% of teachers are female, 64% are White, and 32% are African American. Thirty-four percent of teachers hold a master's degree, and teachers earn, on average, US$42,320 annually. In recent years, the district has received national recognition, including the 2011 Broad Prize for Urban Education.

We use a comprehensive administrative data set from 2000–2001 through 2009–2010. These data contain test records for state end-of-grade exams in mathematics and reading in third through eighth grade as well as demographic characteristics, student enrollment records, and teacher employment histories. We link student achievement data to teachers using a course enrollment file that contains both teacher and school IDs. Similar to past research, preliminary analyses revealed both larger average returns to teaching experience and substantially greater individual variation in mathematics than in reading (Boyd, Lankford, et al., 2008; Harris & Sass, 2011). This led us to concentrate on returns to experience as measured by teachers' contributions to students' mathematics achievement.

We combine these data with teachers' responses on the *North Carolina Teacher Working Conditions* Survey, which was administered in 2006, 2008, and 2010. This 100-plus item survey, developed by Eric Hirsch of the New Teacher Center, solicits teachers' opinions on a broad range of questions about the social, cultural, and physical environment in schools. These survey data present new opportunities to measure elements of the work context that play a central role in shaping teachers' experiences, but that are much more difficult to quantify than indices of traditional working conditions such as school resources and physical infrastructure. Survey response rates in the district increased with each administration from 46%, to 67%, to 77%. The survey contains identifying information on the schools where teachers work, but not unique IDs for teachers. Thus, we merge these survey records to our administrative data using unique school identifiers.

Our analytic sample consists of all students who can be linked to their mathematics teachers in fourth through eighth grades, the grades in which the necessary baseline and outcome testing data are available. This includes more than 280,000 student-year observations and 3,145 unique teachers.[1]

Measures

Our primary outcome consists of students' scaled scores on their end-of-grade examinations in mathematics, standardized within each grade and year ($\mu = 0$, $\sigma = 1$). Although test scores do not capture the full contribution that teachers make to children's intellectual and emotional development, we proceed with this narrow measure because it enables us to quantify one aspect of teacher productivity.

Our primary question predictor is the interaction of teaching experience, *EXPER*, and an overall measure of the professional environment in schools, *PROF_ENV*. We measure a teacher's level of experience using her step on the state salary scale. Because teachers receive salary increases for each year of experience they accrue, this provides a reasonable measure of actual on-the-job experience.

Because we examine the within-teacher returns to experience (i.e., we use teacher-fixed effects), we must make a methodological

assumption to fit our models. The reason is that teachers with standard career patterns gain an additional year of experience with every calendar year. In other words, teachers who start in the district in the fall of 2001 will have 10 years of experience in the fall of 2011. Thus, within-teacher, we cannot separate the effect of differences in achievement across school years (e.g., from the introduction of a new curriculum) from the returns to teaching experience without making a methodological assumption (Murnane & Phillips, 1981). The nature of this assumption can lead to substantial bias in the estimated returns to teaching experience (see Papay & Kraft, 2013, for a detailed discussion).

However, in this article, we focus on *differences* in the within-teacher returns to experience across individual teachers and schools, not the shape of the average returns-to-experience profile. Thus, the specific assumption we make is a second-order concern. As a result, we adopt Rockoff's (2004) simple and widely used identifying assumption by censoring experience at 10 years.[2] This approach enables us to examine the returns to experience for early- to mid-career teachers. We test the sensitivity of our results to alternative identifying assumptions and find that they are unchanged.[3] In our main models, we code experience as a continuous predictor up to 10 years, while in supplementary models we use a set of indicator variables to reflect teacher experience.

We create our measure of the professional environment by drawing on both the theoretical and empirical literature concerning the work context in schools reviewed above. We first identified elements of the work context characterized in the literature as important for creating an environment that provides opportunities for teachers to improve their effectiveness. We then restricted our focus to those elements for which we could find supporting empirical evidence, and which were included as topics on the survey (see Johnson et al., 2012, for a detailed description of this process). These elements of the professional environment include the following:

- *ORDER and DISCIPLINE:* The extent to which the school is a safe environment where rules are consistently enforced and administrators assist teachers in their efforts to maintain an orderly classroom.
- *PEER COLLABORATION:* The extent to which teachers are able to collaborate to refine their teaching practices and work together to solve problems in the school.
- *PRINCIPAL LEADERSHIP:* The extent to which school leaders support teachers and address their concerns about school issues.
- *PROFESSIONAL DEVELOPMENT:* The extent to which the school provides sufficient time and resources for professional development and uses them in ways that enhance teachers' instructional abilities.
- *SCHOOL CULTURE:* The extent to which the school environment is characterized by mutual trust, respect, openness, and commitment to student achievement.
- *TEACHER EVALUATION:* The extent to which teacher evaluation provides meaningful feedback that helps teachers improve their instruction, and is conducted in an objective and consistent manner.

To measure these elements, we selected 24 items from the survey, all of which were administered with identical or very similar question stems and response scales across the 3 years (see Online Appendix A). A principal components analysis of all 24 items suggested strongly that teachers' responses represented a single unidimensional latent factor in each survey year.[4] Internal-consistency reliability estimates across all items exceeded .90 in each year. Consequently, we focused our analysis on a single composite measure of the professional environment. We created this composite for each teacher in each year by taking a weighted average of their responses to all 24 items, using weights from the first principal component. Decomposing the variance of this composite measure, we find that differences in professional environments across schools account for approximately 30% of the total variance in teachers' responses in each year.

We then create a school-level measure of the professional environment by averaging these composite scores at the school-year level. We restrict our school-year averages to those derived from 10 or more teacher survey responses in each year. To arrive at our preferred overall measure of the professional environment in a school, we take the average of these school-year values in 2006,

2008, and 2010 and standardize the result. Our preferred models include this time-invariant average teacher rating of the overall professional environment in a school, *PROF_ENV*.[5] Recognizing that some of the differences in the measure across years may be due to real changes in schools' professional environment, we conduct supplementary analyses that use a time-varying measure. Results from these models are quite consistent with our primary findings, although less precise because they are limited to 3 years of data.

Finally, we include a rich set of student, peer, and school-level covariates in our models to account for observed individual differences across students as well as the sorting of students and teachers across and within schools. Student-level measures include dichotomous indicators of gender, race, limited English proficiency, and special-education status. Peer-level measures include the means of all student-characteristic predictors, and prior-year achievement in mathematics and reading for each teacher-by-year combination. School-level measures mirror peer-level measures averaged at the school-by-year level and also include the percent of students eligible for free or reduced price lunch in each year.[6]

Data Analysis

We examine the relationship between teacher effectiveness and teacher experience using an education production function in which we model student achievement as a function of prior test scores, student and teacher demographics, and school characteristics (Kane & Staiger, 2008; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Todd & Wolpin, 2003). Following previous studies of returns to experience using multilevel cross-classified data, we adopt a covariate-adjusted model as our preferred specification, which we then modify to answer each of our research questions. Our baseline model is as follows:

$$A_{itjg} = \eta_g\left(f\left(A_{i,t-1}\right)\right) + \varphi\left(f\left(EXPER^*_{jt}\right)\right) + \gamma \boldsymbol{X}_{it} + \pi_{gt} + \delta_j + \varepsilon_{itjg} \quad (1)$$

The outcome of interest, A_{itjg}, is the end-of-year mathematics test score for student *i* in year *t* in grade *g* taught by teacher *j*.[7] We include cubic functions of prior-year achievement in both math and reading, and allow the relationship between prior and current achievement in math to differ across grade levels by interacting our linear measure of prior achievement with grade-level indicators.[8] The vector $\boldsymbol{X}_{it}$ represents the student, peer, and school-level covariates described above. We include grade-by-year fixed effects, π_{gt}, to control flexibly for average differences in achievement across grades and school years, such as the introduction of new policies in certain grades. We specify the average effect of experience as a quartic function. We present results below that demonstrate how a quartic polynomial approximates well a non-parametric specification of experience.

Including teacher-fixed effects, δ_j, in our models is critical because it isolates the within-teacher returns to teaching experience, thereby avoiding many of the selection biases that arise in cross-sectional comparisons of teachers with different experience levels. Models that omit teacher-fixed effects compare less experienced teachers to their more experienced peers. Instead, we explicitly compare teachers' effectiveness to their own effectiveness earlier their careers.

Estimating heterogeneity. We modify the baseline specification described above to examine the variability in returns to teaching experience across *individual* teachers and schools. Here, we are interested in the variance of these estimated returns to experience. As a result, we depart from the fixed-effect modeling approach described above and adopt a multilevel random intercepts and random-slopes framework that provides more robust, model-based variance estimates.[9] In the new model, we specify individual teacher effects as random (rather than fixed) intercepts, α_j, and allow each teacher's return to experience to deviate from the average profile by including a random slope for each teacher, β_j. In other words, we estimate the returns to teaching experience separately for each teacher and summarize the variation across these estimates by examining the variance of β_j. These additions result in the following generic multilevel model:

$$A_{itjs} = \eta_g\left(f\left(A_{i,t-1}\right)\right) + \varphi\left(f\left(EXPER^*_{jt}\right)\right) + \gamma \boldsymbol{X}_{it} + \pi_{gt} + \left[\alpha_j + \beta_j EXPER^*_{tj} + \varepsilon_{itjs}\right], \quad (2)$$

where $\begin{bmatrix}\alpha_j \\ \beta_j\end{bmatrix} \sim N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma^2_{\alpha_j} & \sigma_{\alpha_j\beta_j} \\ \sigma_{\beta_j\alpha_j} & \sigma^2_{\beta_j}\end{bmatrix}\right)$

Here, the structural part of the model remains quite similar to Equation 1.[10] Again, we model a common returns to experience profile as a quartic function of experience, $f\left(EXPER^{*}_{jt}\right)$, but we allow returns to experience to vary across individual teachers as linear deviations from this average curvilinear trend. Sensitivity analyses presented below demonstrate this approach fits our data well. The random coefficients, β_j, characterize these individual deviations from the average profile. If the variance of these random slopes, $\sigma^2_{\beta j}$, is statistically significant, it will suggest that there exists heterogeneity in returns to teaching experience across individual teachers. In other words, it will indicate that some teachers do improve more rapidly than others. We extend this framework to examine whether the average returns to teaching experience differ across schools. We add a random effect for schools, μ_s, and replace the teacher-specific random slopes for experience with school-specific random slopes, β_s.

$$A_{itjs} = \eta_g\left(f\left(A_{i,t-1}\right)\right) + \varphi\left(f\left(EXPER^{*}_{jt}\right)\right) + \gamma \boldsymbol{X}_{it} + \pi_{gt} + \left[\mu_s + \beta_s EXPER^{*}_{tj} + \alpha_j + \varepsilon_{itjs}\right], \quad (3)$$

$$\text{where } \begin{bmatrix}\mu_s \\ \beta_s\end{bmatrix} \sim N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma^2_{\mu_s} & \sigma_{\mu_s\beta_s} \\ \sigma_{\beta_s\mu_s} & \sigma^2_{\beta_s}\end{bmatrix}\right),$$

$$\alpha_j \sim N\left(0, \sigma^2_{\alpha_j}\right)$$

As before, estimates of the random slopes, β_s, capture the average deviation from the average returns to experience profile for all teachers in a given school. A statistically significant estimate of the population variance of these random slopes, $\sigma^2_{\beta_s}$, will suggest that there are systematic differences in the pace at which teachers in different schools improve over time.

Here, our focus is on quantifying the total variance in returns to experience across individuals or schools, rather than producing estimates for each individual teacher. As such, our approach allows us to obtain consistent, model-based estimates of the true population variance.[11] However, while this specification accounts for measurement and other error appropriately, it also imposes several strong assumptions. First, we have assumed that all random effects are normally distributed. Second, the model requires that the random effects (including teacher effects) are independent of the large set of covariates we include in the model. This assumption would be violated and could produce biased estimates of our parameters if, for example, more effective teachers tended to teach certain types of students. As a result, we return to the widely used fixed-effect modeling framework to relax these assumptions as well as to facilitate a more direct comparison of our results with related estimates from the prior literature.

Examining heterogeneity across professional environments. We conclude our analyses by exploring whether differences in the professional environment help to explain variation in returns to experience across schools. In other words, we seek to understand whether teachers in more supportive environments improve more rapidly than teachers in less supportive schools. We do this by adding our measure of the professional environment and its interaction with experience ($EXPER^{*}_{tj} \times PROF\_ENV_s$) to Model I. This specification allows us to answer our third research question by interpreting a single parameter of interest, β_2.

$$A_{itjs} = \phi_g\left(f\left(A_{i,t-1}\right)\right) + \varphi\left(f\left(EXPER^{*}_{jt}\right)\right) + \beta_1 PROF\_ENV_s + \beta_2\left(EXPER^{*}_{tj} \times PROF\_ENV_s\right) + \gamma \boldsymbol{X}_{it} + \delta_j + \eta_{gt} + \varepsilon_{itjs} \quad (4)$$

Estimates of β_2 capitalize on variation in the average returns to teaching experience of teachers across schools with different professional environments. In effect, we are comparing the within-teacher returns to experience of teachers in schools with more supportive professional environments to those of their peers in schools with less supportive environments. A positive and statistically significant estimate for β_2 then indicates that teachers become relatively more effective over time when teaching in schools with more supportive professional environments. As before, we estimate an average curvilinear return to experience using a quartic polynomial. We assume that differences across professional environments accelerate (or decelerate) this underlying pattern by the same amount per year over the first 10 years of their career.

In addition to these primary analyses, we also test the robustness of our modeling approaches and explore a variety of alternative explanations for our findings. We model differences in returns to experiences across individuals, schools, and professional environments using polynomial and non-parametric functional forms. We re-estimate our models across different time periods and using alternative constructions of our professional environment measure to test for non-response bias, self-report bias, and reverse causality. We allow for differential returns to experience across a variety of teacher and student-body characteristics. Finally, we test for patterns of differential teacher retention related to rates of improvement and dynamic student sorting that might account for our findings. As discussed below, these analyses all confirm our central results.

Findings

We begin by presenting estimates of the average returns to experience in our sample as a relative benchmark for our estimates of the variation in returns to experience, as well as an illustration of the fit of our quartic function in experience. These estimates rely on a specific identifying assumption that teachers do not improve after 10 years. As we discuss in detail in a separate article (Papay & Kraft, 2013), we recommend that researchers who are concerned primarily with estimating the exact magnitude and functional form of the average returns to experience profile conduct parallel analyses using several alternative identifying assumptions. We find that the average returns to teaching experience after 10 years in our sample is almost 0.11 standard deviations (*SD*s) of the student test-score distribution based on estimates from Model I. In Figure 1, we illustrate the shape and magnitude the average returns to teaching experience profile, showing that a quartic function closely approximates the profile suggested by the flexible, but less precisely estimated, set of indicator variables. Importantly, the magnitudes of these returns to teaching experience are likely biased downward because we assume that teachers do not improve after 10 years.

The average returns to teaching experience after 10 years are large when compared with the overall distribution of teacher effectiveness in our sample estimated from Model II. Consistent with prior estimates (e.g., Hanushek & Rivkin, 2010), we find that a 1 *SD* difference in the distribution of teacher effectiveness represents approximately a 0.18 *SD* difference in student test scores (see Table 1, column 1). Thus, a prototypical teacher who as a novice was at the 27th percentile of the distribution of overall effectiveness moves to approximately the median after 10 years of experience. As Boyd, Lankford, et al. (2008) make clear, it makes sense to compare the effects of interventions affecting teachers with the *SD* of gain scores (in effect, 0.18 *SD* here).

Do the Returns to Teaching Experience Differ Across Individual Teachers and Schools?

Estimates from Model II confirm that the average returns-to-teaching-experience profile obscures a large degree of heterogeneity in individual teachers' changes in effectiveness over time. In the first column of Table 1, we present the estimated *SD*s of each of the random effects included in Model II. We find that the estimated *SD* of the random slopes for returns to experience across individual teachers (σ_{β_j}) is 0.025 test-score *SD*s ($p < .001$). This suggests that a teacher who is at the 75th percentile of returns to experience is improving her effectiveness by almost 2% of a test-score *SD* more *annually* than a teacher whose improvement follows the average returns to experience trajectory. Specifying Model II with random intercepts for schools leads to almost identical results (column 2).

We illustrate this heterogeneity visually by plotting—in Panel A of Figure 2—the fitted returns-to-teaching-experience profiles for a random sample of 25 early-career teachers who had taught for at least 7 years. Each teacher's predicted random intercept serves as an estimate of her initial effectiveness level as a novice teacher. Individual returns-to-experience profiles are obtained from our fitted models by combining the estimated average returns to experience and the estimated teacher-specific deviations from this average pattern. In panel B, we center each teacher's random intercept on zero to focus attention on how much individuals improve relative to their effectiveness as a novice teacher.

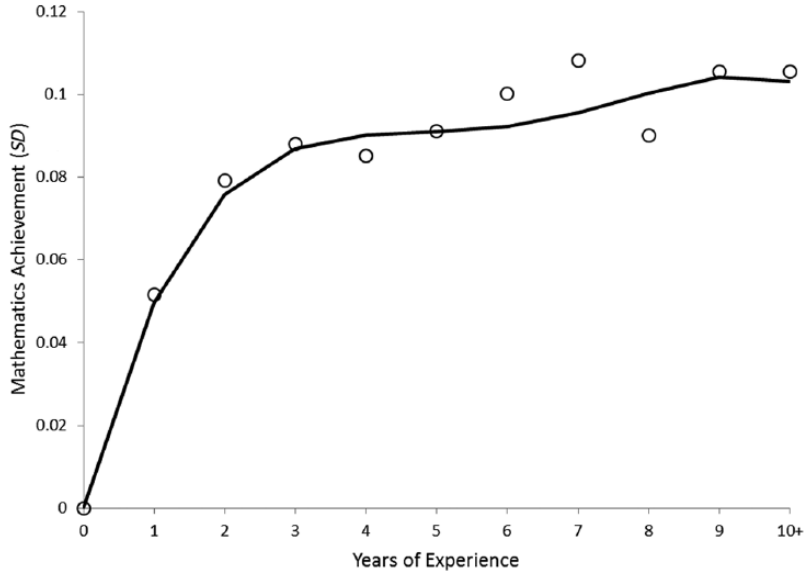

FIGURE 1. *Estimated average returns to teaching experience, over 10 years.*
Note. This figure is based on parameter estimates from Model I, which specified achievement as a non-parametric function of teaching experience, with results from the quartic functional form overlaid.

TABLE 1
Standard Deviations of Random Intercepts and Slopes from Multilevel Models Examining Heterogeneity in Returns to Teaching Experience

	(1)	(2)	(3)
Teacher intercepts ($\sigma_{\alpha j}$)	0.177***	0.170***	0.143***
	(0.004)	(0.004)	(0.002)
Teacher slopes ($\sigma_{\beta j}$)	0.025***	0.024***	
	(0.001)	(0.009)	
School intercepts ($\sigma_{\mu s}$)		0.068***	0.073***
		(0.006)	(0.007)
School slopes ($\sigma_{\beta s}$)			0.007***
			(0.001)
Residual (σ)	0.443***	0.443***	0.444***
	(0.001)	(0.001)	(0.001)
Correlation ($\sigma_{\alpha j}$, $\sigma_{\beta j}$)	−.625	−.658	
Correlation ($\sigma_{\mu s}$, $\sigma_{\beta s}$)			−.353
Observations	280,687	280,687	280,687

Note. Standard errors are reported in parentheses.
*p < .05. **p < .01. ***p < .001.

Two overall patterns emerge from this figure. As is now widely documented in the literature, Panel A depicts substantial differences in individual teacher effectiveness across teachers. Second, the figure also demonstrates how returns to teaching experience differ widely across teachers. The intersecting profiles in Panel A demonstrate how these differences in the rate of improvement cause some teachers to become more effective than others over time. Panel B

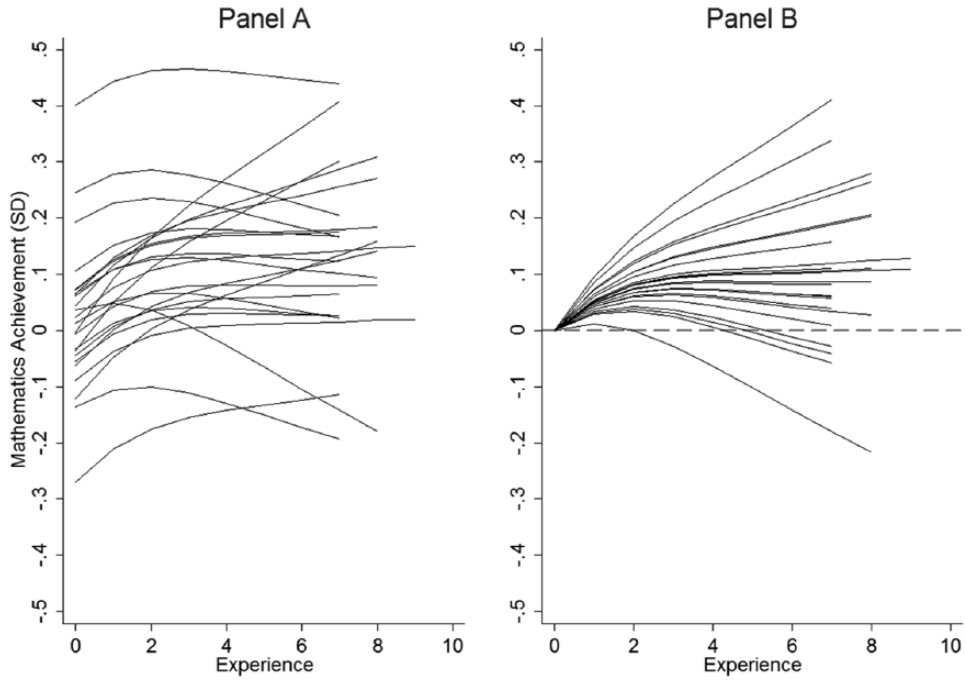

FIGURE 2. *Estimated individual returns to teaching experience for a random sample of 25 teachers with at least 7 years of classroom experience.*
Note. We depict in this figure the returns to experience profiles of 25 teachers based on post hoc estimates from Model II. In Panel B, we center these profiles at zero to illustrate differences in the returns to experience over time across teachers.

helps to illustrate this point. Relative to each teacher's initial effectiveness, some teachers improve much more rapidly than others.

We also find strong evidence of variation in the average returns to experience among teachers across individual schools. In the final column of Table 1, we present results from Model III, in which we include both random intercepts and slopes for schools. Here, we estimate that the *SD* of the school-specific random slopes is 0.007 *SD* ($p < .001$), or almost 30% of the estimated variation in returns to experience across individual teachers. In other words, teachers in certain schools tend to improve more than teachers in other schools.[12]

Descriptive Findings on Professional Environments in Schools

We now examine whether the quality of the professional environment in schools accounts for the estimated differences in returns to experience across schools described above. Overall, there exist meaningful differences in the quality of the professional environment in which teachers work in CMS. To illustrate this point, we present the sample distribution of teachers' average responses within a school to three individual survey items from 2008 in Figure 3. For example, teachers' average perceptions of whether "There is an atmosphere of trust and mutual respect within the school" and whether "School administrators support teachers' efforts to maintain discipline in the classroom" differ widely across schools, with long left-hand tails suggesting that some schools struggled in these areas. These distributions also reveal that teachers on the whole felt only slightly more positive than neutral about these statements.

Not surprisingly, a school's professional environment is also related to the characteristics of its students and teachers. In Table 2, we compare school-level averages of selected student and teacher characteristics by the quartiles of the overall rating of professional environment. Schools with more supportive professional environments serve students who are higher achieving, more likely to be White, less likely to be

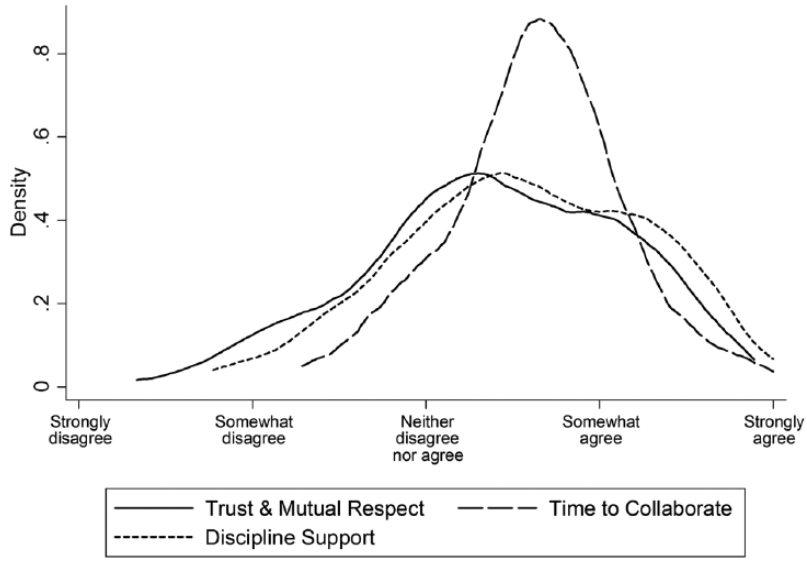

FIGURE 3. *Sample distributions of unstandardized school-average responses to three survey items from the 2008 North Carolina Teacher Working Conditions Survey.*
Note. Full item stems are "There is an atmosphere of trust and mutual respect within the school," "Teachers have time available to collaborate with their colleagues," and "School administrators support teachers' efforts to maintain discipline in the classroom."

from low-income families, and more likely to attend school. On average, students at schools in the top quartile of the professional environment outperform their counterparts in the lowest quartile by three fourths of a *SD* in both mathematics and reading and are absent over 3 days less per year. White students make up over half of the student population at top quartile schools compared with less than 15% at bottom quartile schools.

Schools with the most supportive professional environments also employ more highly qualified teachers. Teachers who are experienced, earned National Board certification, hold master's degrees, and graduated from competitive colleges are more likely to teach in top quartile schools. Teacher sorting by race mirrors the same patterns found among students. Schools in the bottom quartile of the professional environment employ less experienced teachers on average and more than twice as many alternatively certified teachers as all other schools.

These strong associations between student characteristics—and to a lesser degree teacher characteristics—and the professional environment in schools pose important challenges for analysts using observational data. They illustrate the difficulties many past researchers have faced when attempting to disentangle the effect of working conditions from the characteristics of students or teachers in a school. They also highlight the importance of including our rich set of controls for student characteristics and teacher-fixed effects in our statistical models, as well as examining whether the returns to experience differ by these teacher and student characteristics (as we do in sensitivity tests).

Do Teachers in Schools With Stronger Professional Environments Improve More Over Time?

We find substantial heterogeneity in returns to experience across schools with different professional environments. A 1 *SD* difference in the quality of the professional environment in which teachers work is associated with an additional 0.0026 *SD* ($p = .024$) increase in the *annual* returns to teaching experience (Table 3, column 1). This becomes a 0.0052 *SD* difference after 2 years, a 0.0078 *SD* difference after 3 years, and eventually a 0.0260 *SD* difference after 10 years. In Figure 4, we illustrate the magnitude of these differences as they compound over time by plotting the within-teacher returns-to-teaching-experience profiles of

TABLE 2
Sample Averages of Student and Teacher Characteristics by Quartiles of the Overall Professional Environment in Schools

		Quartiles of professional environment			
	M	Bottom	2nd	3rd	Top
Panel A: School-average student characteristics					
Student enrollment	807	889	639	844	837
Mathematics achievement in Grades 4–8 (*SD*)	−0.09	−0.45	−0.38	0.03	0.36
Reading achievement in Grades 4–8 (*SD*)	−0.09	−0.42	−0.37	0.03	0.35
Days absent per student	7.44	9.41	7.94	6.64	6.03
Limited English proficient	0.15	0.17	0.17	0.15	0.12
Special education	0.11	0.14	0.11	0.10	0.10
Free and reduced price lunch	0.56	0.71	0.74	0.50	0.32
African American	0.45	0.61	0.59	0.38	0.24
White	0.30	0.13	0.15	0.36	0.53
Hispanic	0.16	0.18	0.18	0.16	0.13
Panel B: Teacher characteristics					
0 to 3 years of experience	0.37	0.42	0.41	0.36	0.30
4 to 10 years of experience	0.33	0.31	0.35	0.34	0.34
11 years of experience or more	0.30	0.26	0.24	0.31	0.36
African American	0.24	0.39	0.28	0.22	0.13
White	0.73	0.58	0.70	0.76	0.85
National Board Certified	0.07	0.04	0.06	0.06	0.12
Master's degree	0.27	0.24	0.26	0.28	0.30
Alternative pathway certification	0.12	0.23	0.08	0.09	0.07
Less competitive college	0.15	0.22	0.18	0.13	0.08
Competitive college	0.50	0.49	0.53	0.49	0.49
Very competitive college	0.24	0.20	0.20	0.26	0.27
Highly competitive college	0.07	0.06	0.06	0.06	0.11
Most competitive college	0.04	0.03	0.03	0.05	0.05

Note. We estimate statistics for student characteristics using school-year averages for those schools in the analytic sample. We estimate statistics for teacher characteristics using all teacher-year observations that are represented in the analytic sample.

three prototypical teachers, those at schools which are rated as average as well as at the 75th and 25th percentiles of the professional environment ratings.

On average, after 3 years, teachers working in schools at the 75th percentile of professional environment ratings have improved their effectiveness by 0.010 *SD* more than teachers working in schools at the 25th percentile, a 12% improvement gap. After 5 years, teachers working at schools at the 75th percentile have improved their effectiveness by 0.017 *SD*, on average, a 20% gap. As Figure 4 shows, by year 10, a prototypical teacher at a school with a very strong professional environment will have improved by 0.035 *SD* more on average than a teacher in a school with a very weak professional environment, a 38% gap. Thus, after 10 years, teachers at a school with a more supportive professional environment move upward in the distribution of overall teacher effectiveness by approximately one fifth of a *SD* more than teachers who work in less supportive professional environments.

We extend this analysis by refitting Model IV with each of our six conceptually distinct elements of the professional environment and present these exploratory results in Online Appendix

TABLE 3
Parameter Estimates of the Differential Returns to Teaching Experience Across Schools With More Supportive and Less Supportive Professional Environments

	(1)	(2)	(3)	(4)
EXPER × PROF_ENV*	0.0026*	0.0025*	0.0039*	0.0027
	(0.0011)	(0.0011)	(0.0016)	(0.0027)
*EXPER**	0.0664***	0.0685***	0.0673***	0.0664***
	(0.0087)	(0.0085)	(0.0087)	(0.0087)
EXPER^2*	−0.0174***	−0.0188***	−0.0180***	−0.0174***
	(0.0039)	(0.0039)	(0.0040)	(0.0039)
EXPER^3*	0.0020**	0.0022***	0.0021**	0.0020**
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
EXPER^4*	−0.0001*	−0.0001**	−0.0001*	−0.0001*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
PROF_ENV	−0.0087			−0.0090
	(0.0083)			(0.0092)
EXPER^2 × PROF_ENV*				−0.0000
				(0.0002)
Teacher-fixed effects	Yes	Yes		Yes
School-fixed effects		Yes		
Teacher-by-school fixed effects			Yes	
Observations	280,687	280,687	280,687	280,687

Note. Standard errors clustered by school-grade-year are reported in parentheses. All student-level models include grade-by-year fixed effect as well as vectors of student, peer, and school-level covariates.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 1. Peer collaboration and school culture are among the strongest predictors, but we emphasize that each element captures a large degree of common variance and that all six parameter estimates are statistically indistinguishable from each other.

Assessing Model Assumptions

In Table 3, we present parameter estimates from our preferred model as well as alternative specifications of our education production function which show that our primary results are not driven by our modeling decisions.[13] We begin by augmenting our preferred specification of Model IV by including school fixed effects, implicitly removing the effect of all time-invariant student or teacher characteristics that differ systematically across schools. Although we must remove the main effect of the professional environment from the model because it does not vary within school, we can still estimate average differences in returns to teaching experience across professional environments. We find that our parameter estimate describing the differential returns to experience across professional environments remains virtually unchanged (column 2).

Second, we replace teacher-fixed effects with teacher-by-school fixed effects. Including teacher-by-school effects restricts our estimates of the returns to experience to within teacher-school combinations, eliminating the threat that specific patterns of teacher-transfer across schools could create the effects we find. This approach produces somewhat larger and statistically significant estimates of the differential returns to teaching experience across professional environments, suggesting that the more conservative results from our primary approach using Model IV may understate the potential effect of the professional environment (column 3).

Finally, we relax our assumption that the differential returns to experience across professional

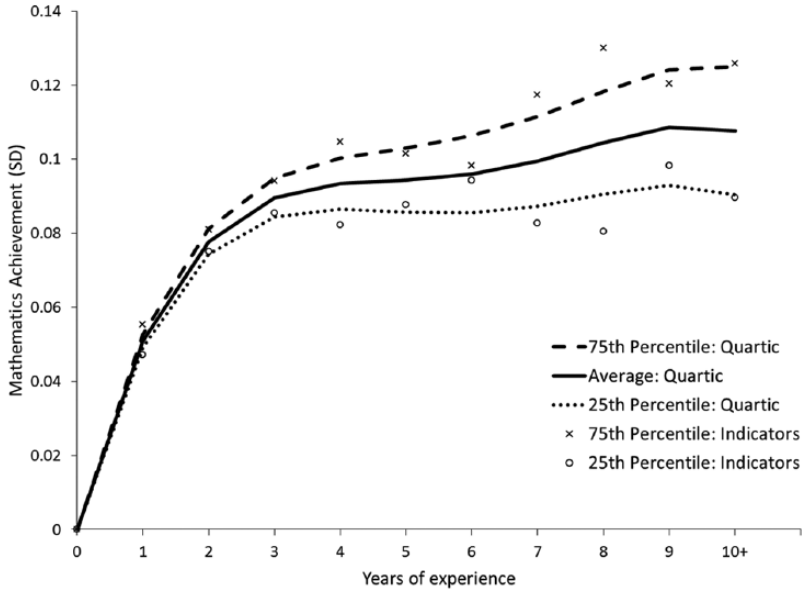

FIGURE 4. *Fitted returns to teaching experience for prototypical teachers, across school professional environments.*

Note. In this figure, we plot fitted estimates from Model IV in which we specify the main effect of experience as a quartic function, and allow for individual linear deviations from the average relationship between experience and achievement. We also include fitted estimates from a model in which we specify experience as a vector of dichotomous indicator variables up to 10 years interacted with our measure of the overall professional environment for teachers at the 25th and 75th percentile schools.

environments are linear. In column 4, we report results from a model that allows for the *differential* returns to experience to take on a quadratic functional form. Again, we continue to specify the average returns-to-experience using a quartic polynomial; we simply model deviations from this average trend using a quadratic relationship. The point estimate on the interaction of our measure of the overall professional environment and the square of experience is not statistically significant and is precisely estimated as very close to zero, which suggests that the underlying pattern is linear. To be even more conservative, we also fit a model that uses a completely flexible set of indicator variables to model these deviations. As seen above in Figure 4, these non-parametric point estimates are well approximated by our preferred model.

Alternative Explanations

The analytic methods discussed above allow us to show clearly that teachers in schools with stronger professional environments experience greater returns to experience over time. Ultimately, we also want to know whether it is the work environment itself that causes this additional improvement over time. Thus, we examine the most plausible alternative explanations for the patterns we observe in our data to further our understanding of the potential causes of this observed relationship. However, we cannot make definitive causal statements about the relationship between the professional environment and teacher development given the lack of exogenous variation in the professional environment in our data.

Our construction and use of the professional environment measure presents several possible alternative explanations. First, the relationship between returns to experience and the professional environment could be a product of nonresponse bias, self-report bias, or reverse causality. Second, other unobserved characteristics that are correlated with the professional environment in a school could be the underlying cause of the observed relationship. In addition, a pattern where schools with more favorable professional environments recruit, select, and retain teachers with greater potential for improvement over time could account for our results. A final

TABLE 4

Parameter Estimates of Differential Returns to Teaching Experience Across Schools Using Alternative Measures of the Overall Professional Environment in Schools

	Restricted sample of schools with combined 50% response rate across surveys (2001–2010)	Professional environment measure constructed using only veteran teachers (2001–2010)	2006 measure of professional environment (2006–2010)	Average professional environment measure in 2006, 2008, 2010	Time-varying professional environment measure in 2006, 2008, 2010
	(1)	(2)	(3)	(4)	(5)
EXPER × PROF_ENV*	0.0032*	0.0025*	0.0071*	0.0022	0.0024
	(0.0013)	(0.0012)	(0.0031)	(0.0022)	(0.0023)
*EXPER**	0.0637***	0.0662***	0.0748***	0.0474*	0.0400*
	(0.0094)	(0.0087)	(0.0142)	(0.0199)	(0.0192)
EXPER^2*	−0.0151***	−0.0174***	−0.0216***	−0.0083	−0.0080
	(0.0042)	(0.0039)	(0.0061)	(0.0081)	(0.0080)
EXPER^3*	0.0015*	0.0020**	0.0029**	0.0009	0.0009
	(0.0007)	(0.0006)	(0.0010)	(0.0012)	(0.0012)
EXPER^4*	−0.0001	−0.0001*	−0.0001**	−0.0000	−0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)
PROF_ENV	−0.0138	−0.0077	0.0046	0.0024	0.0224
	(0.0095)	(0.0087)	(0.0241)	(0.0140)	(0.0165)
Teacher-fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	243,178	280,687	125,302	86,343	86,343

Note. Standard errors clustered by school-grade-year are reported in parentheses. All models include grade-by-year fixed effects as well as vectors of student, peer, and school-level covariates. Model 1 restricts the sample to the 91% of schools that have at least a 50% combined response rate across all three Teacher Working Condition Surveys. Model 2 uses an alternative measure of the professional environment constructed using only responses from teachers with 11 or more years of experience. Model 3 uses only data from 2006 to construct a measure of the professional environment and only includes schools with a valid 2006 measure. Model 4 uses our preferred overall measure of the professional environment in a restricted sample that includes the 3 years the Teacher Working Condition Survey was administered. Model 5 uses a time-varying measure of the professional environment and restricts the sample to the same 3 years as Model 4. In Model 5, we impute missing values of the time-varying measure of the professional environment with the average professional environment among years with valid data for each school to facilitate a comparison of results with Model 4 which is not confounded by sample differences. We account for this imputation by including an indicator for school years with missing values for the professional environment and its interaction with *EXPER**. Missing values are concentrated in 2006 and represent 4.7% of the school-year observations in Model 5.

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

alternative explanation could be that student assignment patterns to individual teachers over time differ across schools in ways that relate systematically to the work environment.

Endogeneity in the Measurement of the Professional Environment

The construction of our measure of the overall professional environment using teacher survey data presents the potential for three types of endogeneity: non-response bias, self-report bias, and reverse causality. We present evidence to assess the contributions of these biases to our results in Table 4. First, it could be that the opinions of teachers who responded to the Working Conditions Survey do not reflect the general opinion of teachers in their school. This issue would be of particular concern in schools with low response rates. To test this, we restrict our

sample to include only those schools with at least a 50% response rate across the three survey administrations. Results reported in column 1 of Table 4 demonstrate that our findings are slightly stronger using this restriction, suggesting that measurement error due to low response rates may be attenuating our results. We also examine the demographic characteristics of teachers who responded to the survey and find that they are quite similar to those of the district's workforce as a whole.

A second concern is that, although the survey was both anonymous for teachers and its results were not considered in any school-evaluation process, individual teachers' responses to the working conditions survey may be systematically biased. Here, the issue is not that teachers overall rated schools systematically higher (or lower), but that teachers in schools where early-career teachers were improving at greater rates had systematically inflated responses. We construct a test of this self-report bias by creating an alternative measure of *PROF_ENV* using only the self-reported data from teachers with 11 years of experience or more. This allows us to make inferences about the improvement of teachers in their first 10 years of the career without relying on their own self-reported data to measure the professional environments in their schools. As seen in column 2, results using this alternative measure of *PROF_ENV* are nearly identical to our preferred estimates, demonstrating that our findings do not appear to be subject to a self-reporting bias.

A third potential concern is that, by employing survey data from 2006 to 2010 to characterize the professional environment in previous and concurrent years, our findings may be the result of reverse causality. We examine this threat by refitting Model IV in three ways. First, we construct our measure of the professional environment using data from the first survey in 2006 and limit our analysis to data from 2006 to 2010. Using this 2006 measure of the working environment, we confirm that a prior measure of the work environment predicts large and statistically significant differential returns to experience in future years (column 3). Here, estimates are substantially larger than in our preferred model. Second, we then restrict our sample to include only the 3 years during which the Working Condition survey was administered (2006, 2008, and 2010). We fit two models in this restricted sample, one with our preferred school-level average measure of the professional environment across these 3 years and one with a time-varying measure. Parameter estimates in both models are quite similar in magnitude to our preferred estimates, but are not statistically significant. These imprecise estimates are likely the result of our reduced analytic sample from 10 to 3 years of data, although we cannot rule out the possibility that these coefficients are zero in the population (columns 4 and 5). In short, the consistent pattern of results across these different specifications suggests that non-response bias, self-report bias, and reverse causality are not driving our findings.

Omitted Variable Bias

Another concern is that our estimates of the differential returns to experience may not be driven by the professional environment in schools, but instead capture differences due to unobserved teacher or student-body characteristics that are positively correlated with both the quality of professional environment and student achievement. For example, it could be that certain types of teachers are more likely than others to improve with experience and to work in stronger professional environments. Or, perhaps teachers improve their effectiveness more rapidly when they teach certain types of students who are likely to attend schools with stronger professional environments. To test for these alternative explanations, we also allow for differential returns to experience related to individual teacher characteristics as well as average student characteristics in a school.

We refit Model IV with an additional interaction term of $EXPER^{*}$ and one of several teacher or student-body characteristics and report the results in Table 5. If our estimates are the result of omitted variable bias, then including these terms should attenuate our point estimates substantially. Instead, our estimates of the differential returns to experience across professional environments remain practically unchanged by the addition of these

TABLE 5
Sensitivity Analyses of the Differential Returns to Teaching Experience Across Schools With More Supportive and Less Supportive Professional Environments

Panel A: Models that also allow for differential returns by individual teacher characteristics

	White	African American	National Board Certification	Alternative certification	Master's degree	Competitive college
	(1)	(2)	(3)	(4)	(5)	(6)
*EXPER** × *PROF_ENV*	0.0028*	0.0028*	0.0024*	0.0026*	0.0026*	0.0030*
	(0.0012)	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0012)
*EXPER** × *Teacher characteristic*	−0.0042	0.0046	0.0033	0.0013	−0.0007	−0.0021
	(0.0029)	(0.0030)	(0.0033)	(0.0013)	(0.0028)	(0.0028)
Observations	278,169	278,169	280,687	274,481	280,687	257,202

Panel B: Models that also allow for differential returns by student-body characteristics

	% FRPL	% African American	% limited English proficient	% special education	Average mathematics achievement in prior year	Average reading achievement in prior year
	(1)	(2)	(3)	(4)	(5)	(6)
EXPER PROF_ENV*	0.0021†	0.0021†	0.0025*	0.0025*	0.0022†	0.0023†
	(0.0012)	(0.0013)	(0.0011)	(0.0011)	(0.0013)	(0.0013)
*EXPER** × *Student body characteristic*	−0.0014	−0.0009	−0.0006	−0.0032	0.0015	0.0018
	(0.0011)	(0.0010)	(0.0010)	(0.0025)	(0.0023)	(0.0030)
Observations	280,687	280,687	280,687	280,687	280,687	280,687

Note. Standard errors clustered by school-grade-year are reported in parentheses. All models include teacher-fixed effects and grade-by-year fixed effects as well as vectors of student, peer, and school-level covariates. We omit the main effect of time-invariant teacher characteristics from models. We include the main effect of the respective student-body characteristic in all models that allow for differential returns to experience by student-body characteristics. FRPL = free and reduced price lunch.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

interactions. Overall, these results confirm that differences in returns to teaching experience across professional environments are not driven by the strong correlations between the professional environment and observed teacher characteristics or student-body characteristics in a school.

Dynamic Teacher Sorting Across Schools

Finally, we must be concerned that teachers who will improve at greater rates selectively sort into schools with stronger professional environments. Patterns of highly qualified teachers sorting to more affluent, suburban, and White communities are widely documented in the literature on teacher mobility (Clotfelter, Ladd, Vigdor, & Wheeler, 2007; Lankford, Loeb, & Wyckoff, 2002). However, our inclusion of teacher-fixed effects removes the possibility that a pattern where more *effective* teachers sorting to schools with better professional environments would produce our results. Instead, the concern is only that schools with more favorable professional environments selectively recruit and hire teachers with greater potential for *improvement* over time. Although we cannot rule out this alternative explanation, we find

it more likely that schools would search for effective teachers, rather than teachers who will improve. Furthermore, a large body of literature documents the quite weak relationship between observable teacher characteristics and future effectiveness (Clotfelter, Ladd, & Vigdor, 2007; Wayne & Youngs, 2003), as well between measures of teachers' conscientiousness and self-efficacy and future effectiveness (Rockoff et al., 2011).

Positive estimates of the differential returns to teaching experience across professional environments could also be the result of differential retention. Again, our results will only be biased if schools are retaining teachers selectively who are improving more over time, not simply that schools retain teachers who are more (or less) effective on average. We are able to examine this possibility by testing whether the relationship between the probability of leaving a school and estimates of an individual's returns to teaching experience differ by the quality of the professional environment in a school. To do this, we fit our teacher-specific random slopes and intercepts model (Model II) and obtain estimates of individual teachers' pace of improvement, relative to the average returns to experience, from the fitted model. We then use these best linear unbiased predictions of the degree to which an individual teacher is improving ($\hat{\beta}_j$) as a predictor in the following linear probability model that at time *t* teacher *j* leaves school *s*:

$$\begin{aligned} LEAVE_{tjs} &= \gamma_1 \hat{\beta}_j + \gamma_2 PROF\_ENV_s \\ &+ \gamma_3 \left(\hat{\beta}_j \times PROF\_ENV_s \right) + \tau_t \\ &+ \psi \left(f\left(EXPER^*_{jt} \right) \right) + r_{tjs} \end{aligned} \tag{5}$$

We include fixed effects for calendar year and teacher experience to account for any district-wide trends in student achievement or teacher employment patterns. The parameter estimate on the $(\hat{\beta}_j \times PROF\_ENV_s)$ interaction term, γ_3, tests whether teachers who are improving more rapidly are less likely to leave schools with more supportive professional environments. We find that teachers in more supportive professional environments are less likely to leave, but the probability of leaving one's school is unrelated to both changes in a teacher's effectiveness over time as well as the interaction of changes in effectiveness and our measure of the professional environment. In other words, γ_3 is not statistically different from zero ($p = .41$).[14] Thus, we find no evidence that dynamic attrition explains our results.

Differential Student Sorting Within Schools and Teachers

Finally, it is possible that our results are the product of differential student sorting patterns to individual teachers as they gain experience, which are related to the professional environment. Although more senior teachers and teachers in schools with better professional environments are often assigned higher achieving students (Clotfelter, Ladd, & Vigdor, 2006; Kalogrides, Loeb, & Beteille, 2011), these patterns alone could not explain our results. Instead, our findings could only be caused by a differential sorting pattern where schools with better professional environments were more likely to assign such students to more experienced teachers than schools with worse environments. Furthermore, because we include selected observable student characteristics (including prior test scores) as covariates in our model, this alternative explanation would require a pattern of differential sorting over time on *unobserved* characteristics that are positively correlated with test scores. In fact, evidence suggests that the opposite pattern may hold (Loeb et al., 2012).

A true test of this alternative explanation is impossible to conduct because, by definition, the variables we would like to examine are unobserved. Instead, we attempt to understand the nature of dynamic sorting on observed student characteristics to gain insights into the potential sorting on unobserved characteristics. To do this, we modify model (IV) by using student characteristics as our outcomes, S_CHAR, and removing all student and peer level-covariates as follows:

$$\begin{aligned} S\_CHAR_{itjs} &= \varsigma\left(f\left(EXPER^*_{jt} \right) \right) \\ &+ \lambda_1 PROF\_ENV_s \\ &+ \lambda_2 \left(EXPER^*_{tj} \times PROF\_ENV_s \right) \\ &+ \theta S_{ts} + \omega_j + \xi_{gt} + \upsilon_{itjs} \end{aligned} \tag{6}$$

Again, the coefficient on the interaction of $EXPER^*$ and $PROF\_ENV$, λ_2, is our parameter of interest, as it examines whether certain types of students are more likely to be assigned

to teachers as they gain experience in schools with more supportive professional environments than in schools with less supportive environments. In Online Appendix Table 2, we show that estimates of λ_2 are near zero and not statistically significant across nearly all student characteristics. The only statistically significant estimate points in the opposite direction than the type of sorting that might indicate potential upward bias in our estimates. This suggests that patterns of student sorting to teachers over time appear to be consistent across schools and dynamic sorting on unobserved student characteristics is unlikely to explain away our findings.

Conclusion and Policy Implications

With this study, we have sought to document heterogeneity in the returns to teaching experience and to examine whether this heterogeneity can be explained, in part, by the professional environment in which teachers work. We find strong evidence of such heterogeneity, establishing that there is not only substantial variation in teacher effectiveness but also in the pace at which teachers improve their effectiveness. Some teachers are improving two or three times faster than others and continue these rapid gains in effectiveness throughout their first 5 to 10 years on the job. This large variation in returns to experience across teachers has important implications for research and policy on teacher effectiveness.

Researchers often treat teacher effectiveness as fixed, attributing year-to-year fluctuations to classroom-peer effects or sampling error. This approach assumes away an important element of teacher effectiveness dynamics, how it changes over time with experience. Teachers are also commonly characterized as having a fixed level of effectiveness in the popular press and in education policy reform initiatives. For example, if Ms. Smith is an effective teacher, she should be recruited, rewarded, and retained. If Ms. Jones is an ineffective teacher, we should avoid hiring her and she should not be granted tenure. That some teachers are far more effective than others is an empirical fact. However, these characterizations fail to consider the substantial degree to which individual teachers improve over their careers and the large variation in this improvement. The frequent crossing of returns to experience profiles plotted in Panel A of Figure 2 demonstrates how the rank order of teacher effectiveness changes as teachers improve at different paces over time. A novice teacher who struggles at first but makes sustained improvements over time may become more effective overall than an average novice teacher who fails to improve with experience.

Our findings also illustrate how policies aimed at improving teacher effectiveness that focus on the individual, ignoring the role of the organization, fail to recognize or leverage the potential importance of the school context in promoting teacher development. We show that the degree to which teachers become more effective over time varies substantially by school. In some schools, teachers improve at much greater rates than in others. We find that this improvement is strongly related to the opportunities and supports provided by the professional context in which they work. For example, we estimate that teachers who work in schools at the 75th percentile of professional environment ratings increase their effectiveness by over 0.035 test-score *SD* more over the course of 10 years than a similar teacher at a school at the 25th percentile, a 38% difference in total improvement.

Although these findings are not definitive, causal evidence that improving the professional environment will accelerate teacher development, they are consistent with recent evidence that the school context has lasting effects on teachers' practice and career decisions. For example, Ronfeldt (2012) finds that pre-service teachers who have field placements in easier-to-staff schools become more effective teachers and are less likely to leave the profession after 5 years, a result that is not driven by student characteristics or teacher sorting.

While our estimates of the differences in returns to experience across professional environments are small in absolute magnitude, they are substantial given the overall distribution of teacher effectiveness in the district. A difference of 0.035 test-score *SD* is approximately 20% of a *SD* in the distribution of overall teacher effectiveness and represents over 30% of the average

total improvement teachers make in their first 10 years on the job. As Boyd, Grossman, Lankford, Loeb, and Wyckoff (2008) note, estimates of the effects of many interventions designed to improve teacher effectiveness are overwhelmingly of similar or smaller magnitude (e.g., Boyd, Lankford, et al., 2008; Goldhaber, Liddle, & Theobald, 2013; Kane et al., 2008; Koedel, Parsons, Podgursky, & Ehlert, in press). Furthermore, these results are likely to be a lower bound estimate for several important reasons. First, measurement error inherent in the survey response data we use to quantify the professional environments in schools will necessarily attenuate our findings. Second, CMS's district-wide efforts to improve schools' professional environments are not captured by our estimates, as we only examine variation in environments across schools. Finally, our measure of effectiveness based on teachers' contributions to student achievement on standardized tests does not fully capture many aspects of teachers' professional practice or the ways in which veteran teachers contribute to the effectiveness of their peers and assume important leadership roles (Jackson & Bruegmann, 2009; Johnson & Birkeland, 2003).

Ultimately, comparing point estimates across studies fails to capture a central difference between supportive professional environments and many interventions intended to improve teacher effectiveness. In contrast to a one-time investment in teacher skills, teachers have the potential to benefit from the learning opportunities provided by a supportive professional environment every day. Our findings suggest that working in a more supportive environment is related to improvement which accumulates throughout the first 10 years of the career.

Furthermore, our study's findings that strong professional environments are related to teacher improvement align with the growing recognition that such environments benefit teachers and students systematically. For example, if teachers in more supportive environments improve more and feel more successful because of this improvement, this "sense of success" can increase the likelihood they remain at their schools (Johnson & Birkeland, 2003). A large body of research finds that strong professional environments are directly related to teacher retention (Allensworth, Ponisciak, & Mazzeo, 2009; Boyd et al., 2011; Johnson et al., 2012; Ladd, 2011; Loeb et al., 2005). As effective teachers remain in schools, opportunities for meaningful peer collaboration and a positive organizational culture become even more possible. This positive cycle can lead to effective school organizations, while the opposite pattern can occur in hard-to-staff schools. Poor working conditions may stifle teachers' efforts to improve their practice, promoting turnover and contributing to staffing challenges.

Scholarly research is just beginning to discover why some teachers improve more than others and the importance of school organizational environments for systemic improvement. Practice and research have started to highlight promising avenues for promoting improvement among teachers, such as providing teachers with actionable feedback about their instruction, creating opportunities for productive and sustained peer collaboration, supporting teachers' efforts to maintain an orderly and disciplined school environment, and investing in a school culture characterized by high expectations, trust, and mutual respect. Transforming schools into organizations that support the learning of both students and teachers will be central to any successful effort to increase the human capital of the U.S. teaching force.

Authors' Note

Andrew Baxter and Thomas Tomberlin at Charlotte–Mecklenburg Schools and Eric Hirsch at the New Teacher Center generously provided the data for our analyses. The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, or the Spencer Foundation.

Acknowledgment

We thank Heather Hill, Lawrence Katz, Susan Moore Johnson, Richard Murnane, Doug Staiger, Eric Taylor, and John Willett for their valuable feedback on earlier drafts.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Spencer Foundation and the Institute of Education Sciences, U.S. Department of Education Grant (R305C090023) to the President and Fellows of Harvard College.

Notes

1. We define teachers in our data as individuals in the Human Resources employment files who are paid based on the teacher salary schedule, who have titles that indicate they are classroom teachers, and who are uniquely identified as the math teacher of record in the course file. This results in a total of 3,922 fourth- through eighth-grade teachers who taught students in math. We restrict our sample to only those teachers who teach regular education classes and who have at least five students with valid current and prior test scores in math, removing 2.6% of our sample. We then drop an additional 1.1% of teachers who work in schools that do not have at least 10 respondents to the Working Conditions Survey in any year. Next, we remove teachers from the data for whom we observe irregular jumps on the salary experience scale in back-to-back years. These irregular jumps are likely caused by human resource processing delay, retroactive credits awarded for relevant outside experience, or measurement error, and could bias our estimates of the return to teaching experience. This eliminates 5.5% of our sample. Finally, we restrict our estimates to those teachers who have continuous experience profiles, meaning that they do not leave the district and return. This removes 10.5% of the sample. Importantly, relaxing either of these two final sample restrictions, or both, does not change the character of our results.

2. Formally, $EXPER^{*}_{ij} = \begin{cases} k, k \leq 10 \\ 10, k > 10 \end{cases} \forall k \in [0,39]$.

Rockoff justifies this assumption by citing previous literature that finds no evidence of returns to teaching experience beyond the first several years on the job. We also choose 10 years because we are primarily interested in early-career returns to experience, where we possess the most data; teachers with 10 years of experience or fewer comprise 70% of our estimation sample.

3. We estimate models using two alternative identifying assumptions. First, we censor experience at 20 years. Second, we adopt a two-stage modeling approach that we have developed in a separate article (Papay & Kraft, 2013). As expected, we find that our results are quite consistent across modeling approaches.

4. In each survey year, the Eigenvalue of the first principal component was greater than 12, with item loadings ranging between 0.14 and 0.24 across all 24 times, while the Eigenvalue of the second principal component was less than 2. Visual inspection of the corresponding scree plots shows a clear kink at the second Eigenvalue.

5. Several other factors contributed to this decision. It is unclear whether differences in our measure of the professional environment across survey administrations are capturing true changes over time, or whether these differences are due to the changing composition of survey respondents, differential survey response rates across years, or changes in the response anchors across years.

6. We obtain these data from publicly available records maintained by the North Carolina Department of Instruction. These state records cover 90.5% of the school-years observations in our analytic sample. We impute school-specific values for missing data by taking the average of all available school-year observations for a given school. We include a dichotomous indicator for school years in which we imputed missing data in all our models.

7. We estimate standard errors after clustering students by school-by-grade-by-year to account for possible unobserved correlations among the residuals of students in the same grade cohort within a school.

8. Including lagged student test scores, $A_{i,t-1}$, as independent variables is potentially problematic because these scores are an imperfect measure of true achievement. It is possible that measurement error in lagged test scores will be correlated with measurement error in current-year test scores. This potential for serial correlation between individual students' error terms over time could result in biased estimates. We examine this potential threat by instrumenting for lagged test scores using twice-lagged test scores as proposed by Todd and Wolpin (2003). This alternative estimation strategy does not affect the character of our results.

9. See Murnane and Willett (2010) for a clear discussion of the trade-offs between using fixed- and random-effects models.

10. We assume that the teacher-specific random intercepts and slopes are distributed non-independently, bivariate normal with mean zero, and appropriate population variances and covariance. Fitting our models results in moderate to strong negative estimates of the correlations between the teachers' initial effectiveness and their change in effectiveness over time as reported in Table 1. These are asymptotically unbiased estimates of the population correlation between teachers' true initial status and change, estimated *within* the model rather than estimated by predicting these values for individual teachers and then

correlating these predictions ex-post. Our unbiased estimate of a negative correlation between true change and initial status are consistent with Atteberry, Loeb, and Wyckoff (2012) who found that teachers in "the lowest two quintiles [of initial value-added] exhibit the most improvement."

11. Estimating random-effects variance components within our model using full-information maximum likelihood allows us to obtain consistent estimates of the true population variances and covariance. Alternative estimators such as the corresponding sample variances of ordinary least squares or empirical Bayes estimates of individual teacher intercepts and slopes would result in bias—an overestimate and underestimate, respectively (see Bryk & Raudenbush, 2002).

12. We find nearly identical results across a variety of specifications. First, we change our identifying assumption by censoring the main effect of experience at 20 years instead of 10. Second, we restrict our sample to include only teachers who taught for at least 5 years in the district to ensure that teachers for whom we have very few years of data are not inflating estimated variances. Finally, we relax our assumption that individual and school-specific deviations from the common returns to experience profile are linear by allowing these deviations to take on a quadratic functional form.

13. An alternative approach to modeling these returns to experience involves estimating the model in two steps. Here, we could fit a model similar to that in Model I, omitting the experience predictor and estimating teacher-year effects rather than teacher-fixed effects. We could then regress these teacher-year effects (which essentially reflect an estimate of the teacher's productivity in each year) on a quartic function of teacher experience (uncensored) and an interaction between our measure of the professional environment and a linear measure of experience through the first 10 years of a teacher's career. When we implement this approach, we find that our parameter of interest is 0.0023 (p = .023), nearly identical to the estimate of 0.0026 from our preferred model. We estimate standard errors using the bootstrap method to account appropriately for this two-stage modeling approach.

14. We estimate standard errors using the bootstrap method to account appropriately for the two-stage modeling approach used in Model V.

References

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*, 1034–1037.

Allensworth, E., Ponisciak, S., & Mazzeo, C. (2009, June). *The schools teachers leave: Teacher mobility in Chicago public schools*. Consortium on Chicago School Research. Retrieved from http://ccsr.uchicago.edu/sites/default/files/publications/CCSR_Teacher_Mobility.pdf

Atteberry, A., Loeb, S., & Wyckoff, J. (2012). *Do first impressions matter? Improvement in early career teacher effectiveness* (No. w19096). Cambridge, MA: National Bureau of Economic Research.

Blase, J., & Blase, J. (1999). Implementation of shared governance for instructional improvement: Principals' perspectives. *Journal of Educational Administration*, *37*, 476–500.

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*, 303–333.

Boyd., D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Measuring effect sizes: The effect of measurement error*. Paper prepared for the National Conference on Value-Added Modeling, University of Wisconsin–Madison.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, *27*, 793–818.

Bryk, A., & Raudenbush, S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: SAGE.

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

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

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*, 778–820.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effect. *Economics of Education Review*, *26*, 673–682.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Wheeler, J. (2007). High-poverty schools and the distribution of teachers and principals. *North Carolina Law Review*, *85*, 1345–1379.

Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting sustainability: teachers' advice networks and ambitious instructional reform. *American Journal of Education*, *119*, 137–182.

Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis*, *29*, 262–295.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*, 181–199.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*, 81–112.

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

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Glazerman, S., Dolfin, S., Bleeker, M., Johnson, A., Isenberg, E., Lugo-Gil, J., . . . Britton, E. (2008). *Impacts of comprehensive teacher induction: Results from the first year of a randomized controlled study*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review*, *36*, 29–44.

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*, 2552–2585.

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

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

Harris, D., & Sass, T. (2011). Teacher training, teacher quality, and student achievement. *Journal of Public Economics*, *95*, 798–812.

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

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, *39*, 50–79.

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

Johnson, S. M. (2009). *How best to add value? Striking a balance between the individual and the organization in school reform* (EPI Briefing Paper #249). Retrieved from http://www.epi.org/publication/bp249/

Johnson, S. M., & Birkeland, S. E. (2003). Pursuing a "sense of success": New teachers explain their career decisions. *American Educational Research Journal*, *40*, 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 [Special issue]. *Teachers College Record*, *114*(10), 1–39.

Kalogrides, D., Loeb, S., & Beteille, T. (2011). *Power play? Teacher characteristics and class assignments* (CALDER Working Paper No. 59). Retrieved from http://www.urban.org/publications/1001530.html

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*, 615–631.

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (NBER Working Paper No. 14607). Retrieved from http://www.nber.org/papers/w14607

Kanter, R. M. (1984). *The change masters*. New York, NY: Simon & Schuster.

Kaufman, D., Johnson, S. M., Kardos, S. M., Liu, E., & Peske, H. (2002). "Lost at sea": New teachers' experiences with curriculum and assessment. *Teachers College Record*, *104*, 273–300.

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

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (in press). Teacher preparation programs and teacher quality: Are there real differences across programs. *Education Finance and Policy*.

Ladd, H. (2009). *Teachers' perceptions of their working conditions: How predictive of policy-relevant outcome?* (CALDER Working Paper No. 33). Retrieved from http://www.urban.org/url.cfm?id=1001440&RSSFeed=UI_EducationPolicyCenter.xml

Ladd, H. (2011). Teachers' perceptions of their working conditions: How predictive of planned and

actual teacher movement? *Educational Evaluation and Policy Analysis*, *33*, 235–261.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*, 37–62.

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.

Loeb, S., Kalogrides, D., & Beteille, T. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy*, *7*, 269–304.

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

Matsumura, L. C., Garnier, H. E., & Resnick, L. B. (2010). Implementing literacy coaching: The role of school social resources. *Educational Evaluation and Policy Analysis*, *32*, 249–272.

May, H., & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly*, *47*, 332–352.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. A. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*, 67–101.

Murnane, R. J., & Phillips, B. R. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review*, *1*, 453–465.

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal*, *46*, 532–566.

Papay, J. P., & Kraft, M. A. (2013). *Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career growth* (Harvard Graduate School of Education Working Paper). Retrieved from http://scholar.harvard.edu/mkraft/publications/productivity-returns-experience-teacher-labor-market-methodological-challenges-a

Papay, J. P., West, M. R., Fullerton, J. B., & Kane, T. J. (2012). Does an urban teacher residency increase student achievement? Early evidence from Boston. *Education Evaluation and Policy Analysis*, *34*, 413–434.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*, 247–252.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, *6*, 43–74.

Ronfeldt, M. (2012). Where should student teachers learn to teach? Effects of field placement school characteristics on teacher retention and effectiveness. *Educational Evaluation and Policy Analysis*, *34*, 3–26.

Sass, T., Hannaway, J., Xu, Z., Figlio, D., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower-poverty schools. *Journal of Urban Economics*, *72*, 104–122.

Steinberg, M. P., Allensworth, E., & Johnson, D. W. (2011, May). *Student and teacher safety in Chicago Public Schools: The roles of community context and school social organization*. Consortium on Chicago School Research. Retrieved from http://ccsr.uchicago.edu/downloads/8499safety_in_cps.pdf

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

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

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

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher*, *37*, 469–479.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89–122.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, *100*, 61–78.

Authors

MATTHEW A. KRAFT is an assistant professor of education at Brown University. His research and teaching interests include the economics of education, education policy analysis, and applied quantitative methods for causal inference. Specifically, he studies

human capital policies in education with a focus on teacher effectiveness and organizational change in K-12 urban public schools.

JOHN P. PAPAY is an assistant professor of education and economics at Brown University and a research affiliate with the Project on the Next Generation of Teachers at Harvard. His research focuses on teacher policy, the economics of education, and teacher labor markets.

Manuscript received March 25, 2013
First revision received September 27, 2013
Second revision received December 11, 2013
Accepted December 11, 2013